Published in final edited form as:

Phys Biol. ; 6(1): 15001. doi:10.1088/1478-3975/6/1/015001.

What have we learned from the studies of two-state folders, and what are the unanswered questions about two-state protein folding?

Doug Barrick

T. C. Department of Biophysics, The Johns Hopkins University, 3400 N. Charles St., Baltimore MD 21218

Abstract

Small proteins with globular structures often fold by simple all-or-none mechanisms, both in an equilibrium and a kinetic sense, despite the very large number of partly folded conformations available. This type of “two-state” folding will be discussed in terms of experimental tests, underlying molecular mechanisms, and limits to two-state behavior. Factors that appear to be important for two-state folding include topology (sequence distance of contacts in the native structure), molecular cooperativity, and local energy distribution. Because their local stability distributions and cooperativities can be dissected and analyzed separately from topological features, recent studies of the folding of symmetric proteins will be discussed as a means to better understand the origins of two-state folding.

Keywords

Folding; two-state; cooperativity; topology

Biological systems are epitomized by a high level of organization, both in space and in time. This organization has both inherited and spontaneous components. Genome organization and development takes temporal cues passed down over the course of evolution and reproduction. However, organization of macromolecules (both intra- and intermolecular) into complex structures is often spontaneous. Protein folding is one of the more spectacular examples of self-organization seen in biological systems. Often without the input of external energy, small monomeric proteins can fold into very organized, compact, structures that show little structural variation around the biologically active “native state”. This folding process often involves a hundred or more residues, corresponding to thousands of protein (and solvent) atoms.

A key process in folding is the selection of narrow ranges of values for hundred of backbone and side-chain dihedral angles along the polypeptide chain, which uniquely determine the protein fold. Even in an equilibrium (long-time) sense, protein folding is remarkable: the number of degrees of freedom available to the polypeptide chain is very large, yet a narrow, well-defined ensemble of structures (the native ensemble) is the most stable (lowest free energy) ensemble. More remarkably, proteins fold quickly: for many small globular proteins, the search for the native state happens in a time-scale of between seconds, milliseconds, and even microseconds (Yang & Gruebele, 2003; Arora *et al.*, 2004; Kubelka *et al.*, 2004). Similarly remarkable, the protein folding reaction is often *mechanistically* simple, proceeding in what appears in experimental studies as a concerted all-or-none reaction.

This review will focus on aspects of the concerted, all-or-none nature of the protein folding reaction. This abrupt transition is often referred to as a “two-state” reaction, meaning that only two macroscopically distinct “states” are populated (the denatured and the native ensembles), and along with the rate-limiting transition-state ensemble, determine the rate of folding and unfolding. Two-state behavior is seen both at equilibrium and in kinetic studies. Owing to the differences between equilibrium and kinetic studies, the two state character has different manifestations in the two experimental formats. For this reason, a distinction will be made between “equilibrium” and “kinetic” two-state behavior. However, it may be expected that there are structural and thermodynamic features (recent studies point to cooperativity and to the compensation of entropy and enthalpy) that are common to the equilibrium and kinetic facets of two-state behavior.

Equilibrium two-state behavior

If only two distinct structural states are populated, structurally sensitive probes (e.g. fluorescence, CD, shape-sensitive coefficients) should have contributions solely from signals of the native or denatured ensembles. Although this discrete variation should be seen at the single molecule level (Figure 1A), in most folding studies the observed signal is averaged over a large number of molecules, where signals from the two ensembles are incompletely resolved. Thus, even for a strict equilibrium two-state mechanism, the observable signal reports a population-weighted average of the native and denatured ensembles, and assumes intermediate values between the two limits (Figure 1B).

Although ensemble measurements of a two-state equilibrium lead to intermediate signals, equilibrium transitions induced between the native and denatured ensembles still have features suggestive of an all-or-none reaction. Most notably, when conditions are adjusted to favor the denatured ensemble (often by increasing temperature, pressure, or adding chemical denaturants (Privalov & Gill, 1988; Royer, 2002; Street *et al.*, 2008)), a sharp sigmoidal transition is observed to span the signals of the native and denatured ensembles (Figure 1B). The position of the observed signal reflects the relative populations of the native and denatured ensembles, and thus on the equilibrium constant and free energy difference. The steepness of the unfolding transition has been shown to roughly reflect the size of the cooperative unit (Privalov & Gill, 1988; Myers *et al.*, 1995); in the two-state limit, the cooperative unit spans the entire molecule. The avoidance of partly folded states in the equilibrium two-state limit is equivalent to existence of a (free) energy barrier between the native and denatured ensembles, when energy is represented as a function of a relevant conformational coordinate (Lumry & Biltonen, 1966; Oliveberg & Wolynes, 2005).

Kinetic two-state behavior

An alternative molecular view to equilibrium two-state behavior, which helps motivate kinetic two-state folding, is that if the conformation of a single polypeptide is followed for a long time (Figure 1A), it will spend nearly all of its time within the native and denatured ensembles, interconverting very rapidly between the two. The relative amount of time (on average) spent between the two conformations determines the equilibrium constant. The mean dwell-time in the denatured and native ensembles determines the folding and unfolding rates. For a few proteins, such traces have been recorded at the single-molecule level, either using optical tweezers at constant force (Cecconi *et al.*, 2005), or using fluorescence resonance energy transfer (see (Schuler & Eaton, 2008)). According to the ergodic hypothesis of statistical thermodynamics (Hill, 1986), the rapid interconversion between the native and denatured states is consistent a rarification of partly folded states in a long-term average, either as a result of high (unfavorable) molecular energies, or from a restriction in the number of partly folded

states. Collectively, this rarefaction can be viewed as a free energy barrier (either enthalpic or entropic) that interconnect the native and denatured ensembles.

The kinetic two-state mechanism interprets this barrier as a single transition state ensemble (TSE) which is common to folding (from the denatured ensemble) and unfolding (from the native ensemble; Figure 2A), rather than a set of high energy intermediates and/or barriers which contribute to kinetics, depending on the direction and conditions in which the reaction takes place. Mechanistically the kinetic two-state model can be described with following scheme:

When combined with a conservation of mass relation, the rate laws implied by this simple kinetic scheme lead to single exponential kinetics in both the folding and unfolding direction, with an exponent $k_f + k_u$, (the sum of the microscopic rate constants k_f and k_u , respectively; (Fersht, 1999)). In the steady state, a defining relationship between rate and equilibrium constants is established:

$$\frac{d[N]}{dt} = +k_f[D] - k_u[N] = 0; \quad (1A)$$

$$\frac{k_u}{k_f} = \frac{[D]}{[N]} = K_{eq} \quad (1B)$$

In other words, if there is only a single kinetic barrier between the folded and unfolded states, the equilibrium constant should match the ratio of the folding and unfolding rate constants.

