The folding of single domain proteins — have we reached a consensus?

Tobin R Sosnick[1] and Doug Barrick[2]

Rather than stressing the most recent advances in the field, this review highlights the fundamental topics where disagreement remains and where adequate experimental data are lacking. These topics include properties of the denatured state and the role of residual structure, the nature of the fundamental steps and barriers, the extent of pathway heterogeneity and non-native interactions, recent comparisons between theory and experiment, and finally, dynamical properties of the folding reaction.

Addresses
[1] Department of Biochemistry and Molecular Biology, Institute for Biophysical Dynamics, University of Chicago, Chicago, IL 60637, USA
[2] T.C. Jenkins Department of Biophysics, Johns Hopkins University, 3400 N. Charles St., Baltimore, MD 21218, USA

Corresponding author: Sosnick, Tobin R (trsosnic@uchicago.edu)

Introduction

The field of protein folding generally is considered to be mature. This maturation process has seen the development of a number of methods and theoretical constructs for studying folding, and their application has led to some agreement on general principles. Yet, current views and results addressing many fundamental issues remain remarkably diverse even for the folding behavior of single domain proteins. Although some of the diversity can be attributed to a difference in folding properties of individual proteins, much controversy still remains on many fundamental topics including the importance of residual structure in denatured states, the nature of folding steps and kinetic barriers as well the extent of pathway diversity. Rather than stressing the generally agreed-on principles (which, as often as not, vary depending on who is doing the agreeing), this review is intended to highlight the fundamental topics where disagreement remains in the context of recent studies. It is hoped that focusing on these disagreements will help to bring about critical tests, and ultimately, a more coherent picture of the folding process.

We have organized this review to parallel the folding reaction. We begin by discussing the starting point for proteins, the denatured state ensemble (DSE). We then focus on folding events, and highlight where experiments have provided a rich picture of the reaction, and where they have yet to shed light. Next we discuss the nature of intermediates connected by fundamental folding events, and the extent to which intermediates are structured and solvated. We next focus the principle of topology as a major determinant of folding kinetics followed by a discussion the cases for and against multiple folding paths, which provides a major route to testing the theories and simulations. Here, there are some real (and perhaps unexpected) consistencies between theory and experiment. We finally turn to recent work on solvent viscosity and internal friction, a relatively new, or at least resurgent, experimental approach that connects to theory.

Residual structure and its significance

A considerable amount of recent work has focused on the denatured state and it serves as a logical starting point. Considerable evidence exists for residual structure in the DSE [1] although its importance in guiding the folding process is unclear. Building on earlier site-resolved studies, the power of NMR methods has been exploited in a number of recent works [2]. For example, the cold [3] and acid denatured states [4] of C-terminal domain of the ribosomal protein L9 are seen to contain significant native- and non-native secondary structures. Chemical shift analysis indicates that three native helical segments are mutually stabilized in the denatured state of a four-helix bundle, acyl coenzyme A binding protein (ACBP) [5]. A combination of biophysical methods has been utilized to generate partially structured, compact DSE in drk N-terminal SH3 [6]. Thermodynamic studies indicate that differences in the amount of residual structure are related to thermostability for homologs of RNase H1 [7,8].

In spite of these and previous demonstrations of residual structure, the generality and productivity of residual structure remains an unresolved issue in protein folding. For example, the aforementioned transient helical assembly in denatured ACBP is not the same as in the transition state ensemble (TSE) which contains a different combination of three helices. In addition, another chemical shift analysis conducted on all β protein, c-src SH3, observed only non-native helical structure [9]. Likewise, no coherent three-dimensional structure was observed in phosphoglycerate kinase using paramagnetic relaxation

enhancement [10]. Generally, proteins select amino acids because of their native-like secondary structure propensities [11] and these biases may be strong enough to influence the denatured state, although the prevalence of polyproline II conformers in isolated peptides suggests that non-native conformations also are well sampled [12]. Residual structure can be productive, particularly in local structures such as helices or turns, but whether it is a significant factor in the folding of many proteins is debatable.

Another factor to consider is that residual structure is much less prevalent at high denaturant concentrations yet folding behavior often is similar under folding and unfolding conditions. Many kinetically two-state proteins have a robust TSE, a property that underlies the wide-spread application of the denaturant dependent 'chevron' analysis. The robustness of the TSE is directly supported by hydrogen exchange EX1-style measurements for proteins which fold on the millisecond [13] and microsecond time scale [14]; for these two proteins, folding and unfolding rates under native conditions match rates extrapolated from high denaturant concentration. Extremely linear chevron 'folding arms' are frequently observed even down to zero denaturant concentrations [15]. Nonlinear behavior would be expected if there are qualitative changes in folding behavior, such as movement of the TS [16,17], or the formation of residual structure, collapsed denatured states or intermediates. However, transfer models can retain linearity even with compaction of the DSE implying that linearity is insufficient to conclude that there are no qualitative changes in folding behavior [18–20].

Compaction in the DSE

A major unresolved topic is the compactness of the denatured state under native conditions, for example, right after dilution of denaturant in a refolding experiment. This issue has broad implications to folding mechanisms, computer simulations, thermodynamics as well as the interpretation of a wide variety of new and old experiments. Because hydrophobic interactions and intra-protein hydrogen bonds increase in strength at lower denaturant concentrations, the unfolded protein chain often is assumed to rapidly collapse to a more compact conformation having a radius of gyration (R_g) smaller than the random coil value observed at high denaturant [21]. In support of this plausible view are numerous single molecule fluorescence resonance energy transfer (smFRET) studies where properties of the DSE are readily distinguished from those of the native population [19,22–29]. Collapse also is observed in ensemble kinetic [30–33] and equilibrium studies [34]. Random sequences have increased diffusion coefficients under aqueous conditions [35]. These results are supported by the observation compact denatured states and early collapse phases in many theoretical studies [19,22,25,32,36•].

However, no early collapse phase is observed after denaturant dilution in a wide variety of small-angle X-ray scattering measurements (SAXS). These studies have been conducted on kinetically two state proteins including Protein L [37], ubiquitin and acyl-phosphatase [38] as well as on monomeric SOD [39] where revised studies indicate that the reduction in the R_g, if any, is quite small upon transfer to 1.5 M urea (C Kayatekin, O Bilsel, unpublished data). For a denaturant jump measurement on Rnase A, all but ~1 Å (the statistical error) of total 10 Å decrease in R_g upon folding to the native state is accounted for in the observed folding process [40]. In these kinetic measurements, the R_g is measured immediately after denaturant has been diluted while nearly all the molecules are still unfolded. Hence, these nonequilibrium measurements provide an alternative strategy to smFRET for measuring properties of the DSE under native conditions without the native state contributing to the observed signal. The SAXS results on Protein L are particularly notable as collapse has been observed in the smFRET studies although to different degrees [22,26].

The lack of compaction at low denaturant is also found in equilibrium SAXS studies on nonfolding analogs of RNase A [41,42] and hen egg white lysozyme [43]. These analogs, created by the reduction of the disulfide bonds, remain extended under aqueous conditions. Although the truncated version of *S. aureus* nuclease ($\Delta131\Delta$) is often cited as evidence for a compact denatured state, it can be destabilized by a single point mutant to generate a DSE under fully aqueous conditions that has the same R_g as the wild-type version in 8 M urea [44]. This system provides an important example as it displays both behaviors — a compact intermediate and an expanded DSE — and hence, it serves as an excellent control. Although these results indicate that many proteins can have expanded DSEs under aqueous conditions, SAXS measurements on other proteins have observed a compaction of the DSE. These proteins include drk SH3 [6,45] and in the burst phase for pH jumps from the acid denatured states of cytochrome *c* [46] and monellin [47]. It is unclear why the DSEs of these proteins behave differently than the other eight.

For the present purposes focusing on properties of denatured proteins, we distinguish the DSEs of globular proteins from intrinsically disordered proteins having hydrophilic sequences [48,49•,50,51] which are likely to behave differently [52], and from specific intermediates which are separated from the denatured state by free energy barrier. In larger proteins that fold in a multi-state manner, intermediates with stable secondary structures often are observed to be collapsed by SAXS [53–56].

In summary, SAXS measurements on a significant number of proteins indicate that compact denatured states and early collapse phases do not always occur. Hence, these

events are not obligatory. The origin of the substantial disparity between these results and those observed with other methods is unknown. This issue should be resolved as it impacts a wide range of ongoing experimental and theoretical studies and their interpretation.

Folding events and pathways

Despite the deceptively simple two-state mechanism by which many proteins fold, the complexity of even small proteins implies the folding reaction does not involve the simultaneous collision and isomerization of thousands of atoms and bonds. From an experimental perspective, resolving the underlying events from a simple exponential folding reaction remains a major challenge. Characteristics of the TSE can be deduced from rate-equilibrium free energy relationships, either through chemical modification (including isotope effects [57,58], ϕ [59,60] and ψ analyses [61–64]) or by physical [65,66] or solvent perturbation.

However, the transient events leading to and from the major barrier are particularly difficult to access experimentally (Figure 1). This experimental limitation is particularly problematic for identifying the early, weakly populated species going up the (initial) barrier leading to the first stably collapsed species [67,68]. As a result, the

portion of the pathway is largely hidden going from the DSE to the TS for two-state proteins, and from the DSE to the TS leading to the first stable intermediate for multistate proteins. For many single domain proteins, this ignorance is particularly significant, as a majority of the structure and topology is formed before the TS [69,70,71,72•] and one can only speculate on the nature of the early transient steps. For example, even if a helix with tertiary interactions is present in the TS, one cannot identify whether the helix forms before or after it makes tertiary contacts [72•,73].

