Early events in protein folding

Kalyan K. Sinha and Jayant B. Udgaonkar*

National Centre for Biological Sciences, Tata Institute of Fundamental Research, Bangalore 560 065, India

Many proteins take at least a few seconds to fold, but almost all proteins undergo major structural transitions within the first millisecond (ms) of folding. Understanding the nature of the product of the first ms of folding is important because it sets the stage for the major folding reaction that follows. The past decade has seen major advances in methodologies that have enabled temporal and structural resolution of events happening in the first ms of folding. An important very early event appears to be the collapse of the polypeptide chain to form a compact globule. A specific structure also appears to form within the first ms, and it appears for several proteins that this happens only after the initial chain collapse reaction. Hence, when studied at the first ms of folding, the compact globule appears to be a specific folding intermediate. The accumulated kinetic data suggest that structure formation in the first ms may be highly non-cooperative and may occur in many steps. Multiple folding routes appear to be available, and the nature and extent of structure formation in the first ms may depend on the dominant route utilized under a particular folding condition. There is now evidence suggesting that the energy barrier encountered by the collapsing polypeptide chain can be as small as ~k_BT, bringing out the possibility that initial chain collapse and structure formation may even be gradual transitions. Understanding how such continuous transitions can still lead to the development of a specific structure during sub-ms folding reactions poses a difficult experimental challenge.

Keywords: Cooperativity, intermediate, polypeptide chain collapse, protein folding, specificity.

FORM begets function in life. At the molecular level, the information required to sustain life is stored only in a one-dimensional form, namely the DNA sequence. This information becomes of potential use only when it is transferred first to a RNA sequence and then to a protein sequence. Precision in these steps is ensured by the use of templates. The final productive step in information transfer is from the one-dimensional protein sequence to a precise three-dimensional protein structure, which confers a specific function to the protein. In the final step, it is the unique amino acid sequence of a protein that appears to serve as a self-guiding template for folding to the unique structure. Understanding the development of a significant form as a protein folds, has remained one of the fundamental problems of modern biology.

The problem really is to determine how an unfolded polypeptide chain searches out its final native conformation from an inconceivably large number of available conformations. A polypeptide chain of 101 amino acid residues would have to sample $3^{100} = 5 \times 10^{47}$ conformations, if each bond connecting two consecutive residues has only three possible configurations. If the sampling takes place at a rate equal to that of bond vibrations, i.e. 10^{13} s^{-1}, then it would take 10^{27} years for an unfolded polypeptide chain to complete the search for its native conformation[1]. The discrepancy between this large time estimate and the real folding times of proteins, which are in the seconds timescale or faster, is commonly referred to as the Levinthal paradox[2]. Although not all possible conformations are accessible to a polypeptide chain[3], and hence not sampled, it is commonly assumed that the solution to the Levinthal paradox lies in the protein using specific pathways to fold. A folding pathway defines a particular sequence of structural events, and understanding this sequence of events has been a long-standing challenge for experimental biochemists.

The most poorly understood aspects of protein folding reactions are the initial sub-millisecond (ms) folding reactions which result in partial[4] or, in a few cases, complete[5] folding of proteins. Sub-ms folding reactions that lead to partial folding of proteins, involve a fast collapse of the unfolded polypeptide chain. In this review, the current status of knowledge about sub-ms folding events is presented, with emphasis on the initial collapse reaction. Its cooperativity and specificity are discussed. Experimental methodologies are described briefly. The contributions of structure in the unfolded state, as well as of the elementary events that start-off structure formation, in the initiation of folding reactions are also discussed.

How do proteins fold?

Experimental studies of protein-folding mechanisms have been steered by several conceptual models[6] (Figure 1). The framework model suggests that during the initial stages of protein folding, local interactions dominate and guide the formation of secondary structural elements[7–9]. This is followed by random diffusion–collision of these local elements of the secondary structure until stable native tertiary contacts are made[10–12]. The hydrophobic collapse

*For correspondence. (e-mail: jayant@ncbs.res.in)