Although the test for kinetic two-state behavior (equation 1B) is very simple, there are some experimental issues that prevent direct evaluation. The kinetic quantity that is experimentally accessible, $k_{app} = k_f + k_u$, does not uniquely determine the two underlying rate constants. Instead, k_f and k_u can be determined from folding and unfolding transients under conditions that strongly favor the native and denatured states, respectively (typically low and high denaturant concentrations). However, owing to their denaturant dependences (see below), values of k_f and k_u determined independently under native and denaturing conditions cannot be directly compared using equation (1B). Thus, the kinetic two-state test requires the denaturant dependence of the rate (and equilibrium) constants to be accounted for.

When plotted as a function of denaturant concentration, the log of k_{app} is V-shaped. Owing to this shape, these rate-denaturant plots are often called “chevron plots” (Matthews, 1987). The minimum in the V-shape corresponds closely to the unfolding midpoint C_m . At denaturant concentrations well below C_m , the observed dependence of k_{app} is dominated by an exponential decrease in k_f with denaturant, whereas well above C_m , the denaturant dependence is dominated by an exponential increase in k_u . Assuming that the log-linear denaturant dependence seen for K_{eq} translates into k_f and k_u , the denaturant dependence of the microscopic and apparent rate constants can be given as:

$$\log k_f = \log k_{f,H_2O} + m_f[x]/2.3RT \quad (2A)$$

$$\log k_u = \log k_{u,H_2O} + m_u [x] / 2.3RT \quad (2B)$$

$$\log (k_f + k_u) = \log (k_{f,H_2O} \times 10^{m_f [x] / 2.3RT} + k_{u,H_2O} \times 10^{m_u [x] / 2.3RT}) \quad (2C)$$

For a chevron plot with linear arms (Figure 2B), m_f and m_u are constants of opposite sign. The factor $2.3RT$ in the denominator sets the scale to match that typically used for the equilibrium denaturant dependence, m_{eq} . The values k_{f,H_2O} and k_{u,H_2O} are folding and unfolding rate constants extrapolated to zero denaturant concentration. Equation 2C can be fitted to a chevron plot to give k_{f,H_2O} , k_{u,H_2O} , m_f , and m_u . Such analysis provides the basis to apply the test for kinetic two-state behavior (equation 1B) at any denaturant concentration, and as a further test of the generality of the mechanism, to evaluate the equality

$$m_u - m_f = m_{eq} \quad (3)$$

This relationship among the kinetic and equilibrium m -values generalizes the kinetic two-state model to the entire range of accessible denaturant concentrations.

Over the last decade, several small globular proteins have been shown to fold and unfold with simple exponential kinetics, and have chevron plots with linear folding and unfolding arms (see (Maxwell *et al.*, 2005) for an extensive gallery of chevron plots collected under standard conditions), and satisfy the simple kinetic two-state criteria above (equations 1B, 3). An early and compelling example of this is exhibited by CI2, a small (86 residue) protein (Jackson & Fersht, 1991). Following this example, a number of proteins have been found to exhibit similar behavior over a broad range of experimental conditions.

A thorough comparison of two-state folding kinetics for a large number of different proteins (Jackson, 1998) shows that kinetic two-state folding can be achieved for a variety of folds (e.g. α vs. β - vs mixed α/β), and over a variety of timescales (Chen *et al.*, 1989; Arora *et al.*, 2004). Likewise, kinetic two-state folding is exhibited for proteins that vary significantly in the sequence proximity of native-state (contact order, see below). The clearest qualitative distinction separating kinetic two-state proteins from those with more complicated folding mechanisms is chain length. In the compilation by (Jackson, 1998), the average chain-length among two-state folders, not counting close structural homologues, is 85 residues (rmsd=19, max=126). An updated list from Bai *et al.* ((Bai *et al.*, 2004) gives a slightly smaller average (75 residues). In contrast, for multistate folders the average chain-length is 136 (rmsd=38, max=219; Jackson, 1998). Although the average length of kinetically two-state folders is significantly less than that of multistate folders, the two distributions have clear overlap. In other words, no chain length cutoff neatly separates the two behaviors. Thus, there must be additional factors that promote (or prevent) kinetic two-state folding. As a clear demonstration that length is not the only factor controlling folding mechanism, both two- and three-state kinetics have been seen for the same protein depending on conditions (e.g. temperature, sulfate; (Khorasanizadeh *et al.*, 1993; Spudich *et al.*, 2004)) or point substitution (Laub *et al.*, 1995; Spudich *et al.*, 2004) (Friel *et al.*, 2004). In a few cases, two-state and multistate kinetic pathways have been shown to coexist simultaneously (Kiefhaber, 1995; Matagne *et al.*, 1997; Thirumalai & Hyeon, 2005).

Although a number of studies have added to both the two- and three-state lists in the last decade, the above picture—an approximate limit to the size that key elements of structure can be organized in a concerted step—remains intact. Current questions that are motivated by

consideration of two-state folding are: 1) What are the key elements of structure that contribute to two-state behavior? 2) What can we learn from exceptional two-state folders? 3) How does equilibrium cooperativity relate to two-state folding? 4) What is this size limit for two-state folding? Although these questions remain unanswered, research over the last decade has provided a number of important clues, which will be described below.

Structural features the transition state ensemble

As depicted in Figure 2A, the structure of the TSE is often considered to be midway between the denatured and native ensembles. Given the key role of the TSE in determining the kinetics of folding and unfolding, and the role the TSE plays, along with the rest of the energy landscape, in simplifying the mechanism to a two-state reaction, it seems appropriate to begin to address the questions listed above by describing the current view of structure in the TSEs of kinetic two-state folders.

By definition, TSEs are high in energy, and thus cannot be isolated and characterized directly. Thus, much of the information available on the structure of the TSE comes from comparing the sensitivity of rate and equilibrium constants for folding to physical perturbation. Some types of perturbations influence stability in a global, nonspecific way, including variations in temperature (which provides access to entropy, enthalpy, and most significantly, heat capacity of the TSE (Chen *et al.*, 1989)), denaturant concentration (which provides information on surface area burial and backbone exposure in the TSE), and more recently pressure (which provides information on the molar volume of the TSE, and indirectly, the extent of hydration (Royer, 2008)). These global probes provide a picture in which the heat capacity and surface area burial from solvent in the TSE are midway between the native and denatured state (but vary considerably from protein to protein, and can be rather uncorrelated (Plaxco *et al.*, 1998)). The molar volume of the TSE is often close to the native state, indicating substantial dehydration (Royer, 2008), although this quantity seems to be surprisingly sensitive to sequence perturbation (Mitra *et al.*, 2007).