Nevertheless, evidence has been obtained for native-like long-range strand–strand interactions in unstable, preTS intermediate using ψ-analysis with engineered bihistidine (biHis) binding sites (Figure 2) [74]. For a biHis site between the terminal β strands of ubiquitin, the ψ value is unity indicating that metal binding exerts its full thermodynamic effect during the passage over the kinetic barrier. The establishment of the binding equilibrium can only occur if the otherwise unstable intermediate having the interstrand binding site is present often enough before the TS that it can encounter and bind the free metal ions.

On the native side of the barrier, the post-TS pathway has been easier to characterize, partly because the native state

Figure 1

Current Opinion in Structural Biology
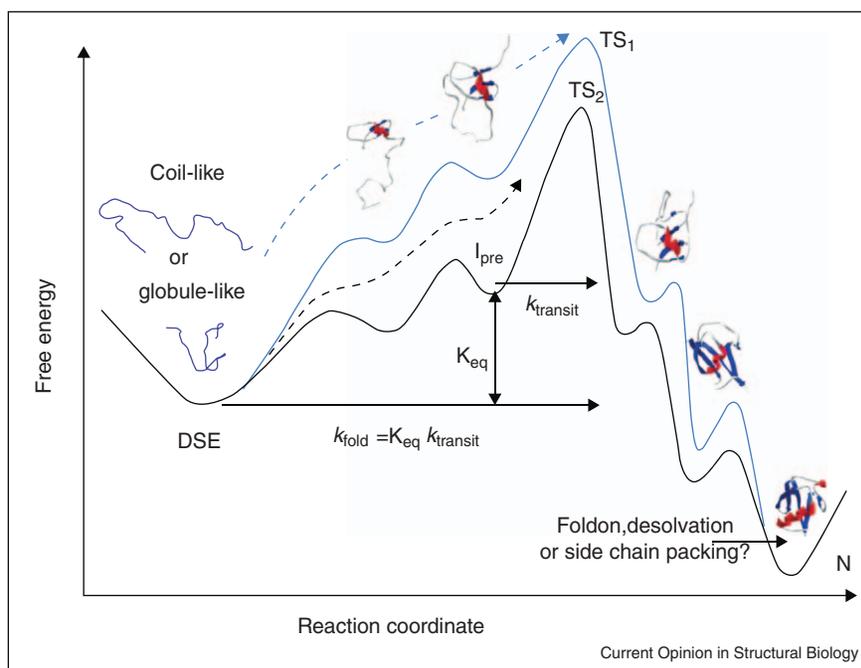
Kinetically two-state reaction surface illustrating possible folding events. The DSE may be collapsed or expanded and there may be more than one pathway to and from the TS (black and blue traces). Steps between the TS and the native state represent the addition of foldons and are likely to be added in a particular sequence. These behaviors also may apply to the steps leading up to the TS, which are largely hidden from experimental characterization. Different views describe the last folding events as the addition of a foldon, desolvation of the core or the locking of side chains from a dry MG state. In the scheme depicted here, k_{fold}, the observed folding rate, is the product of K_{eq}, the equilibrium constant for I_{pre}, the highest energy intermediate in fast equilibrium with the DSE, times k_{trans}, the transit rate going over the barrier at the top.

Figure 2
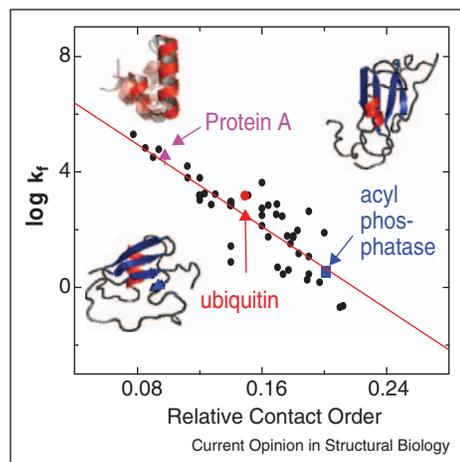
Correlation between RCO and folding rate, and TS structures determined using ψ analysis. The line indicates the best fit to the data, which may represent proteins which form ~70% of the native topology in the TS. This level is based on the RCO value observed in the three proteins whose representative TS structures were determined using Langevin dynamic simulations constrained by experimental ψ values [62,72,140].

can act as both an experimental and conceptual reference for the analysis of structure and because partly folded states can be more stable than the unfolded state. Relevant methods to characterize these species include native state hydrogen exchange (NSHX) [75] and thiol exchange [76,77] as well as NMR relaxation dispersion (RD) experiments [78–81]. The exchange studies indicate that the pathway down from the TS occurs via the addition of foldons, defined as cooperative formation of pieces of secondary structure or loops, often contiguous in sequence, in a process of sequential stabilization [75,82–87,88*] (*note*: the term 'foldon' is also used to refer to a elemental building block analogous to a subdomain [89]). In this well-established picture, protein folding is not a completely cooperative process even for proteins with apparent two-state kinetics or thermodynamic behavior; rather, folding often is punctuated with the smaller scale, subglobal folding events [82].

As previously noted, the early events going from the DSE to the TS are largely uncharacterized. Nevertheless, the principles of foldon substructure and sequential stabilization observed in the post-TS folding steps are likely to apply to the early portion of the pathway [62,64,90]. In this view, the early folding process involves an initial search through a multitude of mostly unproductive unstable conformations. Some are more populated and can serve as better foundations for additional foldons. These folding steps involve the build-up of well-formed local and long-range structure with commensurate levels of hydrogen bonding and surface burial [57,58,62]. These

principles can be used to predict folding pathways and native structures with high accuracy in a strategy involving iterative fixing of secondary structures [91,92]. The nucleation condensation model puts a stress on the formation of a critical subset of long-range tertiary contacts in the TS while the remainder of the protein makes a variety of diffuse interactions [93,94].

Intermediates and desolvation
Characterization of partly folded intermediates is a key approach to understanding the protein folding pathways [95]. Major strategies range from characterizing transient intermediates during refolding experiments to studying equilibrium species populated through the use of unnatural conditions (e.g. low pH and cosolvents) or mutagenesis [96–99]. Another major strategy involves the detection of sparsely populated equilibrium intermediates, for example, using NSHX and NMR RD methods [78–81,100,101,102*]. These methods have provided structural information on a large number of intermediates, particularly in the last few years.

Here we describe the current picture, with emphasis on more recent novel work, and again highlight where interpretations and generalizations agree, and where they differ. Three variables relevant to the structure of intermediates are the extent to which they firstly, have rigid packing versus fluid side chains; secondly, have desolvated cores; and thirdly, have native versus non-native structural features.[a] Intermediate conformations that span these structural descriptors have been characterized for some time, and have had complementary predictions from theory and computation. Molten globule (MG) intermediates are an important class of intermediates that are undergoing a resurgence [103*]. Structural features include a retention of a near-native amount of secondary structure, but the loss of near-UV circular dichroism signal resulting from the disruption of the tertiary packing of aromatic side chains, and considerably decreased chemical shift dispersion in NMR spectra.

The reported degree of hydration ranges from highly solvated (wet molten globules) to highly desolvated (dry molten globules). Solvent penetration into the core of native proteins has been observed in all-atom simulations and solvent expulsion is postulated to be a late step in folding [104–107]. Recent experimental data have been interpreted in the context of a dry MG [108,109], although direct evidence of side chain fluidity is lacking in these studies unlike the original NMR investigations [110,111]. The proposition that the dry MG is a general,

---

[a] Here, 'non-native' structure is not meant to include structure that is indistinguishable from the denatured state ensemble, but rather to secondary structure formation, secondary structure association, or packing that has been acquired during folding (from the denatured state ensemble) and is different from the native structure.

well populated unfolding intermediate [103[•]] is at odds with the apparent two-state folding behavior observed for many proteins, for example, as demonstrated using chevron analysis where all the free energy and the surface burial of the equilibrium reaction are accounted for in the single observed reaction.

A simplifying view of hydration and side chain packing can be found in the studies of an equilibrium analog of a late folding intermediate of ubiquitin created using a 'Protein Vivisection' strategy [86]. An aliphatic core residue is substituted with acidic residue so that at pH 7, only the foldon having the substitution is disrupted. Upon charge neutralization under mildly acidic conditions, the protein returns to its native form, as indicated by ^{1}H–^{15}N HSQC NMR measurements. In the intermediate at pH 8, half of the residues retain their native chemical shifts, one third of the resonances move and the remaining resonances disappear due to dynamics on the NMR time scale. The retention of the native resonances for residues on three β strands and the associated helix indicates that a part of the protein has native-like packing with a desolvated core, while the remainder of the protein is partially or fully hydrated and dynamic. Kinetic and NSHX measurements on this intermediate suggest that is a kinetically relevant species on the folding pathway.

These NMR data on the ubiquitin intermediate indicate that neither global side chain locking nor core dehydration is the last folding step. Hence, ubiquitin does not follow either the wet or dry MG model. Rather, this protein folds via successive folding steps involving the addition of foldons with native-like packing and dehydration occurring in the more organized regions before the final step. This view is consistent with the capillarity picture [112] as well as experimental results for IM7 where desolvation occurs early during folding [113], and structural studies of other equilibrium intermediates. These intermediates have well packed regions with dehydrated cores while other regions are disordered or dynamic, or have non-native conformations [80,102[•],114–116]. Although chemical shifts can be less dispersed than in the native state, the fact that intermediates often have unique chemical shifts that deviate from random coil values and narrow linewidths [3,84,102[•],114] indicates a greater level of structure than would be expected for a MG. Nevertheless, other intermediates display poor NMR spectra due to dynamics [98,99]. It would be of great interest to use these NMR methods to test the extent of secondary and tertiary structures in the potentially dry MGs described above.