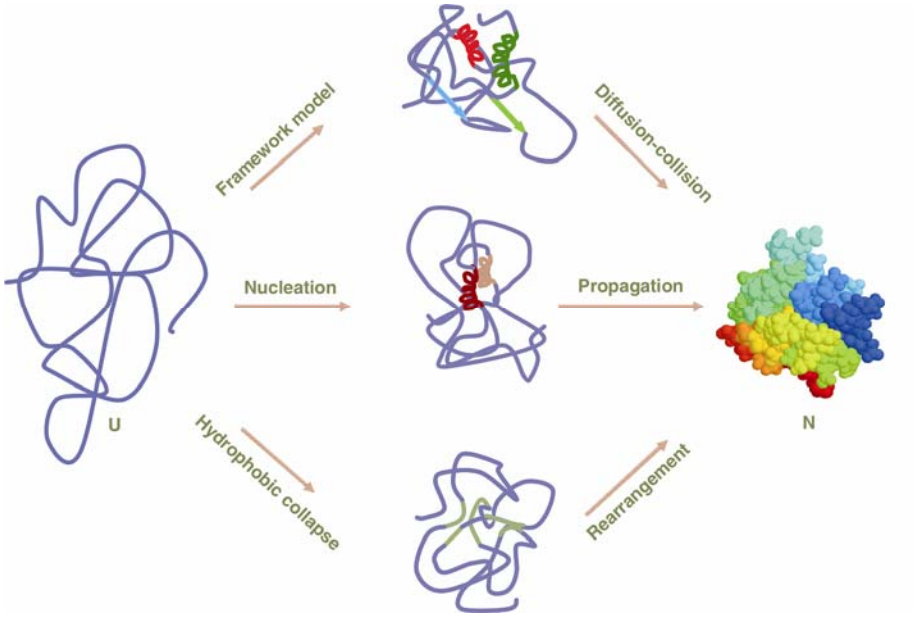

Figure 1. Models of protein folding. In the nucleation model, a structure develops around a local nucleus of the secondary structure. In the framework model, the secondary structure forms before the development of any tertiary contacts. In the hydrophobic collapse model, the folding reaction begins with a collapse of the unfolded polypeptide chain which precedes the formation of any structure.

mechanism postulates that the folding of a protein begins with an entropically driven clustering of the hydrophobic amino acid residues[13–15]. The formation of a collapsed intermediate with restricted conformational space facilitates the formation of the secondary structure and consolidation of tertiary contacts. The nucleation model postulates the formation of a local nucleus of the secondary structure by a few key residues in the polypeptide chain. The rest of the structure then propagates around the nucleus without encountering any energy barrier[16]. Thus, in this model, the formation of a nucleation site is the rate-limiting step. An extension of the nucleation model, the nucleation–condensation model, envisages a more diffuse nucleus, and that the collapse of the polypeptide chain occurs in parallel with structure formation[17,18]. This model was proposed to describe the mechanism of folding of proteins that appear to fold by a 'two-state' $N \leftrightarrow U$ mechanism, in which all physical interactions appear to develop in a concerted manner[19].

A more recent view of the protein folding mechanism is the energy landscape model[20,21]. According to this model, all folding protein molecules are guided by an energy bias to traverse an energy landscape towards the native conformation. The concomitant decrease in conformational entropy leads to a funnel-shaped energy landscape (Figure 2). Many different folding trajectories for individual protein molecules are envisaged and hence, multiple folding pathways are expected to be operative. Intermediates, when present, are considered as kinetic traps which slow down the folding reaction. This view also predicts the existence of downhill folding scenarios in which the energy barrier to folding disappears and the folding reaction occurs at its speed limit.

It is useful to examine the different models of protein folding in the context of how proteins begin their search for the native conformation. The nucleation model appears to be inapplicable to folding reactions, because it does not predict the early intermediate forms seen during the folding of many proteins[8]. Secondary structural elements do not appear to form unless some stabilizing tertiary contacts are made. Hence, a framework model is also unlikely to be a common mechanism by which proteins fold[22]. By contrast, the observation that a fast (sub-ms) collapse reaction precedes the formation of a secondary structure during the folding reactions of several proteins[4,23–28], suggests that many proteins indeed fold by the hydrophobic collapse mechanism. Nonetheless, the simultaneous occurrence of collapse and structure formation in the case of a few apparently two-state folding proteins[19,29] is difficult to explain by the classical hydrophobic collapse mechanism. For such proteins, the nucleation–condensation mechanism may better describe how folding occurs[17,18].

Protein folding: timescales and barriers

Proteins fold on a timescale that ranges from a few microseconds (μs) to several hundred seconds. It is difficult to explain billion-fold differences in the folding times seen with different proteins. According to transition state theory, which is commonly applied for analysing protein folding kinetics, a dominant free-energy barrier describ-

able on a single reaction coordinate, slows down the folding reaction from a folding speed limit[30,31]. The differences seen in the folding rates of different proteins can be explained by different barrier heights encountered by different folding polypeptide chains. But there is no direct way to calculate the absolute size of this free-energy barrier because a reliable value of the pre-exponential factor is not known for protein folding reactions.