Structural information on the TSE can be obtained at much higher structural resolution by introducing local perturbations to the protein sequence. This is done by comparing the rates and equilibrium constants of proteins with side-chain substitutions at a single localized sites. Early studies of this type were largely based on point-mutations isolated from genetic screens (Matthews, 1987). With the advent of oligonucleotide-mediated site-directed mutagenesis techniques (Shortle *et al.*, 1981; Zoller & Smith, 1983), substitutions could be focused at the level of individual sites and high chemical precision (Goldenberg *et al.*, 1989; Matouschek *et al.*, 1989). To analyze these substitutions quantitatively, Fersht and coworkers introduced an index, referred to as the Φ -value, to represent the extent of structure formation at the site of substitution in the TSE (Matouschek *et al.*, 1989). A Φ -value of 1, which results when the folding rate and equilibrium constants are equally perturbed by substitution, suggests substantial, native-like structure at the site of substitution in the TSE. A Φ -value of 0, which results when the folding rate is insensitive to a (typically) destabilizing substitution, suggests a lack of structure. This approach has since been used by a large number of laboratories to characterize the details of structure formation in the TSE of two-state (and non-two-state) protein folding (see (Daggett & Fersht, 2003; Zarrine-Afsar & Davidson, 2004; Raleigh & Plaxco, 2005; Royer, 2008) for reviews).

Although the details of TSE structures deduced from Φ -value analysis on two-state proteins are diverse, some recurrent themes have emerged. Frequently, there is a localized region of structure formation (moderate Φ -values) in the interior of the protein (termed the nucleus), often involving only partly formed elements of secondary structure, and the extent of structure formation becomes gradually weaker extending out the nucleus. This pattern has been

interpreted as a “nucleation-condensation” mechanism for folding. Examples of proteins with TSEs matching this mechanism are CI2, FKBP12, and the FN10 domain of fibronectin (Itzhaki *et al.*, 1995; Fulton *et al.*, 1999; Cota *et al.*, 2001). In other small globular proteins, full units of secondary structure appear to be fully formed in the TSE, consistent with the “framework” or “diffusion-collision” model for folding (Mayor *et al.*, 2003). The relative contribution of the nucleation-condensation and framework mechanisms depends, in large part, on local secondary structure propensities (White *et al.*, 2005). In still other proteins, including several SH3 domains, the TSE seems more polarized, with residues at a few sites showing significant structure (high Φ -values), and other sites showing little structure (Grantcharova *et al.*, 1998; Martinez & Serrano, 1999; Northey *et al.*, 2002). Studies in which side-chain identity is systematically varied have suggested loose burial of hydrophobic side-chains stabilizes the TSEs of some proteins (Northey *et al.*, 2002; Anil *et al.*, 2005), whereas other TSEs show stabilization by native-like hydrogen bond networks (Grantcharova *et al.*, 1998).

The measurement of Φ -values by a number of laboratories on a number of proteins has provided a unique, highly detailed experimental picture of TSE structure in the folding of two-state proteins. Perhaps because of the descriptive nature of Φ -value analysis and the diversity of TSE properties observed using Φ -value analysis, it is hard to generalize or to make specific predictions regarding transition state structure and absolute folding rates based on primary sequence and native structure alone. Instead, the large data set resulting from Φ -value analysis of two-state protein folding has served as a benchmark to develop theoretical (Clementi *et al.*, 2000; Weikl & Dill, 2003) (Alm *et al.*, 2002) and computational (Daggett *et al.*, 1996; Paci *et al.*, 2002; Schaeffer *et al.*, 2008) models for protein folding, which in turn have filled in with even more structural detail. What (if any) general features of protein structures and sequence can be used to predict folding rates and learn about TSE structures?

Contact order and mechanism

Given the rather diverse, case-by-case view of TSE structure provided by Φ -value analysis, are there any general structural properties of two-state proteins that determine the kinetics of folding and thus relate to TSE structure? One key observation that holds general insights into the structural determinants of folding kinetics is that for kinetically two-state proteins, folding rates are inversely correlated with the sequence distance between contacts in the native state (Plaxco *et al.*, 1998; Plaxco *et al.*, 2000). This dependence on native-state topology has been parameterized as “relative contact order”, and was initially used to help explain a large ($\sim 10^6$ -fold) difference in k_f values among two-state folders. Subsequently, a variety of indices based on contact distribution in the native state have been shown to lead to similar correlations (Gromiha & Selvaraj, 2001; Dokholyan *et al.*, 2002; Zhou & Zhou, 2002), as has secondary structure content (Gong *et al.*, 2003). Such metrics have also been extended to three-state folding (Ivankov *et al.*, 2003), and in some cases, appear to discriminate between kinetic two- and three-state folders, at least at the level of adjustable parameters (Zhang & Sun, 2005).

How can the observed contact order correlation for kinetic two-state proteins be combined with has been learned about TSE structure from Φ -value analysis? As described below, the contact order correlation suggests that a close approximation to the native topology must be realized in the transition state ensemble. This view is most consistent with a nucleation-condensation mechanism, in which most of the protein contributes, albeit weakly, to the structure of the TSE, and to a high denaturant sensitivity to the folding rate (a “late” TSE), but is less consistent with framework or polarized mechanisms. Experimental studies have provided additional insight into what the contact-order correlation means for the rate-limiting steps in folding (see (Sosnick, 2008) for a recent review). Through a novel two-site mutagenesis scheme analogous to Φ -value analysis (referred to as ψ -value analysis), Sosnick and coworkers have probed the overall topology of the TSEs of three different proteins that span the range of contact order

observed for globular proteins (Krantz *et al.*, 2004; Pandit *et al.*, 2006; Sosnick, 2008). The transition states of these three proteins all appear to have the same fraction of the native-state topology (~70 percent), suggesting that this native state metric may indeed play a general role in transition states in protein folding, although the full native-state topology may not be required. A similar conservation of transition state topology has been seen in simulations of protein folding that have been constrained by experimental Φ -values (Paci *et al.*, 2005; Wallin & Chan, 2006).

What does this striking contact order correlation tell us about the mechanism of two-state folding? A relationship between topology and protein folding rates has been viewed to be consistent with a model of protein folding in which aspects of the energy landscape are funneled toward the native state, with pairwise native (but *not* non-native) interactions making the dominant contribution to stability (Onuchic *et al.*, 2000; Oliveberg & Wolynes, 2005; Thirumalai & Hyeon, 2005). In this view, the barrier required to produce exponential kinetics has been described as a mismatch between decreases in the internal energy of the chain, free energy of the solvent, and chain entropy. The observed dependence of the folding rate on contact order (Plaxco *et al.*, 2000) suggests that the free energy barrier results from an entropy bottleneck that is encountered upon adopting the native topology (Shea *et al.*, 1999; Thirumalai & Klimov, 2007). This topological “frustration” is consistent with some (Clementi *et al.*, 2000; Koga & Takada, 2001) but not all (Cieplak & Hoang, 2003; Jewett *et al.*, 2003) simulations of folding using funneled (Gō-type) energy functions.