Non-native structures
Evidence for kinetically significant non-native interactions continues to grow. Highly structured yet off-pathway intermediates can accumulate due to prematurely formed interactions [117], suggesting that subdomain competition for hydrophobic residues is a general feature

of proteins with multiple modules [118]. Simulations [119] indicate that the complicated folding kinetics seen for the designed protein TOP7 [120] result from non-native interactions involving an excessively hydrophobic β strand. Similarly, non-native interactions can influence the kinetics of Fyn SH3 domain [121].

Intermediates can have non-native contacts as a result of the chain readjusting to maximize hydrophobic burial in the absence of all the native structural elements. The slow folding of the well-known apomyoglobin intermediate arises in part due to non-native interactions between the G and H helices [122,123]. The 3000-fold slowing for two versions of α-spectrin, a simple three-helix bundle, is postulated to result from helix misdocking even though folding remains kinetically two-state [124[•]]. This interpretation is significant because it implies that a non-native interaction formed even in an unstable species can slow folding and hence, the intermediate is an obligate species on the folding pathway.

The kinetic barrier for the stable IM7 intermediate, which also appears to be obligate and productive, involves the partial unfolding of a three-helix bundle [125]. Potentially, partial unfolding of an obligate intermediate also applies to RNase H1 folding because the two-state and three-state folding behavior can be manipulated by point mutants yet folding occurs with the same order [126]. For both a three-helix [127] and four-helix bundle [102[•]], an intermediate with the majority of the native structure forms in microseconds. But, the rate-limiting step requires the breaking of non-native interactions. These studies support the contention that intermediates generally fold slowly due to misfold/reorganization barriers formed in the initial folding phases [68,128].

Transition state topology
The two-state folding of single domain proteins has been proposed to be a nucleation process with the chain attaining a coarse version of the native topology in the TS [67,129–131]. There is broad agreement that topology is a critical feature in the determination of folding rates because of the correlation between folding rates and relative contact order (RCO) for many single domain proteins which fold in a two-state manner (Figure 2) [132,133]. This topic continues to be an active research area [119,126] with a wide variety of models recapitulating the observed RCO correlation [134–139].

For three proteins with disparate RCO values, ubiquitin, acyl phosphatase and the B-domain of Protein A, their TSEs have acquired a similar level of native topology, $RCO^{TS} \approx 0.7 \cdot RCO^{N}$ [62,72[•],140]. These studies used ψ analysis, a method that is extremely well suited for determining the topology of the TS because it directly identifies interresidue contacts (Figure 3). Such contacts were used as constraints in all-atom simulations to create

Figure 3
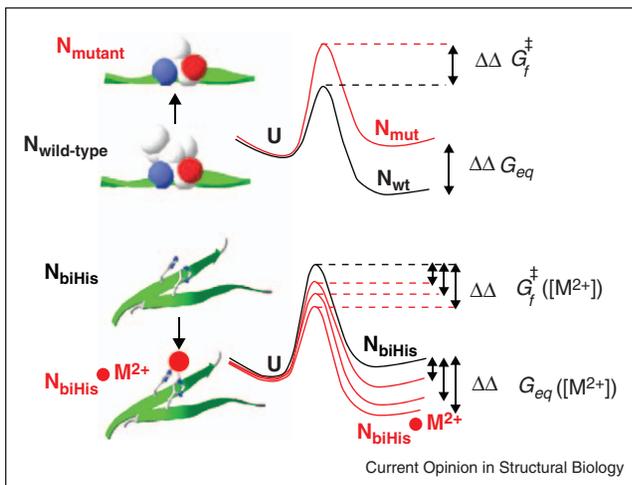
Characterizing transition states using ϕ and ψ analyses. In ϕ analysis, a side chain is mutated and the change in stability of the TS relative to the native state is quantified as $\phi = \Delta\Delta G_f^{\ddagger}/\Delta\Delta G_{eq}$. In ψ-analysis, a bihistidine (biHis) site is introduced at a location that is stabilized upon binding divalent metal ions (M^{2+}). The stabilization of the TS relative to the native state as a function of $[M^{2+}]$ is $\psi = \partial\Delta\Delta G_f^{\ddagger}([M^{2+}])/\partial\Delta\Delta G_{eq}([M^{2+}])$. Whereas a ϕ value can reflect native and non-native interactions as well as differences in secondary structure propensities of the two amino acids in the TSE, a ψ value reflects the binding of metal to the biHis site which is only possible when the two histidines are in proximity (e.g. β strand pairing or helix formation, depending on biHis location). Hence, this newer method identifies residue–residue contacts and it is especially powerful in the determination of TS structures and topologies.

atomic-level models of the TS (Figure 2) [64,90,140]. If the $RCO^{TS} \approx 0.7 \cdot RCO^N$ relationship observed for these three proteins is applicable to other proteins that obey the RCO–k_f trend, it would provide a simple rationalization for the trend as well as a constraint for possible TS structures for the other proteins. Similar relationships between the topology of the TS and the native state are found in theoretical studies [69,70,141].

A lower RCO fraction is observed in the study utilizing ϕ values [141], potentially because they underreport of chain–chain contacts [142] due to chain relaxation and accommodation [63,72•,143] and non-native interactions in the TS [144,145]. Underreporting by ϕ also would explain why TSs have been characterized as small and polarized in ubiquitin [146] and other proteins that obey the RCO correlation [147–152], in disagreement with the $RCO^{TS} \approx 0.7 \cdot RCO^N$ relationship deduced using ψ analysis. The application of ψ analysis to these other proteins could provide a critical test of the universality of the relationship as well as a useful comparison between the two methods for characterizing TSEs.

In addition, the experimental ϕ values for ubiquitin differ from those calculated using the TS structures obtained

from ψ-constrained all-atom simulations [143]. This difference likely is because the experimental values reflect energies while the simulated values reflect contacts. Hence, one should not always expect agreement between theory and experiment for ϕ values. Other concerns with ϕ relate to a separation of secondary structure propensities from tertiary interactions [153]. A recent large-scale analysis of ϕ values questioned their information content and concluded that the TSE is stabilized by local rather than specific tertiary interactions [154]. This conclusion is in contradiction to the aforementioned $RCO^{TS} \approx 0.7 \cdot RCO^N$ relationship.

Pathway diversity

There is a significant difference in opinion concerning whether TSs are heterogeneous and folding follows multiple pathways. Given the 1000s of atoms in the system, and the extensive amount of pathway diversity seen in many simulations with both simplified and all-atom models, it is plausible that there is considerable pathway diversity [89]. Nevertheless, direct experimental evidence is rather limited for TS heterogeneity for the same protein sequence as defined by the participation of different set of helices or strands in the TS, as opposed to local 'microscopic heterogeneity' such as a frayed helix [71]. Changes in TS structure have been observed for circular permutated versions of the S6 and α-spectrin SH3 proteins [89]. However, such sequences should be considered to be different proteins as they have a different topology and hence, they do not provide authentic examples of TS heterogeneity.

Obtaining direct evidence of TS heterogeneity in kinetically two-state systems is extremely challenging even with single-molecule methods. A viable strategy using ensemble methods involves the destabilization of a structured region in one TS and testing for a shift in flux to an alternative TS by remeasuring a ϕ or ψ value at a location which is likely to be structured in the alternative TS. This strategy was applied to CI2 [155], a three-helix bundle [72•], and α-spectrin SH3 [156], but no shift in flux to another TS was observed. Furthermore, the high percentage of native topology present in the TS for proteins obeying the RCO correlation ($RCO^{TS} \approx 0.7 \cdot RCO^N$), restricts the diversity of possible TS structures.

However, for an immunoglobulin domain, the pathway shifts from one TS to another, albeit one containing a subset of the same secondary structures [157]. Proteins with strong symmetries are other candidates for having multiple TSs [61,158–160], in particular those containing multiple subdomains such as hairpin + helix motifs [89,161]. The TS of Protein G can be shifted from the C-terminal to the N-terminal hairpin upon changes in their relative stability, implying that some intermediate sequence could fold with flux going through both TSs. TS heterogeneity is observed in the dimeric GCN4 coiled

coil, but it is lost when the translational symmetry is broken upon the introduction of a crosslink at one end [61,162]. For larger, linear repeat proteins, which also have translational symmetry, TSE shifts can be brought about readily by tipping the local stability distribution from one end of the molecule to the other [163,164].

Even when the TS is homogenous, the pathways to and from the TSE could be diverse. The principle of sequential stabilization implies that most foldons form in a well-defined sequence after crossing the initial barrier leading to collapse [88•]. But, when two foldons can be added along a pathway independently or with comparable energy, the pathway can temporarily bifurcate [62,64,165,166]. On the way up to the TS from the denatured state, the degree of pathway diversity is less clear. A multitude of unstable conformations can be sampled, but some structures provide more suitable templates for the addition of other elements. Accordingly, the uphill steps may still involve a largely sequential accretion of structure [62] although the energetic biases will not be as pronounced as for the post-TS pathways. Hence, the uphill steps may contain more diversity than the downward steps. Overall, it is likely that sequential stabilization in combination with topological and energetic biases result in a reaction surface with a dominant pathway. Nevertheless, the Boltzman distribution implies that many options are possible, and alternative paths are undoubtedly sampled.

Even for proteins that fold with multiple phases and the kinetic analysis appears to require parallel pathways, the order of native structure formation may still be the same along each pathway. In this scenario, the protein encounters different optional misfolding steps. But from the standpoint of the accretion native-like structure, the pathway still occurs with the same sequence (e.g. Helix$_A$→Helix$_B$→Helix$_C$ and Helix$_A$→Helix$_B$→(misfold & resolution)→Helix$_C$). This model, termed 'Predetermined pathways with optional errors' has been successfully applied to proteins which previously had been proposed to have parallel pathways, hen egg white lysozyme [88•] and Snase [85].