In the presence of a sizeable barrier, the folding reaction is expected to be cooperative, which implies that the detection of intermediate conformations, which a folding polypeptide chain adopts en route, would be extremely difficult. Several proteins show single exponential folding kinetics consistent with a cooperative barrier-crossing reaction[19]. On the other hand, the folding kinetics of many other proteins is multi-exponential[32–34], which implies that intermediates populate the folding pathways of these proteins. There are several known examples in which a folding reaction appears non-cooperative when studied using either multiple spectroscopic probes[35,36] or site-specific probes[37–40]. Thus, the folding reactions of proteins appear to be far too complex to be understood completely in terms of a simple first-order, small-molecule chemical reaction like the barrier-crossing mechanism[41]. Nevertheless, in the absence of easily applicable alternative models, transition state theory is commonly used to describe protein-folding kinetics.

A major recent thrust of protein-folding research has been the identification of downhill folding scenarios predicted by the energy landscape view of protein folding. While the classical transition-state folding scenario involves only two co-existing and inter-converting states separated by a dominant free-energy barrier, a barrier-less folding scenario is expected to involve multiple steps with small distributed barriers instead of a single dominant activation barrier (Figure 3)[20]. The downhill folding mechanism allows a detailed description of different steps involved in the folding reaction of a protein[20,38,42,43]. On the basis of experimental observations, a folding speed limit of $N/100$ μs has been suggested for the folding reaction of a protein of N residues[31]. Many recent studies have examined ultra-fast folding reactions, which result in complete or partial folding of different proteins within a few μs. The height of the activation barrier for such fast-folding reactions is expected[44,45] to approach k_BT, a regime where diffusive conformational search is expected to dominate the folding reaction, because the barrier is readily crossed[46]. It is not clear how suitable transition state theory is for describing a process with a marginal activation energy barrier, because the species at the barrier top might populate to an extent greater than that envisaged for a first-order chemical reaction[43,44].

It is important to study protein-folding reactions which occur on a timescale between the molecular speed limit of folding and that of a folding reaction involving a dominant activation barrier (> 3 k_BT). The folding reactions of several ultra-fast folders, and the sub-ms folding reactions that result in the partial folding of many non two-state folding proteins, are examples of such reactions.

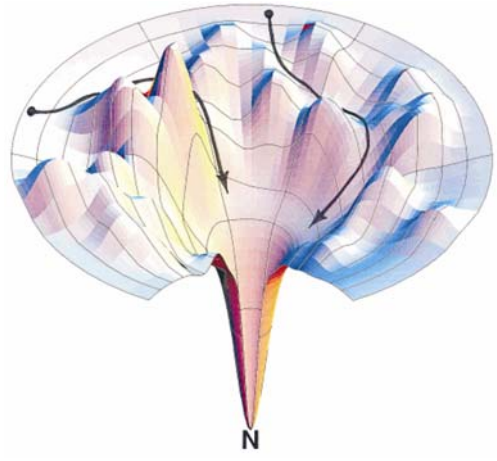

Figure 2. The energy landscape view of protein folding. A protein can find its native conformation via multiple routes. A feature common to all the folding trajectories in the landscape view is a reduction in conformational entropy as the folding molecules travel along the downward slope of the landscape. Different routes may differ in the degree of ruggedness they present to a polypeptide chain en route to folding. From Chan and Dill[221] with permission.

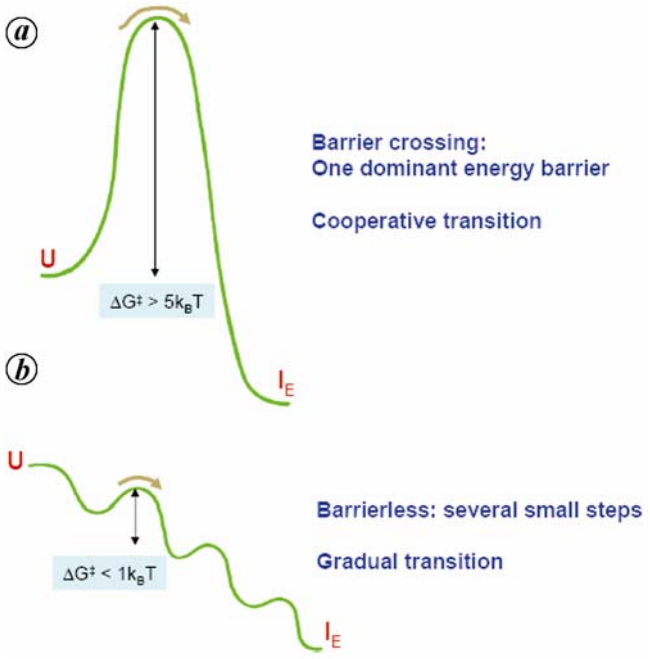

Figure 3. Energy surface for (***a***) two-state and (***b***) downhill protein folding scenarios. In the presence of a dominant free energy barrier (>5 k_BT) separating two species (here shown as the completely unfolded form, U and an early folding intermediate, I_E), the folding transition appears to be cooperative because the intermediates do not populate. In a downhill folding scenario, there is no significant barrier between U and I_E. In such a case, the transition between U and I_E is expected to be a gradual one with multiple steps, each involving small free enerrgy barriers (~k_BT).