Both the general concepts behind energy landscape theory and the case-by-case simulations using funneled potentials lend consistency to the importance of topology in two-state folding. However, they do not pinpoint the precise physical interactions that underlie the topological correlation. Can the underlying interactions be abstracted from physical models to capture the correlation in a general way? In one early study predating the observed contact order correlation, Langevin dynamics of a simplified, albeit reasonable protein model governed by a general energy showed a pathway for direct folding in which the critical nucleus involved structuring “topology-preserving” regions near turns (Guo & Thirumalai, 1995). Like experimental two-state folders, this nucleation-collapse mechanism leads directly to the native state without intermediates via a single-exponential process. More recently, two groups have modeled the rate limiting step in two-state folding as a diffusive “topomer search” to a conformer that roughly approximates the native-state topology (Debe *et al.*, 1999; Makarov & Plaxco, 2003). Using a Gaussian chain model in their topomer search, Makarov *et al.* (Makarov *et al.*, 2002) were able to reproduce the observed topology dependence of two-state folding kinetics.

Influences of specific sequence and local stability

The topomer search models above are consistent with the entropic bottleneck postulated in landscape theory of two-state protein folding, and help provide in a mechanistic picture. But is this the whole story? Recent Langevin dynamics simulations of C α -models suggest such a search would be too inefficient to lead to productive folding (Wallin & Chan, 2005), suggesting that one or more general factors not included in the topomer-search model remain to be identified (although the fractional contact order results described above could improve the search problem). Consistent with this suggestion, experiments have shown clear exceptions to the contact order correlation. Proteins with homologous folds, and even single residue substitutions, can have very different folding rates (Jackson & Fersht, 1991; Munson *et al.*, 1997), although the entropic bottlenecks are expected to be quite similar.

Perhaps the clearest examples indicating that specific interactions impart an energetic bias that influences folding rate comes from experimental studies of proteins with internal symmetry.

Φ -value analysis in protein-L and protein-G, which have an approximate two-fold rotation symmetry, shows a clear preference for one of the two symmetry-related pathways. Moreover, different symmetry-related pathways are used in the two proteins, reflecting sequence (and presumably local stability) differences (Kim *et al.*, 2000; McCallister *et al.*, 2000). Linear repeat proteins, with rough translational symmetry, have also been shown to fold by discrete pathways (Tang *et al.*, 2003; Bradley & Barrick, 2005). For two ankyrin repeat proteins, folding pathways have recently been shown to be determined by local energetics, and can be shifted by changing local stabilities (Lowe & Itzhaki, 2007; Kloss *et al.*, 2008; Tripp & Barrick, 2008). These findings are consistent with earlier results in globular proteins (Englander *et al.*, 2002). Building such energetic factors into simulations of two-state folding remains an ongoing challenge (Karplus & Weaver, 1994; Burton *et al.*, 1998; Klimov & Thirumalai, 1998; Clementi *et al.*, 2003; Merlo *et al.*, 2005; Barrick *et al.*, 2008).

Recently, models have been developed that can account for the influence of local sequence variation, while retaining the global influence of chain topology. Bai *et al.* searched for minimal fragments that retain native state topology, and showed a correlation of the size of such fragments with folding rate (Bai *et al.*, 2004). This type of partial topological control would shorten the search time, and would provide a mechanism by which different regions of the protein could have different effects on transition state energetics. In a more mechanistic approach, Weikl and Dill (Weikl & Dill, 2003) modeled folding as a series of reactions in which native state contact clusters are formed in discrete steps. By constructing loop closure entropies that are dependent on pre-formed intervening loops, which decrease the “effective contact order”, the topology dependence of two-state folding kinetics was recovered, along with coarse agreement with transition state structures mapped from experimental Φ -value analysis. Importantly, by analyzing the time-dependence of structure formation using a master equation approach, single exponential kinetics was recovered. The barriers identified from the master equations correspond to formation of local clusters, which in turn facilitate the formation of otherwise slow nonlocal clusters (Weikl *et al.*, 2004).

Cooperativity and two-state folding

Cooperativity is the central defining feature of equilibrium two-state folding. Distant regions of proteins stabilize each other through indirect interactions mediated by the intervening protein matrix. An extreme example of cooperativity is seen in the linear repeat proteins (Zweifel & Barrick, 2001; Mosavi *et al.*, 2002; Courtemanche & Barrick, 2008a; Kloss *et al.*, 2008), where elongated modular proteins (as large as 400 residues; (Kloss & Barrick, 2008) have been shown to fold by a concerted equilibrium mechanism. Surprisingly, this cooperativity is even higher than for globular proteins, which at similar chain-length, tend to unfold in multistate equilibrium reactions. For two ankyrin repeat constructs, analysis of the length dependence of equilibrium stability shows this high level of cooperativity to result from strongly stabilizing interfaces between unstable repeat units (Mello & Barrick, 2004; Wetzel *et al.*, 2008). Moreover, this cooperativity has been shown to be sensitive to the balance of local stability distribution: asymmetric shifts in the energy landscape result in highly populated equilibrium intermediates (Street *et al.*, 2007).

Is this extreme cooperativity enough to enforce two-state kinetics? For most repeat proteins, the answer is “no”. Most of the repeat proteins studied to date show multistate folding kinetics; often an on-pathway kinetic intermediate separates the native and denatured states (Tang *et al.*, 1999; Mello *et al.*, 2005; Low *et al.*, 2007). To date, the sole exception is the leucine-rich repeat (LRR) domain of internalin-B (Inl-B, ~210 residues), composed of seven repeats and an N-terminal α -helical cap (Courtemanche & Barrick, 2008a). What permits this large¹ modular domain to fold by a concerted mechanism? The stability distribution of this LRR domain is somewhat bipolar: both ends of the molecule appear to be essential for equilibrium

stability (N. Courtemanche & DB, unpublished), which would be expected to provide a balanced stability distribution across the molecule, promoting equilibrium two-state behavior (Street *et al.*, 2007). Despite this balance, only the N-terminal region of the molecule appears to be structured in the TSE (Courtemanche & Barrick, 2008b), perhaps a result of the N-terminal helical cap. To conform to a two-state kinetic scheme, the C-terminus must become structured late in the reaction (based on measured Φ -values), must add quickly to the already structured N-terminus (and dissociate quickly from it in unfolding), but must not serve as a kinetically competent folding route without structure in the N-terminal domain, in addition to being essential for native state stability. The difference in kinetics of the two termini, despite their mutual stabilization of the domain, may be essential to avoid parallel folding pathways and/or persistent partly folded kinetic intermediates (Lowe & Itzhaki, 2007). The extent to which the observed bipolar stability distribution promotes kinetic cooperativity in this large protein remains to be determined.