Theoretical studies
Landscape theory has investigated folding behavior on funnel-like landscapes having minimal energetic frustration [167]. Several gross features of landscape theory are consistent with the foldon-based mechanism. For example, both approaches emphasize assembly from native elements of structure, and neither view incorporates non-native structures as kinetically important *per se*. An apparent difference between the foldon description and the funneled landscape view is that the former uses a specific or preferred pathway to describe the acquisition of structure, whereas the latter suggests a highly parallel flux through many different nonoverlapping substructures [135,167–169]. This difference is certainly very real

at the conceptual level, and one would hope it would be a means for resolving the merits of the two views. Indeed it has motivated a number of experimental studies specifically designed to identify pathway heterogeneity, as discussed above.

However, when landscape theory is used to examine the folding of cytochrome *c* whose folding pathway has been resolved in great detail experimentally [166], the similarities seem more conspicuous than the differences. On a funneled energy landscape with multibody terms to heighten cooperativity [170,171•], discrete foldons assemble in a sequential folding mechanism consistent with the experimental picture. It is important to know whether such models having added complexity and heightened cooperativity produce similar behavior for nonheme containing proteins.

Other general features of experimental protein folding data, including kinetic two-state behavior and high thermodynamic cooperativity are not well represented using the simple native-centric models. These shortcomings can be rectified in large part by including energy terms with explicit desolvation [172] or cooperative terms that strengthen interactions as folding progresses [170,171•,173,174].

Although there has been considerable progress, identifying the structural content of the TSE remains a challenge. For BdpA, only one [175] of a dozen studies generated a TS model similar to the experimental TSE [72•]. Potentially, the TSE of this small three-helix bundle is challenging to characterize precisely because of its small size and symmetry so that subtle changes in the energetic balance between secondary and tertiary structure formation can greatly influence the order of events.

With an increase in computing power, all-atom simulations are able to fold some proteins completely to the native state [32,36•,176]. In spite of this progress, issues still remain with the accuracy of force fields, for example, the native state is not always the thermodynamic minimum in certain force fields [177] or the entropy/enthalpy balance is incorrect [178]. Improvements have been suggested [179].

From the experimentalist's perspective, a major advance in atomic-level folding simulations is the translation of multiple folding trajectories using Markov models into a single reaction surface with clustered intermediates and defined interconversion rates [36•,180,181•]. In addition to providing testable predictions, the connectivity between states is visualized so that degree of pathway heterogeneity and sequentiality is readily discerned. Using these methods, a 'hub model' was found [36•] that has a particularly nonfunnel-like characteristic. Even though there are numerous folding pathways, they are

determined by the initial chain configuration, rather than the standard view that the DSE reconfigures quickly so that all molecules have time to search for the lowest energy pathways. The postulate that the DSE is kinetically frozen is difficult to reconcile with the commonly held observation that (with the exception of proline isomerization), folding rates and amplitudes are independent of how the DSE was prepared.

Diffusion and internal friction
The dynamical properties of the denatured states have been studied using a variety of techniques [24,31,32,182]. These studies point to a significant decrease in chain diffusion rates in the denatured state under native conditions, which is generally attributed to collapse in the denatured state. Why this slowing of chain diffusion does not affect the linearity of the chevron arms is unclear.

Another topic with conflicting results and interpretations is the contribution of solvent viscosity and internal friction to the overall folding rate [183–185]. Some folding studies obtain a Kramers-like inverse viscosity dependence for the folding rate without a significant contribution of internal friction [186–188]. Internal friction, however, is implicated in the folding reaction rate for villin [189], the small binding domain BBL [182], and the molten globule of cytochrome c [190], while solvent friction alters the folding pathway of tryptophan zipper TZ2 [191]. For homologs of α-spectrin, internal friction is implicated only for the slower folding versions, being attributed to helix misdocking [124•].

For a folding landscape with early, unstable intermediates that are in rapid equilibrium with the DSE relative to the overall folding rate (Figure 1), the relevant dynamical properties for the overall reaction are only those for the steps at the top of the free energy surface ($k_{transit}$ in Figure 1). This step could be a small-scale folding event such as the addition of a single foldon [74]. Accordingly, the standard single barrier implementation of Kramers theory should be modified when it is applied to protein folding.

Downhill folding remains a contentious issue [192–194]. It should be appreciated that a variety of downhill folding scenarios exists. In one scenario, no free energy barrier exists between the DSE and the native state. In another scenario applicable to fast folding proteins, the height of the free energy barrier is small relative to thermal energies so that chain diffusion times over the barrier are significant and folding rates are probe dependent [195]. The precise conditions for this scenario are unclear as a monomeric coiled coil can fold in microseconds according to multiple probes located throughout the protein. Hence, the microsecond folding of a simple protein can be limited by a high barrier even under strongly native conditions [14]. Another downhill scenario can

occur if the changes in reaction surface occur quickly relative to population reconfiguration times (e.g. after a rapid temperature jump), so that a subpopulation now resides on a portion of the surface where it no longer has to traverse a free energy barrier in order to fold [196].

Numerous basic issues remain concerning the reaction dynamics in protein folding [185]. These topics include internal friction: What is it and how should it be included in a rate equation? Why is internal friction seen in some systems and not others — does it depend on protein size, topology or the folding speed? Other items relate the validity of the isoenergetic analysis for determining whether rates inversely scale with viscosity [187], what are transit times over the barrier [23]; and finally, does the diffusion constant significantly vary across the reaction surface [32,137,184,189,197]?

Conclusion
Widely differing opinions exist even for the basic principles and interpretation of many folding experiments. Thus, our main conclusion is that although great progress has been made, there is a surprisingly large amount of work to be done.

References and recommended reading
Papers of particular interest, published within the period of review, have been highlighted as:

• of special interest

1. McCarney ER, Kohn JE, Plaxco KW: **Is there or isn't there? The case for (and against) residual structure in chemically denatured proteins**. *Crit Rev Biochem Mol Biol* 2005, **40**:181-189.

2. Meng W, Shan B, Tang Y, Raleigh DP: **Native like structure in the unfolded state of the villin headpiece helical subdomain, an ultrafast folding protein**. *Protein Sci* 2009, **18**:1692-1701.

3. Shan B, McClendon S, Rospigliosi C, Eliezer D, Raleigh DP: **The cold denatured state of the C-terminal domain of protein L9 is compact and contains both native and non-native structure**. *J Am Chem Soc* 2010, **132**:4669-4677.

4. Shan B, Eliezer D, Raleigh DP: **The unfolded state of the C-terminal domain of the ribosomal protein L9 contains both native and non-native structure**. *Biochemistry* 2009, **48**:4707-4719.

5. Bruun SW, Iesmantavicius V, Danielsson J, Poulsen FM: **Cooperative formation of native-like tertiary contacts in the ensemble of unfolded states of a four-helix protein**. *Proc Natl Acad Sci U S A* 2010, **107**:13306-13311.

6. Marsh JA, Forman-Kay JD: **Structure and disorder in an unfolded state under nondenaturing conditions from ensemble models consistent with a large number of experimental restraints**. *J Mol Biol* 2009, **391**:359-374.

7. Ratcliff K, Marqusee S: **Identification of residual structure in the unfolded state of ribonuclease H1 from the moderately thermophilic *Chlorobium tepidum*: comparison with thermophilic and mesophilic homologues**. *Biochemistry* 2010, **49**:5167-5175.

8. Robic S, Berger JM, Marqusee S: **Contributions of folding cores to the thermostabilities of two ribonucleases H**. *Protein Sci* 2002, **11**:381-389.

9. Rosner HI, Poulsen FM: **Residue-specific description of non-native transient structures in the ensemble of acid-denatured structures of the all-beta protein c-src SH3**. *Biochemistry* 2010, **49**:3246-3253.

10. Cliff MJ, Craven CJ, Marston JP, Hounslow AM, Clarke AR, Waltho JP: **The denatured state of N-PGK is compact and predominantly disordered**. *J Mol Biol* 2009, **385**:266-277.

11. Jha AK, Colubri A, Zaman MH, Koide S, Sosnick TR, Freed KF: **Helix, sheet, and polyproline II frequencies and strong nearest neighbor effects in a restricted coil library**. *Biochemistry* 2005, **44**:9691-9702.

12. Jha AK, Colubri A, Freed KF, Sosnick TR: **Statistical coil model of the unfolded state: resolving the reconciliation problem**. *Proc Natl Acad Sci U S A* 2005, **102**:13099-13104.

13. Sivaraman T, Arrington CB, Robertson AD: **Kinetics of unfolding and folding from amide hydrogen exchange in native ubiquitin**. *Nat Struct Biol* 2001, **8**:331-333.

14. Meisner WK, Sosnick TR: **Barrier-limited, microsecond folding of a stable protein measured with hydrogen exchange: implications for downhill folding**. *Proc Natl Acad Sci U S A* 2004, **101**:15639-15644.

15. Jackson SE, Fersht AR: **Folding of chymotrypsin inhibitor 2.1. Evidence for a two-state transition**. *Biochemistry* 1991, **30**:10428-10435.

16. Sanchez IE, Kiefhaber T: **Evidence for sequential barriers and obligatory intermediates in apparent two-state protein folding**. *J Mol Biol* 2003, **325**:367-376.

17. Gianni S, Brunori M, Jemth P, Oliveberg M, Zhang M: **Distinguishing between smooth and rough free energy barriers in protein folding**. *Biochemistry* 2009, **48**:11825-11830.