Although the high level of cooperativity seen in most large repeat proteins does not appear to be sufficient for kinetic two-state folding, a number of simulations have identified cooperativity as a key issue in kinetic two-state folding. By modifying a simple G $\ddot{o}$ -type potential such that native interactions become more stabilizing as the folded state is reached, the Plaxco and Chan groups showed that the correlation between contact order and folding rates of a 3-D cubic lattice model is significantly improved, that variation in folding rate constants is increased (Jewett *et al.*, 2003; Kaya & Chan, 2003a), and that single-exponential behavior is reasonably approximated (Faisca & Plaxco, 2006). Another set of lattice simulations introducing cooperativity at the level of helix formation and core packing resulted in chevron plots significantly closer to two-state behavior (Kaya & Chan, 2003b), a finding that has recently been extended to off-lattice simulations (Liu & Chan, 2005).

Unresolved questions

Above, four questions relating to two-state folding were posed. As promised, they have not definitively answered. However, significant insight has been developed from theory and experiment. Both topology and local stability appear to be key components of two-state folding of small globular proteins; thus, models that include both of these components are expected to provide significant insight into folding. Cooperativity also appears to be a key component of two-state folding. An important question to be addressed is “what are the cooperative interactions”? There is more than one way to build cooperativity into a molecular model; which ones are most relevant to simplify kinetics? Here, contributions from experiments will be critical. Because constructs of different lengths can be prepared and studied, it is hoped that that simple symmetric proteins can better define the origins of cooperativity and its relationship to folding, and further probe the upper limits of the remarkable two-state behavior of proteins.

References

- Alm E, Morozov AV, Kortemme T, Baker D. Simple physical models connect theory and experiment in protein folding kinetics. *J Mol Biol* 2002;322(2):463–476. [PubMed: 12217703]
- Anil B, Sato S, Cho JH, Raleigh DP. Fine structure analysis of a protein folding transition state; distinguishing between hydrophobic stabilization and specific packing. *J Mol Biol* 2005;354(3):693–705. [PubMed: 16246369]
- Arora P, Oas TG, Myers JK. Fast and faster: a designed variant of the B-domain of protein A folds in 3 microsec. *Protein Sci* 2004;13(4):847–853. [PubMed: 15044721]

¹At 213 residues, Inl-B is more than twice as large as the typical kinetic two-state protein, exceeding the maximum reported by Jackson (1998) by a factor of 1.7.

- Bai Y, Zhou H, Zhou Y. Critical nucleation size in the folding of small apparently two-state proteins. *Protein Sci* 2004;13(5):1173–1181. [PubMed: 15075405]
- Barrick D, Ferreiro DU, Komives EA. Folding landscapes of ankyrin repeat proteins: experiments meet theory. *Curr Opin Struct Biol* 2008;18(1):27–34. [PubMed: 18243686]
- Bradley CM, Barrick D. Effect of multiple prolyl isomerization reactions on the stability and folding kinetics of the notch ankyrin domain: experiment and theory. *J Mol Biol* 2005;352(2):253–265. [PubMed: 16054647]
- Burton RE, Myers JK, Oas TG. Protein folding dynamics: quantitative comparison between theory and experiment. *Biochemistry* 1998;37(16):5337–5343. [PubMed: 9548914]
- Cecconi C, Shank EA, Bustamante C, Marqusee S. Direct observation of the three-state folding of a single protein molecule. *Science* 2005;309(5743):2057–2060. [PubMed: 16179479]
- Chen BL, Baase WA, Schellman JA. Low-temperature unfolding of a mutant of phage T4 lysozyme. 2. Kinetic investigations. *Biochemistry* 1989;28(2):691–699. [PubMed: 2653428]
- Cieplak M, Hoang TX. Universality classes in folding times of proteins. *Biophys J* 2003;84(1):475–488. [PubMed: 12524300]
- Clementi C, Garcia AE, Onuchic JN. Interplay among tertiary contacts, secondary structure formation and side-chain packing in the protein folding mechanism: all-atom representation study of protein L. *J Mol Biol* 2003;326(3):933–954. [PubMed: 12581651]
- Clementi C, Nymeyer H, Onuchic JN. Topological and energetic factors: what determines the structural details of the transition state ensemble and "enroute" intermediates for protein folding? An investigation for small globular proteins. *J Mol Biol* 2000;298(5):937–953. [PubMed: 10801360]
- Cota E, Steward A, Fowler SB, Clarke J. The folding nucleus of a fibronectin type III domain is composed of core residues of the immunoglobulin-like fold. *J Mol Biol* 2001;305(5):1185–1194. [PubMed: 11162123]
- Courtemanche N, Barrick D. Folding thermodynamics and kinetics of the leucine-rich repeat domain of the virulence factor Internalin B. *Protein Sci* 2008a;17(1):43–53. [PubMed: 18156467]
- Courtemanche N, Barrick D. The Leucine-Rich Repeat Domain of Internalin B Folds along a Polarized N-Terminal Pathway. *Structure* 2008b;16(5):705–714. [PubMed: 18462675]
- Daggett V, Fersht AR. Is there a unifying mechanism for protein folding? *Trends Biochem Sci* 2003;28(1):18–25. [PubMed: 12517448]
- Daggett V, Li A, Itzhaki LS, Otzen DE, Fersht AR. Structure of the transition state for folding of a protein derived from experiment and simulation. *J Mol Biol* 1996;257(2):430–440. [PubMed: 8609634]
- Debe DA, Carlson MJ, Goddard WA 3rd. The topomer-sampling model of protein folding. *Proc Natl Acad Sci U S A* 1999;96(6):2596–2601. [PubMed: 10077555]
- Dokholyan NV, Li L, Ding F, Shakhnovich EI. Topological determinants of protein folding. *Proc Natl Acad Sci U S A* 2002;99(13):8637–8641. [PubMed: 12084924]
- Englander SW, Mayne L, Rumbley JN. Submolecular cooperativity produces multi-state protein unfolding and refolding. *Biophys Chem* 2002;101(102):57–65. [PubMed: 12487989]
- Faisca PF, Plaxco KW. Cooperativity and the origins of rapid, single-exponential kinetics in protein folding. *Protein Sci* 2006;15(7):1608–1618. [PubMed: 16815915]
- Fersht, AR. *Structure and Mechanism in Protein Science: a Guide to Enzyme Catalysis and Protein Folding*. New York: W.H. Freeman & Co.; 1999.
- Friel CT, Beddard GS, Radford SE. 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* 2004;342(1):261–273. [PubMed: 15313622]
- Fulton KF, Main ER, Daggett V, Jackson SE. Mapping the interactions present in the transition state for unfolding/folding of FKBP12. *J Mol Biol* 1999;291(2):445–461. [PubMed: 10438631]
- Goldenberg DP, Frieden RW, Haack JA, Morrison TB. Mutational analysis of a protein-folding pathway. *Nature* 1989;338(6211):127–132. [PubMed: 2465497]
- Gong H, Isom DG, Srinivasan R, Rose GD. Local secondary structure content predicts folding rates for simple, two-state proteins. *J Mol Biol* 2003;327(5):1149–1154. [PubMed: 12662937]

- Grantcharova VP, Riddle DS, Santiago JV, Baker D. Important role of hydrogen bonds in the structurally polarized transition state for folding of the src SH3 domain. *Nat Struct Biol* 1998;5(8):714–720. [PubMed: 9699636]
- Gromiha MM, Selvaraj S. Comparison between long-range interactions and contact order in determining the folding rate of two-state proteins: application of long-range order to folding rate prediction. *J Mol Biol* 2001;310(1):27–32. [PubMed: 11419934]
- Guo Z, Thirumalai D. Kinetics of Protein Folding: Nucleation Mechanism, Time Scales, and Pathways. *Biopolymers* 1995;36:83–102.
- Hill, TL. *An Introduction to Statistical Thermodynamics*. Mineola, N.Y: Dover Publications, Inc.; 1986.
- Itzhaki LS, Otzen DE, Fersht AR. The structure of the transition state for folding of chymotrypsin inhibitor 2 analysed by protein engineering methods: evidence for a nucleation-condensation mechanism for protein folding. *J Mol Biol* 1995;254(2):260–288. [PubMed: 7490748]
- Ivankov DN, Garbuzynskiy SO, Alm E, Plaxco KW, Baker D, Finkelstein AV. Contact order revisited: influence of protein size on the folding rate. *Protein Sci* 2003;12(9):2057–2062. [PubMed: 12931003]
- Jackson SE. How do small single-domain proteins fold? *Fold Des* 1998;3(4):R81–R91. [PubMed: 9710577]
- Jackson SE, Fersht AR. Folding of chymotrypsin inhibitor 2. 1. Evidence for a two-state transition. *Biochemistry* 1991;30(43):10428–10435. [PubMed: 1931967]
- Jewett AI, Pande VS, Plaxco KW. Cooperativity, smooth energy landscapes and the origins of topology-dependent protein folding rates. *J Mol Biol* 2003;326(1):247–253. [PubMed: 12547206]
- Karplus M, Weaver DL. Protein folding dynamics: the diffusion-collision model and experimental data. *Protein Sci* 1994;3(4):650–668. [PubMed: 8003983]
- Kaya H, Chan HS. Contact order dependent protein folding rates: kinetic consequences of a cooperative interplay between favorable nonlocal interactions and local conformational preferences. *Proteins* 2003a;52(4):524–533. [PubMed: 12910452]
- Kaya H, Chan HS. Simple two-state protein folding kinetics requires near-levinthal thermodynamic cooperativity. *Proteins* 2003b;52(4):510–523. [PubMed: 12910451]
- Khorasanizadeh S, Peters ID, Butt TR, Roder H. Folding and stability of a tryptophan-containing mutant of ubiquitin. *Biochemistry* 1993;32(27):7054–7063. [PubMed: 8392867]
- Kiefhaber T. Kinetic traps in lysozyme folding. *Proc Natl Acad Sci U S A* 1995;92(20):9029–9033. [PubMed: 7568066]
- Kim DE, Fisher C, Baker D. A breakdown of symmetry in the folding transition state of protein L. *J Mol Biol* 2000;298(5):971–984. [PubMed: 10801362]
- Klimov DK, Thirumalai D. Linking rates of folding in lattice models of proteins with underlying thermodynamic characteristics. *J. Chem. Phys* 1998;109(10):4119–4125.
- Kloss E, Barrick D. Thermodynamics, Kinetics and salt-dependence of unfolding of YopM, a large leucine-rich repeat protein. *J Mol Biol*. Submitted.
- Kloss E, Courtemanche N, Barrick D. Repeat-protein folding: new insights into origins of cooperativity, stability, and topology. *Arch Biochem Biophys* 2008;469(1):83–99. [PubMed: 17963718]
- Koga N, Takada S. Roles of native topology and chain-length scaling in protein folding: a simulation study with a Go-like model. *J Mol Biol* 2001;313(1):171–180. [PubMed: 11601854]
- Krantz BA, Dothager RS, Sosnick TR. Discerning the structure and energy of multiple transition states in protein folding using psi-analysis. *J Mol Biol* 2004;337(2):463–475. [PubMed: 15003460]
- Kubelka J, Hofrichter J, Eaton WA. The protein folding 'speed limit'. *Curr Opin Struct Biol* 2004;14(1):76–88. [PubMed: 15102453]
- Laub PB, Khorasanizadeh S, Roder H. Localized solution structure refinement of an F45W variant of ubiquitin using stochastic boundary molecular dynamics and NMR distance restraints. *Protein Sci* 1995;4(5):973–982. [PubMed: 7663353]
- Liu Z, Chan HS. Solvation and desolvation effects in protein folding: native flexibility, kinetic cooperativity and enthalpic barriers under isostability conditions. *Phys Biol* 2005;2(4):S75–S85. [PubMed: 16280624]