18. O'Brien EP, Brooks BR, Thirumalai D: **Molecular origin of constant *m*-values, denatured state collapse, and residue-dependent transition midpoints in globular proteins**. *Biochemistry* 2009, **48**:3743-3754.

19. O'Brien EP, Ziv G, Haran G, Brooks BR, Thirumalai D: **Effects of denaturants and osmolytes on proteins are accurately predicted by the molecular transfer model**. *Proc Natl Acad Sci U S A* 2008, **105**:13403-13408.

20. Canchi DR, Paschek D, Garcia AE: **Equilibrium study of protein denaturation by urea**. *J Am Chem Soc* 2010, **132**:2338-2344.

21. Kohn JE, Millett IS, Jacob J, Zagrovic B, Dillon TM, Cingel N, Dothager RS, Seifert S, Thiyagarajan P, Sosnick TR *et al.*: **Random-coil behavior and the dimensions of chemically unfolded proteins**. *Proc Natl Acad Sci U S A* 2004, **101**:12491-12496.

22. Merchant KA, Best RB, Louis JM, Gopich IV, Eaton WA: **Characterizing the unfolded states of proteins using single-molecule FRET spectroscopy and molecular simulations**. *Proc Natl Acad Sci U S A* 2007, **104**:1528-1533.

23. Chung HS, Louis JM, Eaton WA: **Experimental determination of upper bound for transition path times in protein folding from single-molecule photon-by-photon trajectories**. *Proc Natl Acad Sci U S A* 2009, **106**:11837-11844.

24. Nettels D, Gopich IV, Hoffmann A, Schuler B: **Ultrafast dynamics of protein collapse from single-molecule photon statistics**. *Proc Natl Acad Sci U S A* 2007, **104**:2655-2660.

25. Nettels D, Muller-Spath S, Kuster F, Hofmann H, Haenni D, Ruegger S, Reymond L, Hoffmann A, Kubelka J, Heinz B *et al.*: **Single-molecule spectroscopy of the temperature-induced collapse of unfolded proteins**. *Proc Natl Acad Sci U S A* 2009, **106**:20740-20745.

26. Sherman E, Haran G: **Coil–globule transition in the denatured state of a small protein**. *Proc Natl Acad Sci U S A* 2006, **103**:11539-11543.

27. Ziv G, Thirumalai D, Haran G: **Collapse transition in proteins**. *Phys Chem Chem Phys* 2009, **11**:83-93.

28. Ziv G, Haran G: **Protein folding, protein collapse, and Tanford's transfer model: lessons from single-molecule FRET**. *J Am Chem Soc* 2009, **131**:2942-2947.

29. Pugh SD, Gell C, Smith DA, Radford SE, Brockwell DJ: **Single-molecule studies of the Im7 folding landscape**. *J Mol Biol* 2010, **398**:132-145.

30. Magg C, Kubelka J, Holtermann G, Haas E, Schmid FX: **Specificity of the initial collapse in the folding of the cold shock protein**. *J Mol Biol* 2006, **360**:1067-1080.

31. Waldauer SA, Bakajin O, Lapidus LJ: **Extremely slow intramolecular diffusion in unfolded protein L**. *Proc Natl Acad Sci U S A* 2010, **107**:13713-13717.

32. Voelz VA, Singh VR, Wedemeyer WJ, Lapidus LJ, Pande VS: **Unfolded-state dynamics and structure of protein L characterized by simulation and experiment**. *J Am Chem Soc* 2010, **132**:4702-4709.

33. Dasgupta A, Udgaonkar JB: **Evidence for initial non-specific polypeptide chain collapse during the refolding of the SH3 domain of PI3 kinase**. *J Mol Biol* 2010, **403**:430-445.

34. Huang F, Lerner E, Sato S, Amir D, Haas E, Fersht AR: **Time-resolved fluorescence resonance energy transfer study shows a compact denatured state of the B domain of protein A**. *Biochemistry* 2009, **48**:3468-3476.

35. Kohn JE, Gillespie B, Plaxco KW: **Non-sequence-specific interactions can account for the compaction of proteins unfolded under ''native'' conditions**. *J Mol Biol* 2009, **394**:343-350.

36. Bowman GR, Pande VS: **Protein folded states are kinetic hubs**.
● *Proc Natl Acad Sci U S A* 2010, **107**:10890-10895.
The authors perform atomic-level folding simulations and translate multiple folding trajectories into a single reaction surface with clustered intermediates and defined interconversion rates using Markov models. Using these methods, a ''hub model'' is found that is non-funnel-like.

37. Plaxco KW, Millett IS, Segel DJ, Doniach S, Baker D: **Chain collapse can occur concomitantly with the rate-limiting step in protein folding**. *Nat Struct Biol* 1999, **6**:554-556.

38. Jacob J, Krantz B, Dothager RS, Thiyagarajan P, Sosnick TR: **Early collapse is not an obligate step in protein folding**. *J Mol Biol* 2004, **338**:369-382.

39. Svensson AK, Bilsel O, Kondrashkina E, Zitzewitz JA, Matthews CR: **Mapping the folding free energy surface for metal-free human Cu, Zn superoxide dismutase**. *J Mol Biol* 2006, **364**:1084-1102.

40. Kimura T, Akiyama S, Uzawa T, Ishimori K, Morishima I, Fujisawa T, Takahashi S: **Specifically collapsed intermediate in the early stage of the folding of ribonuclease A**. *J Mol Biol* 2005, **350**:349-362.

41. Jacob J, Dothager RS, Thiyagarajan P, Sosnick TR: **Fully reduced ribonuclease A does not expand at high denaturant concentration or temperature**. *J Mol Biol* 2007, **367**:609-615.

42. Wang Y, Trewhella J, Goldenberg DP: **Small-angle X-ray scattering of reduced ribonuclease A: effects of solution conditions and comparisons with a computational model of unfolded proteins**. *J Mol Biol* 2008, **377**:1576-1592.

43. Hoshino M, Hagihara Y, Hamada D, Kataoka M, Goto Y: **Trifluoroethanol-induced conformational transition of hen egg-white lysozyme studied by small-angle X-ray scattering**. *FEBS Lett* 1997, **416**:72-76.

44. Flanagan JM, Kataoka M, Fujisawa T, Engelman DM: **Mutations can cause large changes in the conformation of a denatured protein**. *Biochemistry* 1993, **32**:10359-10370.

45. Choy WY, Mulder FA, Crowhurst KA, Muhandiram DR, Millett IS, Doniach S, Forman-Kay JD, Kay LE: **Distribution of molecular size within an unfolded state ensemble using small-angle X-ray scattering and pulse field gradient NMR techniques**. *J Mol Biol* 2002, **316**:101-112.

46. Pollack L, Tate MW, Darnton NC, Knight JB, Gruner SM, Eaton WA, Austin RH: **Compactness of the denatured state of a fast-folding protein measured by submillisecond small-angle X-ray scattering**. *Proc Natl Acad Sci U S A* 1999, **96**:10115-10117.

47. Kimura T, Uzawa T, Ishimori K, Morishima I, Takahashi S, Konno T, Akiyama S, Fujisawa T: **Specific collapse followed by slow hydrogen-bond formation of beta-sheet in the folding of single-chain monellin**. *Proc Natl Acad Sci U S A* 2005, **102**:2748-2753.

48. Crick SL, Jayaraman M, Frieden C, Wetzel R, Pappu RV: **Fluorescence correlation spectroscopy shows that monomeric polyglutamine molecules form collapsed structures in aqueous solutions**. *Proc Natl Acad Sci U S A* 2006, **103**:16764-16769.

49. Mao AH, Crick SL, Vitalis A, Chicoine CL, Pappu RV: **Net charge**
•   **per residue modulates conformational ensembles of intrinsically disordered proteins**. *Proc Natl Acad Sci U S A* 2010, **107**:8183-8188.
Using an implicit solvent model designed to calculate structural features of intrinsically denatured proteins along with FCS measurements of hydrodynamics, the authors demonstrate that significant collapse can be found in polar and charged polyampholyte sequences.

50. Paz A, Zeev-Ben-Mordehai T, Lundqvist M, Sherman E, Mylonas E, Weiner L, Haran G, Svergun DI, Mulder FA, Sussman JL et al.: **Biophysical characterization of the unstructured cytoplasmic domain of the human neuronal adhesion protein neuroligin 3**. *Biophys J* 2008, **95**:1928-1944.

51. Muller-Spath S, Soranno A, Hirschfeld V, Hofmann H, Ruegger S, Reymond L, Nettels D, Schuler B: **Charge interactions can dominate the dimensions of intrinsically disordered proteins**. *Proc Natl Acad Sci U S A* 2010, **107**:14609-14614.

52. Tzul FO, Bowler BE: **Denatured states of low-complexity polypeptide sequences differ dramatically from those of foldable sequences**. *Proc Natl Acad Sci U S A* 2010, **107**:11364-11369.

53. Wu Y, Kondrashkina E, Kayatekin C, Matthews CR, Bilsel O: **Microsecond acquisition of heterogeneous structure in the folding of a TIM barrel protein**. *Proc Natl Acad Sci U S A* 2008, **105**:13367-13372.

54. Arai M, Kondrashkina E, Kayatekin C, Matthews CR, Iwakura M, Bilsel O: **Microsecond hydrophobic collapse in the folding of Escherichia coli dihydrofolate reductase, an alpha/beta-type protein**. *J Mol Biol* 2007, **368**:219-229.

55. Uzawa T, Kimura T, Ishimori K, Morishima I, Matsui T, Ikeda-Saito M, Takahashi S, Akiyama S, Fujisawa T: **Time-resolved small-angle X-ray scattering investigation of the folding dynamics of heme oxygenase: implication of the scaling relationship for the submillisecond intermediates of protein folding**. *J Mol Biol* 2006, **357**:997-1008.