- Low C, Weininger U, Zeeb M, Zhang W, Laue ED, Schmid FX, Balbach J. Folding mechanism of an ankyrin repeat protein: scaffold and active site formation of human CDK inhibitor p19(INK4d). *J Mol Biol* 2007;373(1):219–231. [PubMed: 17804013]
- Lowe AR, Itzhaki LS. Rational redesign of the folding pathway of a modular protein. *Proc Natl Acad Sci U S A* 2007;104(8):2679–2684. [PubMed: 17299057]
- Lumry R, Biltonen R. Validity of the "two-state" hypothesis for conformational transitions of proteins. *Biopolymers* 1966;4(8):917–944. [PubMed: 5975643]
- Makarov DE, Keller CA, Plaxco KW, Metiu H. How the folding rate constant of simple, single-domain proteins depends on the number of native contacts. *Proc Natl Acad Sci U S A* 2002;99(6):3535–3539. [PubMed: 11904417]
- Makarov DE, Plaxco KW. The topomer search model: A simple, quantitative theory of two-state protein folding kinetics. *Protein Sci* 2003;12(1):17–26. [PubMed: 12493824]
- Martinez JC, Serrano L. The folding transition state between SH3 domains is conformationally restricted and evolutionarily conserved. *Nat Struct Biol* 1999;6(11):1010–1016. [PubMed: 10542091]
- Matagne A, Radford SE, Dobson CM. Fast and slow tracks in lysozyme folding: insight into the role of domains in the folding process. *J Mol Biol* 1997;267(5):1068–1074. [PubMed: 9150396]
- Matouschek A, Kellis JT Jr, Serrano L, Fersht AR. Mapping the transition state and pathway of protein folding by protein engineering. *Nature* 1989;340(6229):122–126. [PubMed: 2739734]
- Matthews CR. Effect of point mutations on the folding of globular proteins. *Methods Enzymol* 1987;154:498–511. [PubMed: 3431461]
- Maxwell KL, Wildes D, Zarrine-Afsar A, De Los Rios MA, Brown AG, Friel CT, Hedberg L, Horng JC, Bona D, Miller EJ, Vallee-Belisle A, Main ER, Bemporad F, Qiu L, Teilum K, Vu ND, Edwards AM, Ruczinski I, Poulsen FM, Kragelund BB, Michnick SW, Chiti F, Bai Y, Hagen SJ, Serrano L, Oliveberg M, Raleigh DP, Wittung-Stafshede P, Radford SE, Jackson SE, Sosnick TR, Marqusee S, Davidson AR, Plaxco KW. Protein folding: defining a "standard" set of experimental conditions and a preliminary kinetic data set of two-state proteins. *Protein Sci* 2005;14(3):602–616. [PubMed: 15689503]
- Mayor U, Gudyosh NR, Johnson CM, Grossmann JG, Sato S, Jas GS, Freund SM, Alonso DO, Daggett V, Fersht AR. The complete folding pathway of a protein from nanoseconds to microseconds. *Nature* 2003;421(6925):863–867. [PubMed: 12594518]
- McCallister EL, Alm E, Baker D. Critical role of beta-hairpin formation in protein G folding. *Nat Struct Biol* 2000;7(8):669–673. [PubMed: 10932252]
- Mello CC, Barrick D. An experimentally determined protein folding energy landscape. *Proc Natl Acad Sci U S A* 2004;101(39):14102–14107. [PubMed: 15377792]
- Mello CC, Bradley CM, Tripp KW, Barrick D. Experimental characterization of the folding kinetics of the notch ankyrin domain. *J Mol Biol* 2005;352(2):266–281. [PubMed: 16095609]
- Merlo C, Dill KA, Weikl TR. Phi values in protein-folding kinetics have energetic and structural components. *Proc Natl Acad Sci U S A* 2005;102(29):10171–10175. [PubMed: 16009941]
- Mitra L, Hata K, Kono R, Maeno A, Isom D, Rouget JB, Winter R, Akasaka K, Garcia-Moreno B, Royer CA. V(i)-value analysis: a pressure-based method for mapping the folding transition state ensemble of proteins. *J Am Chem Soc* 2007;129(46):14108–14109. [PubMed: 17960902]
- Mosavi LK, Williams S, Peng Z. Equilibrium Folding and Stability of Myotrophin: A Model Ankyrin Repeat Protein. *J. Mol. Biol* 2002;320:165–170. [PubMed: 12079376]
- Munson M, Anderson KS, Regan L. Speeding up protein folding: mutations that increase the rate at which Rop folds and unfolds by over four orders of magnitude. *Fold Des* 1997;2(1):77–87. [PubMed: 9080201]
- Myers JK, Pace CN, Scholtz JM. Denaturant m values and heat capacity changes: relation to changes in accessible surface areas of protein unfolding. *Protein Sci* 1995;4(10):2138–2148. [PubMed: 8535251]
- Northey JG, Di Nardo AA, Davidson AR. Hydrophobic core packing in the SH3 domain folding transition state. *Nat Struct Biol* 2002;9(2):126–130. [PubMed: 11786916]
- Oliveberg M, Wolynes PG. The experimental survey of protein-folding energy landscapes. *Q Rev Biophys* 2005;38(3):245–288. [PubMed: 16780604]