56. Uzawa T, Akiyama S, Kimura T, Takahashi S, Ishimori K, Morishima I, Fujisawa T: **Collapse and search dynamics of apomyoglobin folding revealed by submillisecond observations of alpha-helical content and compactness**. *Proc Natl Acad Sci U S A* 2004, **101**:1171-1176.

57. Krantz BA, Moran LB, Kentsis A, Sosnick TR: **D/H amide kinetic isotope effects reveal when hydrogen bonds form during protein folding**. *Nat Struct Biol* 2000, **7**:62-71.

58. Krantz BA, Srivastava AK, Nauli S, Baker D, Sauer RT, Sosnick TR: **Understanding protein hydrogen bond formation with kinetic H/D amide isotope effects**. *Nat Struct Biol* 2002, **9**:458-463.

59. Matthews CR: **Effects of point mutations on the folding of globular proteins**. *Methods Enzymol* 1987, **154**:498-511.

60. Matouschek A, Kellis JT Jr, Serrano L, Bycroft M, Fersht AR: **Transient folding intermediates characterized by protein engineering**. *Nature* 1990, **346**:440-445.

61. Krantz BA, Sosnick TR: **Engineered metal binding sites map the heterogeneous folding landscape of a coiled coil**. *Nat Struct Biol* 2001, **8**:1042-1047.

62. 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**:463-475.

63. Sosnick TR, Dothager RS, Krantz BA: **Differences in the folding transition state of ubiquitin indicated by phi and psi analyses**. *Proc Natl Acad Sci U S A* 2004, **101**:17377-17382.

64. Sosnick TR, Krantz BA, Dothager RS, Baxa M: **Characterizing the protein folding transition state using psi analysis**. *Chem Rev* 2006, **106**:1862-1876.

65. Rouget JB, Schroer MA, Jeworrek C, Puhse M, Saldana JL, Bessin Y, Tolan M, Barrick D, Winter R, Royer CA: **Unique features of the folding landscape of a repeat protein revealed by pressure perturbation**. *Biophys J* 2010, **98**:2712-2721.

66. 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**:14108-14109.

67. Sosnick TR, Mayne L, Englander SW: **Molecular collapse: the rate-limiting step in two-state cytochrome c folding**. *Proteins* 1996, **24**:413-426.

68. Krantz BA, Mayne L, Rumbley J, Englander SW, Sosnick TR: **Fast and slow intermediate accumulation and the initial barrier mechanism in protein folding**. *J Mol Biol* 2002, **324**:359-371.

69. Wallin S, Chan HS: **Conformational entropic barriers in topology-dependent protein folding: perspectives from a simple native-centric polymer model**. *J Phys Condens Matter* 2006, **18**:S307-S328.

70. Bai Y, Zhou H, Zhou Y: **Critical nucleation size in the folding of small apparently two-state proteins**. *Protein Sci* 2004, **13**:1173-1181.

71. Pandit AD, Jha A, Freed KF, Sosnick TR: **Small proteins fold through transition states with native-like topologies**. *J Mol Biol* 2006, **361**:755-770.

72. Baxa M, Freed KF, Sosnick TR: **Quantifying the structural**
•   **requirements of the folding transition state of Protein A and other systems**. *J Mol Biol* 2008, **381**:1362-1381.
The authors present a comprehensive picture of the TSE for this highly-studied protein using a battery of methods. In addition, they propose that the TSEs for the proteins obeying the RCO correlation should acquired a similar level of native topology, $RCO^{TS} \approx 0.7 \cdot RCO^N$.

73. Meisner WK, Sosnick TR: **Fast folding of a helical protein initiated by the collision of unstructured chains**. *Proc Natl Acad Sci U S A* 2004, **101**:13478-13482.

74. Bosco G, Baxa M, Sosnick T: **Metal binding kinetics of bi-Histidine sites used in Psi-analysis: evidence for high energy protein folding intermediates**. *Biochemistry* 2009, **48**: 2950-2959.

75. Maity H, Maity M, Krishna MM, Mayne L, Englander SW: **Protein folding: the stepwise assembly of foldon units**. *Proc Natl Acad Sci U S A* 2005, **102**:4741-4746.

76. Sridevi K, Udgaonkar JB: **Unfolding rates of barstar determined in native and low denaturant conditions indicate the presence of intermediates**. *Biochemistry* 2002, **41**:1568-1578.

77. Silverman JA, Harbury PB: **The equilibrium unfolding pathway of a (beta/alpha)8 barrel**. *J Mol Biol* 2002, **324**:1031-1040.

78. Neudecker P, Lundstrom P, Kay LE: **Relaxation dispersion NMR spectroscopy as a tool for detailed studies of protein folding**. *Biophys J* 2009, **96**:2045-2054.

79. O'Connell NE, Grey MJ, Tang Y, Kosuri P, Miloushev VZ, Raleigh DP, Palmer AG 3rd: **Partially folded equilibrium intermediate of the villin headpiece HP67 defined by 13C relaxation dispersion**. *J Biomol NMR* 2009, **45**:85-98.

80. Korzhnev DM, Religa TL, Lundstrom P, Fersht AR, Kay LE: **The folding pathway of an FF domain: characterization of an on-pathway intermediate state under folding conditions by (15)N, (13)C(alpha) and (13)C-methyl relaxation dispersion and (1)H/(2)H-exchange NMR spectroscopy**. *J Mol Biol* 2007, **372**:497-512.

81. Cho JH, O'Connell N, Raleigh DP, Palmer AG 3rd: **Phi-value analysis for ultrafast folding proteins by NMR relaxation dispersion**. *J Am Chem Soc* 2010, **132**:450-451.

82. Bai Y, Englander SW: **Future directions in folding: the multi-state nature of protein structure**. *Proteins* 1996, **24**:145-151.

83. Chamberlain AK, Handel TM, Marqusee S: **Detection of rare partially folded molecules in equilibrium with the native conformation of RNaseH**. *Nat Struct Biol* 1996, **3**:782-787.

84. Feng H, Vu ND, Bai Y: **Detection and structure determination of an equilibrium unfolding intermediate of Rd-apocytochrome b562: native fold with non-native hydrophobic interactions**. *J Mol Biol* 2004, **343**:1477-1485.

85. Bedard S, Mayne LC, Peterson RW, Wand AJ, Englander SW: **The foldon substructure of staphylococcal nuclease**. *J Mol Biol* 2008, **376**:1142-1154.

86. Zheng Z, Sosnick TR: **Protein vivisection reveals elusive intermediates in folding**. *J Mol Biol* 2010, **397**:777-788.

87. Bai Y, Sosnick TR, Mayne L, Englander SW: **Protein folding intermediates: native-state hydrogen exchange**. *Science* 1995, **269**:192-197.

88. Krishna MM, Englander SW: **A unified mechanism for protein**
• **folding: predetermined pathways with optional errors**. *Protein Sci* 2007, **16**:449-464.
The authors argue that pathways previously characterized as parallel also can be described as singular when considering the order of native structure formation, while the observed kinetic diversity results from one or more optional ''error-correction'' barriers.

89. Lindberg MO, Oliveberg M: **Malleability of protein folding pathways: a simple reason for complex behaviour**. *Curr Opin Struct Biol* 2007, **17**:21-29.

90. Sosnick TR: **Kinetic barriers and the role of topology in protein and RNA folding**. *Prot Sci* 2008, **17**:1308-1318.

91. DeBartolo J, Colubri A, Jha AK, Fitzgerald JE, Freed KF, Sosnick TR: **Mimicking the folding pathway to improve homology-free protein structure prediction**. *Proc Natl Acad Sci U S A* 2009, **106**:3734-3739.

92. DeBartolo J, Hocky G, Wilde M, Xu J, Freed KF, Sosnick TR: **Protein structure prediction enhanced with evolutionary diversity: SPEED**. *Protein Sci* 2010, **19**:520-534.

93. 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**:260-288.

94. Fersht AR: **Nucleation mechanisms in protein folding**. *Curr Opin Struct Biol* 1997, **7**:3-9.

95. Brockwell DJ, Radford SE: **Intermediates: ubiquitous species on folding energy landscapes?** *Curr Opin Struct Biol* 2007, **17**:30-37.

96. Vallee-Belisle A, Michnick SW: **Multiple tryptophan probes reveal that ubiquitin folds via a late misfolded intermediate**. *J Mol Biol* 2007, **374**:791-805.

97. Rea AM, Simpson ER, Crespo MD, Searle MS: **Helix mutations stabilize a late productive intermediate on the folding pathway of ubiquitin**. *Biochemistry* 2008, **47**:8225-8236.

98. Whittaker SB, Spence GR, Gunter Grossmann J, Radford SE, Moore GR: **NMR analysis of the conformational properties of the trapped on-pathway folding intermediate of the bacterial immunity protein Im7**. *J Mol Biol* 2007, **366**:1001-1015.

99. Connell KB, Horner GA, Marqusee S: **A single mutation at residue 25 populates the folding intermediate of *E. coli* RNase H and reveals a highly dynamic partially folded ensemble**. *J Mol Biol* 2009, **391**:461-470.

100. Vallurupalli P, Hansen DF, Kay LE: **Structures of invisible, excited protein states by relaxation dispersion NMR spectroscopy**. *Proc Natl Acad Sci U S A* 2008, **105**:11766-11771.

101. Schanda P, Brutscher B, Konrat R, Tollinger M: **Folding of the KIX domain: characterization of the equilibrium analog of a folding intermediate using ^{15}N/^{13}C relaxation dispersion and fast ^{1}H/^{2}H amide exchange NMR spectroscopy**. *J Mol Biol* 2008, **380**:726-741.

102. Korzhnev DM, Religa TL, Banachewicz W, Fersht AR, Kay LE: **A**
• **transient and low-populated protein-folding intermediate at atomic resolution**. *Science* 2010, **329**:1312-1316.
Using a combination of sophisticated NMR and modeling methods, the authors present an atomic level model of a folding intermediate containing kinetically significant non-native interactions.

103. Baldwin RL, Frieden C, Rose GD: **Dry molten globule**
• **intermediates and the mechanism of protein unfolding**. *Proteins* 2010, **78**:2725-2737.
The authors discuss the evidence for, and implications of a dry molten globule model for folding.

104. Cheung MS, Garcia AE, Onuchic JN: **Protein folding mediated by solvation: water expulsion and formation of the hydrophobic core occur after the structural collapse**. *Proc Natl Acad Sci U S A* 2002, **99**:685-690.

105. Garcia AE, Onuchic JN: **Folding a protein in a computer: an atomic description of the folding/unfolding of protein A**. *Proc Natl Acad Sci U S A* 2003, **100**:13898-13903.

106. Fernandez-Escamilla AM, Cheung MS, Vega MC, Wilmanns M, Onuchic JN, Serrano L: **Solvation in protein folding analysis: combination of theoretical and experimental approaches**. *Proc Natl Acad Sci U S A* 2004, **101**:2834-2839.

107. Shea JE, Onuchic JN, Brooks CL 3rd: **Probing the folding free energy landscape of the Src-SH3 protein domain**. *Proc Natl Acad Sci U S A* 2002, **99**:16064-16068.

108. Jha SK, Udgaonkar JB: **Direct evidence for a dry molten globule intermediate during the unfolding of a small protein**. *Proc Natl Acad Sci U S A* 2009, **106**:12289-12294.

109. Reiner A, Henklein P, Kiefhaber T: **An unlocking/relocking barrier in conformational fluctuations of villin headpiece subdomain**. *Proc Natl Acad Sci U S A* 2010, **107**:4955-4960.

110. Kiefhaber T, Labhardt AM, Baldwin RL: **Direct NMR evidence for an intermediate preceding the rate-limiting step in the unfolding of ribonuclease A**. *Nature* 1995, **375**:513-515.

111. Hoeltzli SD, Frieden C: **Real-time refolding studies of 6-^{19}F-tryptophan labeled *Escherichia coli* dihydrofolate reductase using stopped-flow NMR spectroscopy**. *Biochemistry* 1996, **35**:16843-16851.

112. Wolynes PG: **Folding funnels and energy landscapes of larger proteins within the capillarity approximation**. *Proc Natl Acad Sci U S A* 1997, **94**:6170-6175.

113. Bartlett AI, Radford SE: **Desolvation and development of specific hydrophobic core packing during Im7 folding**. *J Mol Biol* 2010, **396**:1329-1345.

114. Zhou Z, Feng H, Ghirlando R, Bai Y: **The high-resolution NMR structure of the early folding intermediate of the *Thermus thermophilus* ribonuclease H**. *J Mol Biol* 2008, **384**:531-539.

115. Kato H, Feng H, Bai Y: **The folding pathway of T4 lysozyme: the high-resolution structure and folding of a hidden intermediate**. *J Mol Biol* 2007, **365**:870-880.

116. Wang T, Zhou Z, Bunagan MR, Du D, Bai Y, Gai F: **Probing the folding intermediate of Rd-apocyt b562 by protein engineering and infrared T-jump**. *Protein Sci* 2007, **16**:1176-1183.

117. Hills RD Jr, Kathuria SV, Wallace LA, Day IJ, Brooks CL 3rd, Matthews CR: **Topological frustration in beta alpha-repeat proteins: sequence diversity modulates the conserved folding mechanisms of alpha/beta/alpha sandwich proteins**. *J Mol Biol* 2010, **398**:332-350.

118. Hills RD Jr, Brooks CL 3rd: **Subdomain competition, cooperativity, and topological frustration in the folding of CheY**. *J Mol Biol* 2008, **382**:485-495.

119. Zhang Z, Chan HS: **Competition between native topology and nonnative interactions in simple and complex folding kinetics of natural and designed proteins**. *Proc Natl Acad Sci U S A* 2010, **107**:2920-2925.

120. Watters AL, Deka P, Corrent C, Callender D, Varani G, Sosnick T, Baker D: **The highly cooperative folding of small naturally occurring proteins is likely the result of natural selection**. *Cell* 2007, **128**:613-624.

121. Zarrine-Afsar A, Wallin S, Neculai AM, Neudecker P, Howell PL, Davidson AR, Chan HS: **Theoretical and experimental demonstration of the importance of specific nonnative interactions in protein folding**. *Proc Natl Acad Sci U S A* 2008, **105**:9999-10004.

122. Nishimura C, Dyson HJ, Wright PE: **Energetic frustration of apomyoglobin folding: role of the B helix**. *J Mol Biol* 2010, **396**:1319-1328.

123. Nishimura C, Dyson HJ, Wright PE: **Identification of native and non-native structure in kinetic folding intermediates of apomyoglobin**. *J Mol Biol* 2006, **355**:139-156.

124. Wensley BG, Batey S, Bone FA, Chan ZM, Tumelty NR, Steward A,
 • Kwa LG, Borgia A, Clarke J: **Experimental evidence for a frustrated energy landscape in a three-helix-bundle protein family**. *Nature* 2010, **463**:685-688.
The authors observe that two members of the family fold slowly due to increase friction related to helix misdocking even though folding remains kinetically two-state. Hence, a non-native interaction formed in the unstable species can slow folding, which implies the species is an obligate intermediate.

125. Capaldi AP, Kleanthous C, Radford SE: **Im7 folding mechanism: misfolding on a path to the native state**. *Nat Struct Biol* 2002, **9**:209-216.

126. Connell KB, Miller EJ, Marqusee S: **The folding trajectory of RNase H is dominated by its topology and not local stability: a protein engineering study of variants that fold via two-state and three-state mechanisms**. *J Mol Biol* 2009, **391**:450-460.

127. McCully ME, Beck DA, Fersht AR, Daggett V: **Refolding the engrailed homeodomain: structural basis for the accumulation of a folding intermediate**. *Biophys J* 2010, **99**:1628-1636.

128. Sosnick TR, Mayne L, Hiller R, Englander SW: **The barriers in protein folding**. *Nat Struct Biol* 1994, **1**:149-156.

129. Sosnick TR, Mayne L, Hiller R, Englander SW: **The barriers in protein folding**. In *Peptide and Protein Folding Workshop.* Edited by DeGrado WF. Philadelphia, PA: International Business Communications; 1995:52-80.

130. Guo ZY, Thirumalai D: **Kinetics of protein-folding: nucleation mechanism, time scales, and pathways**. *Biopolymers* 1995, **36**:83-102.

131. Abkevich VI, Gutin AM, Shakhnovich EI: **Specific nucleus as the transition state for protein folding: evidence from the lattice model**. *Biochemistry* 1994, **33**:10026-10036.

132. Plaxco KW, Simons KT, Baker D: **Contact order, transition state placement and the refolding rates of single domain proteins**. *J Mol Biol* 1998, **277**:985-994.

133. 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**:11177-11183.

134. 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**:1149-1154.

135. Chavez LL, Onuchic JN, Clementi C: **Quantifying the roughness on the free energy landscape: entropic bottlenecks and protein folding rates**. *J Am Chem Soc* 2004, **126**:8426-8432.

136. Makarov DE, Plaxco KW: **The topomer search model: a simple, quantitative theory of two-state protein folding kinetics**. *Protein Sci* 2003, **12**:17-26.

137. Ferguson A, Liu Z, Chan HS: **Desolvation barrier effects are a likely contributor to the remarkable diversity in the folding rates of small proteins**. *J Mol Biol* 2009, **389**:619-636.

138. Portman JJ: **Cooperativity and protein folding rates**. *Curr Opin Struct Biol* 2010, **20**:11-15.

139. Munoz V, Eaton WA: **A simple model for calculating the kinetics of protein folding from three-dimensional structures**. *Proc Natl Acad Sci U S A* 1999, **96**:11311-11316.

140. Pandit AD, Krantz BA, Dothager RS, Sosnick TR: **Characterizing protein folding transition states using Psi-analysis**. *Methods Mol Biol* 2007, **350**:83-104.

141. 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**:495-500.

142. Bulaj G, Goldenberg DP: **Phi-values for BPTI folding intermediates and implications for transition state analysis**. *Nat Struct Biol* 2001, **8**:326-330.

143. Baxa MC, Freed KF, Sosnick TR: **Psi-constrained simulations of protein folding transition states: implications for calculating**. *J Mol Biol* 2009, **386**:920-928.

144. Neudecker P, Zarrine-Afsar A, Choy WY, Muhandiram DR, Davidson AR, Kay LE: **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* 2006, **363**:958-976.

145. Feng H, Vu ND, Zhou Z, Bai Y: **Structural examination of Phi-value analysis in protein folding**. *Biochemistry* 2004, **43**:14325-14331.

146. Went HM, Jackson SE: **Ubiquitin folds through a highly polarized transition state**. *Protein Eng Des Sel* 2005, **18**:229-237.

147. 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**:714-720.

148. Guo W, Lampoudi S, Shea JE: **Temperature dependence of the free energy landscape of the src-SH3 protein domain**. *Proteins* 2004, **55**:395-406.

149. Yi Q, Rajagopal P, Klevit RE, Baker D: **Structural and kinetic characterization of the simplified SH3 domain FP1**. *Protein Sci* 2003, **12**:776-783.