- Onuchic JN, Nymeyer H, Garcia AE, Chahine J, Socci ND. The energy landscape theory of protein folding: insights into folding mechanisms and scenarios. *Adv Protein Chem* 2000;53:87–152. [PubMed: 10751944]
- Otzen DE, Kristensen O, Proctor M, Oliveberg M. Structural changes in the transition state of protein folding: alternative interpretations of curved chevron plots. *Biochemistry* 1999;38(20):6499–6511. [PubMed: 10350468]
- Paci E, Lindorff-Larsen K, Dobson CM, Karplus M, Vendruscolo M. Transition state contact orders correlate with protein folding rates. *J Mol Biol* 2005;352(3):495–500. [PubMed: 16120445]
- Paci E, Vendruscolo M, Dobson CM, Karplus M. Determination of a transition state at atomic resolution from protein engineering data. *J Mol Biol* 2002;324(1):151–163. [PubMed: 12421565]
- Pandit AD, Jha A, Freed KF, Sosnick TR. Small proteins fold through transition states with native-like topologies. *J Mol Biol* 2006;361(4):755–770. [PubMed: 16876194]
- Plaxco KW, Simons KT, Baker D. Contact order, transition state placement and the refolding rates of single domain proteins. *J Mol Biol* 1998;277(4):985–994. [PubMed: 9545386]
- Plaxco KW, Simons KT, Ruczinski I, Baker D. Topology, stability, sequence, and length: defining the determinants of two-state protein folding kinetics. *Biochemistry* 2000;39(37):11177–11183. [PubMed: 10985762]
- Privalov PL, Gill SJ. Stability of protein structure and hydrophobic interaction. *Adv Protein Chem* 1988;39:191–234. [PubMed: 3072868]
- Raleigh DP, Plaxco KW. The protein folding transition state: what are Phi-values really telling us? *Protein Pept Lett* 2005;12(2):117–122. [PubMed: 15723637]
- Royer CA. Revisiting volume changes in pressure-induced protein unfolding. *Biochim Biophys Acta* 2002;1595(1–2):201–209. [PubMed: 11983396]
- Royer CA. The nature of the transition state ensemble and the mechanisms of protein folding: a review. *Arch Biochem Biophys* 2008;469(1):34–45. [PubMed: 17923105]
- Sanchez IE, Kiefhaber T. Evidence for sequential barriers and obligatory intermediates in apparent two-state protein folding. *J Mol Biol* 2003;325(2):367–376. [PubMed: 12488101]
- Schaeffer RD, Fersht A, Daggett V. Combining experiment and simulation in protein folding: closing the gap for small model systems. *Curr Opin Struct Biol* 2008;18(1):4–9. [PubMed: 18242977]
- Schuler B, Eaton WA. Protein folding studied by single-molecule FRET. *Curr Opin Struct Biol* 2008;18(1):16–26. [PubMed: 18221865]
- Shea JE, Onuchic JN, Brooks CL 3rd. Exploring the origins of topological frustration: design of a minimally frustrated model of fragment B of protein A. *Proc Natl Acad Sci U S A* 1999;96(22):12512–12517. [PubMed: 10535953]
- Shortle D, DiMaio D, Nathans D. Directed mutagenesis. *Annu Rev Genet* 1981;15:265–294. [PubMed: 6279018]
- Sosnick TR. Kinetic barriers and the role of topology in protein and RNA folding. *Protein Sci* 2008;17(8):1308–1318. [PubMed: 18502978]
- Spudich GM, Miller EJ, Marqusee S. Destabilization of the Escherichia coli RNase H kinetic intermediate: switching between a two-state and three-state folding mechanism. *J Mol Biol* 2004;335(2):609–618. [PubMed: 14672667]
- Street TO, Bradley CM, Barrick D. Predicting coupling limits from an experimentally determined energy landscape. *Proc Natl Acad Sci U S A* 2007;104(12):4907–4912. [PubMed: 17360387]
- Street TO, Courtemanche N, Barrick D. Protein folding and stability using denaturants. *Methods Cell Biol* 2008;84:295–325. [PubMed: 17964936]
- Tang KS, Fersht AR, Itzhaki LS. Sequential Unfolding of Ankyrin Repeats in Tumor Suppressor p16. *Structure* 2003;11:67–73. [PubMed: 12517341]
- Tang KS, Guralnick BJ, Wang WK, Fersht AR, Itzhaki LS. Stability and Folding of the Tumour Suppressor Protein p16. *J. Mol. Biol* 1999;285(4):1869–1886. [PubMed: 9917418]
- Thirumalai D, Hyeon C. RNA and protein folding: common themes and variations. *Biochemistry* 2005;44(13):4957–4970. [PubMed: 15794634]
- Thirumalai D, Klimov DK. Intermediates and transition states in protein folding. *Methods Mol Biol* 2007;350:277–303. [PubMed: 16957328]

- Tripp KW, Barrick D. Rerouting the folding pathway of the Notch ankyrin domain by reshaping the energy landscape. *J Am Chem Soc* 2008;130(17):5681–5688. [PubMed: 18396879]
- Wallin S, Chan HS. A critical assessment of the topomer search model of protein folding using a continuum explicit-chain model with extensive conformational sampling. *Protein Sci* 2005;14(6):1643–1660. [PubMed: 15930009]
- Wallin S, Chan HS. Conformational entropic barriers in topology-dependent protein folding. *J. Phys. Cond. Matter* 2006;18:S307–S328.
- Weikl TR, Dill KA. Folding rates and low-entropy-loss routes of two-state proteins. *J Mol Biol* 2003;329(3):585–598. [PubMed: 12767836]
- Weikl TR, Palassini M, Dill KA. Cooperativity in two-state protein folding kinetics. *Protein Sci* 2004;13(3):822–829. [PubMed: 14978313]
- Wetzel SK, Settanni G, Kenig M, Binz HK, Pluckthun A. Folding and unfolding mechanism of highly stable full-consensus ankyrin repeat proteins. *J Mol Biol* 2008;376(1):241–257. [PubMed: 18164721]
- White GW, Gianni S, Grossmann JG, Jemth P, Fersht AR, Daggett V. Simulation and experiment conspire to reveal cryptic intermediates and a slide from the nucleation-condensation to framework mechanism of folding. *J Mol Biol* 2005;350(4):757–775. [PubMed: 15967458]
- Yang WY, Gruebele M. Folding at the speed limit. *Nature* 2003;423(6936):193–197. [PubMed: 12736690]
- Zarrine-Afsar A, Davidson AR. The analysis of protein folding kinetic data produced in protein engineering experiments. *Methods* 2004;34(1):41–50. [PubMed: 15283914]
- Zhang L, Sun T. Folding rate prediction using n-order contact distance for proteins with two- and three-state folding kinetics. *Biophys Chem* 2005;113(1):9–16. [PubMed: 15617806]
- Zhou H, Zhou Y. Folding rate prediction using total contact distance. *Biophys J* 2002;82(1 Pt 1):458–463. [PubMed: 11751332]
- Zoller MJ, Smith M. Oligonucleotide-directed mutagenesis of DNA fragments cloned into M13 vectors. *Methods Enzymol* 1983;100:468–500. [PubMed: 6225933]
- Zweifel ME, Barrick D. Studies of the ankyrin repeats of the *Drosophila melanogaster* Notch receptor. 2. Solution stability and cooperativity of unfolding. *Biochemistry* 2001;40(48):14357–14367. [PubMed: 11724547]

Figure 1. Equilibrium two-state folding

(A) On a single-molecule basis, an observable parameter that is sensitive to structure (such as intrinsic fluorescence or far-UV CD) takes on only two values corresponding to the native and denatured state. The average of the time spent at each value reflects the equilibrium constant for folding. For a kinetically two-state mechanism, the transitions are expected to be abrupt in time, as shown. (B) In an ensemble average, a steep but continuous transition is seen between native and denatured baselines (dashed and dotted lines, respectively). Although the average equilibrium signal lies at an intermediate value between these two limits, individual members of the ensemble remain restricted to the native and denatured states (as depicted in panel A).

Data are for CD-monitored unfolding of the Notch ankyrin domain, reproduced from (Street *et al.*, 2008) with permission.

Figure 2. Kinetic two-state folding

(A) A highly simplified reaction diagram for two-state protein folding, in which Gibbs free energy is depicted as a function of a general reaction coordinate ξ , which resolves the native, denatured, and transition state ensembles. In terms of a free energy diagram, a single transition-state ensemble (indicated by the double-dagger) acts as a barrier to folding, determining the rate constant for folding and unfolding. (B) A linear chevron plot for a kinetically two-state protein. The solid line depicts k_{app} ; k_f and k_u are shown as dotted and dashed lines, respectively. Linear folding limbs are obtained if the transition state depicted in panel (A) remains the highest free energy point on the dominant routes to folding (i.e. no new barriers appear (Sanchez & Kiefhaber, 2003)), and does not change its structure with denaturant (Otzen *et al.*, 1999).