150. Garcia-Mira MM, Boehringer D, Schmid FX: **The folding transition state of the cold shock protein is strongly polarized**. *J Mol Biol* 2004, **339**:555-569.

151. Lindberg M, Tangrot J, Oliveberg M: **Complete change of the protein folding transition state upon circular permutation**. *Nat Struct Biol* 2002, **9**:818-822.

152. Riddle DS, Grantcharova VP, Santiago JV, Alm E, Ruczinski II, Baker D: **Experiment and theory highlight role of native state topology in SH3 folding**. *Nat Struct Biol* 1999, **6**:1016-1024.

153. Weikl TR, Dill KA: **Transition-states in protein folding kinetics: the structural interpretation of Phi values**. *J Mol Biol* 2007, **365**:1578-1586.

154. Naganathan AN, Munoz V: **Insights into protein folding mechanisms from large scale analysis of mutational effects**. *Proc Natl Acad Sci U S A* 2010, **107**:8611-8616.

155. Fersht AR, Itzhaki LS, elMasry NF, Matthews JM, Otzen DE: **Single versus parallel pathways of protein folding and fractional formation of structure in the transition state**. *Proc Natl Acad Sci U S A* 1994, **91**:10426-10429.

156. Viguera AR, Serrano L: **Loop length, intramolecular diffusion and protein folding**. *Nat Struct Biol* 1997, **4**:939-946.

157. Wright CF, Lindorff-Larsen K, Randles LG, Clarke J: **Parallel protein-unfolding pathways revealed and mapped**. *Nat Struct Biol* 2003, **10**:658-662.

158. Nauli S, Kuhlman B, Baker D: **Computer-based redesign of a protein folding pathway**. *Nat Struct Biol* 2001, **8**:602-605.

159. Klimov DK, Thirumalai D: **Symmetric connectivity of secondary structure elements enhances the diversity of folding pathways**. *J Mol Biol* 2005, **353**:1171-1186.

160. Shen T, Hofmann CP, Oliveberg M, Wolynes PG: **Scanning malleable transition state ensembles: comparing theory and experiment for folding protein U1A**. *Biochemistry* 2005, **44**:6433-6439.

161. Olofsson M, Hansson S, Hedberg L, Logan DT, Oliveberg M: **Folding of S6 structures with divergent amino acid composition: pathway flexibility within partly overlapping foldons**. *J Mol Biol* 2007, **365**:237-248.

162. Moran LB, Schneider JP, Kentsis A, Reddy GA, Sosnick TR: **Transition state heterogeneity in GCN4 coiled coil folding studied by using multisite mutations and crosslinking**. *Proc Natl Acad Sci U S A* 1999, **96**:10699-10704.

163. Tripp KW, Barrick D: **Rerouting the folding pathway of the Notch ankyrin domain by reshaping the energy landscape**. *J Am Chem Soc* 2008, **130**:5681-5688.

164. Werbeck ND, Rowling PJ, Chellamuthu VR, Itzhaki LS: **Shifting transition states in the unfolding of a large ankyrin repeat protein**. *Proc Natl Acad Sci U S A* 2008, **105**:9982-9987.

165. Krishna MM, Maity H, Rumbley JN, Englander SW: **Branching in the sequential folding pathway of cytochrome c**. *Protein Sci* 2007, **16**:1946-1956.

166. Krishna MM, Maity H, Rumbley JN, Lin Y, Englander SW: **Order of steps in the cytochrome C folding pathway: evidence for a sequential stabilization mechanism**. *J Mol Biol* 2006, **359**:1410-1419.

167. Onuchic JN, Luthey-Schulten Z, Wolynes PG: **Theory of protein folding: the energy landscape perspective**. *Annu Rev Phys Chem* 1997, **48**:545-600.

168. Dill KA, Chan HS: **From Levinthal to pathways to funnels**. *Nat Struct Biol* 1997, **4**:19.

169. Udgaonkar JB: **Multiple routes and structural heterogeneity in protein folding**. *Annu Rev Biophys* 2008, **37**:489-510.

170. Weinkam P, Zimmermann J, Romesberg FE, Wolynes PG: **The folding energy landscape and free energy excitations of cytochrome c**. *Acc Chem Res* 2010, **43**:652-660.

171. Weinkam P, Zong C, Wolynes PG: **A funneled energy landscape**
• **for cytochrome c directly predicts the sequential folding route inferred from hydrogen exchange experiments**. *Proc Natl Acad Sci U S A* 2005, **102**:12401-12406.
The authors show that native-centric energy function with enhanced cooperativity along with a sufficiently detailed model can reproduce experimentally observed ''foldon'' behavior in cytochrome c.

172. Liu Z, Chan HS: **Desolvation is a likely origin of robust enthalpic barriers to protein folding**. *J Mol Biol* 2005, **349**:872-889.

173. Kaya H, Chan HS: **Simple two-state protein folding kinetics requires near-levinthal thermodynamic cooperativity**. *Proteins* 2003, **52**:510-523.

174. Jewett AI, Pande VS, Plaxco KW: **Cooperativity, smooth energy landscapes and the origins of topology-dependent protein folding rates**. *J Mol Biol* 2003, **326**:247-253.

175. Garbuzynskiy SO, Finkelstein AV, Galzitskaya OV: **On the prediction of folding nuclei in globular proteins**. *Mol Biol* 2005, **39**:906-914.

176. Voelz VA, Bowman GR, Beauchamp K, Pande VS: **Molecular simulation of ab initio protein folding for a millisecond folder NTL9(1-39)**. *J Am Chem Soc* 2010, **132**:1526-1528.

177. Freddolino PL, Park S, Roux B, Schulten K: **Force field bias in protein folding simulations**. *Biophys J* 2009, **96**:3772-3780.

178. Day R, Paschek D, Garcia AE: **Microsecond simulations of the folding/unfolding thermodynamics of the Trp-cage miniprotein**. *Proteins* 2010, **78**:1889-1899.

179. Mittal J, Best RB: **Tackling force-field bias in protein folding simulations: folding of Villin HP35 and Pin WW domains in explicit water**. *Biophys J* 2010, **99**:L26-28.

180. Bowman GR, Beauchamp KA, Boxer G, Pande VS: **Progress and challenges in the automated construction of Markov state models for full protein systems**. *J Chem Phys* 2009, **131**:124101.

181. Noe F, Schutte C, Vanden-Eijnden E, Reich L, Weikl TR:
• **Constructing the equilibrium ensemble of folding pathways from short off-equilibrium simulations**. *Proc Natl Acad Sci U S A* 2009, **106**:19011-19016.
The authors perform atomic-level folding simulations and translate multiple folding trajectories into a single reaction surface with clustered intermediates and defined interconversion rates using Markov models.

182. Neuweiler H, Johnson CM, Fersht AR: **Direct observation of ultrafast folding and denatured state dynamics in single protein molecules**. *Proc Natl Acad Sci U S A* 2009, **106**:18569-18574.

183. Klimov DK, Thirumalai D: **Viscosity dependence of the folding rates of proteins**. *Phys Rev Lett* 1997, **79**:317-320.

184. Rhee YM, Pande VS: **Solvent viscosity dependence of the protein folding dynamics**. *J Phys Chem B* 2008, **112**:6221-6227.

185. Hagen SJ: **Solvent viscosity and friction in protein folding dynamics**. *Curr Protein Pept Sci* 2010, **11**:385-395.

186. Plaxco KW, Baker D: **Limited internal friction in the rate-limiting step of a two-state protein folding reaction**. *Proc Natl Acad Sci U S A* 1998, **95**:13591-13596.

187. Bhattacharyya RP, Sosnick TR: **Viscosity dependence of the folding kinetics of a dimeric and monomeric coiled coil**. *Biochemistry* 1999, **38**:2601-2609.

188. Jacob M, Schindler T, Balbach J, Schmid FX: **Diffusion control in an elementary protein folding reaction**. *Proc Natl Acad Sci U S A* 1997, **94**:5622-5627.

189. Cellmer T, Henry ER, Hofrichter J, Eaton WA: **Measuring internal friction of an ultrafast-folding protein**. *Proc Natl Acad Sci U S A* 2008, **105**:18320-18325.

190. Pabit SA, Roder H, Hagen SJ: **Internal friction controls the speed of protein folding from a compact configuration**. *Biochemistry* 2004, **43**:12532-12538.

191. Narayanan R, Pelakh L, Hagen SJ: **Solvent friction changes the folding pathway of the tryptophan zipper TZ2**. *J Mol Biol* 2009, **390**:538-546.

192. DeCamp SJ, Naganathan AN, Waldauer SA, Bakajin O, Lapidus LJ: **Direct observation of downhill folding of lambda-repressor in a microfluidic mixer**. *Biophys J* 2009, **97**:1772-1777.

193. Huang F, Ying L, Fersht AR: **Direct observation of barrier-limited folding of BBL by single-molecule fluorescence resonance energy transfer**. *Proc Natl Acad Sci U S A* 2009, **106**:16239-16244.

194. Li P, Oliva FY, Naganathan AN, Munoz V: **Dynamics of one-state downhill protein folding**. *Proc Natl Acad Sci U S A* 2009, **106**:103-108.

195. Liu F, Gao YG, Gruebele M: **A survey of lambda repressor fragments from two-state to downhill folding**. *J Mol Biol* 2010, **397**:789-798.

196. Chung HS, Tokmakoff A: **Temperature-dependent downhill unfolding of ubiquitin. I. Nanosecond-to-millisecond resolved nonlinear infrared spectroscopy**. *Proteins* 2008, **72**:474-487.

197. Best RB, Hummer G: **Coordinate-dependent diffusion in protein folding**. *Proc Natl Acad Sci U S A* 2010, **107**:1088-1093.