Ultra-fast folders fold completely in less than a millisecond

Several proteins can fold completely in a few μs to several hundred μs[5,31,45,47–50], presumably because their free-energy landscapes are smoother. Some of the ultra-fast folders show non-exponential folding kinetics, which are usually taken as a signature of barrier-less folding[51–53]. It should, however, be noted that downhill folding reactions may also show exponential kinetics, depending on the roughness of the free energy surface[54,55]. Some mutant variants of the λ-repressor fragment, $\lambda_{6\text{-}85}$, fold on a timescale close to the protein folding speed limit[44,53,56]. By tuning the protein stability through mutations, solvent or temperature, the kinetics of a downhill folding reaction can be switched from exponential to non-exponential and vice versa[54]. In a NMR-monitored, atom-by-atom analysis of the equilibrium unfolding transition of Naf-BBL, a protein that folds in the sub-ms time domain, the equilibrium unfolding transition of the protein appeared to be highly non-cooperative as to be a downhill transition[38,57].

Surprisingly, the folding reactions of many of the other ultra-fast folding proteins appear to be 'two-state', and not barrier-less. These proteins show exponential folding kinetics, their folding rates are temperature-dependent, and the equilibrium unfolding reactions appear cooperative[45,58]. Thus, it appears from the available evidence that folding at or near the speed limit may not necessary mean a barrier-less folding transition. However, the size of the free-energy barrier encountered during the microsecond folding of an ultra-fast folder is usually close to k_BT, which would mean that high-energy intermediate conformations are likely to be more populated than in typical folding reactions (energy barrier ~5–10 k_BT)[43]. In a recent experiment with a 35-residue subdomain from the villin headpiece, it was seen that the ultra-fast folding rate constant was independent of the concentration of the chemical denaturant[59]. This observation was explained on the basis of a large movement of a small free-energy barrier (~k_BT)[45,59,60], even though the ultra-fast folding reactions of other proteins do not show this kind of movement of the folding transition state. An equally probable explanation would be that the folding reaction of this protein proceeds in the absence of any free-energy barrier.

Elementary events in protein folding reactions

Since ultra-fast folding proteins can fold completely in <1 ms, it is obvious that elementary folding events such as the formation of loops, turns, α-helices and β-structures, must occur on a shorter or comparable timescale. Several reports have presented estimates of the rate of intramolecular loop formation in short peptides. By laser-flash triplet–triplet energy-transfer measurements over a range of timescales, different motions in unfolded polypeptides could be studied[61], including local motions on the 100 ps timescale, and end-to-end loop formation on the 10–100 ns timescale. The timescale of the formation of loops was seen to correlate with their length[62]. The loop formation rate scales as $N^{3/2}$, where N is the loop length defined by N residues[62–64].

The timescales of other elementary events in the protein folding reaction, namely the formation of the α-helix and the β-hairpin are only marginally slower. Stable alanine-rich α-helices appear to form in ~0.1–0.5 μs and β-hairpins in the ~5–10 μs time domain[25,31,65–67], when detected by temperature-jump experiments. More direct measurements[68] suggest that they may form as slowly as in 150 μs. It should be noted that the helices and sheets present in the folded structures of proteins are not as stable in isolation, and hence, in the context of an intact protein they may form at rates slower than those of their stable, designed counterparts[31].

Methods to study sub-ms protein folding reactions

A variety of physical signals can be used to probe folding reactions (Table 1). In most studies of protein-folding reactions, either one or two spectroscopic probes, which report on population-averaged properties of the molecular conformations, are used. It is possible that much information related to the nature of the early folding events is lost either because of ensemble averaging, or because the spectroscopic probe used remains silent to structural changes in some regions of the protein during folding. An extremely heterogeneous protein folding process may appear simple when monitored by probes which report on global structural changes. For example, it was seen in all-atom Monte Carlo simulations with protein G that, because of the asymmetric location of the only tryptophan residue in the protein, the folding kinetics was mono-exponential even when the actual reaction involved intermediates and multiple pathways[69]. Conformational heterogeneity in folding reactions has been detected in several experimental studies when multiple probes were used[37,38,40,70–72]. Thus, the use of multiple probes and site-specific probes is an important consideration in the study of sub-ms protein-folding reactions.

A sub-ms folding reaction that leads to the partial folding of a protein, manifests itself as an unobservable change in a spectroscopic signal during the burst phase of a kinetic refolding experiment with ms time resolution. For most proteins, the product of the burst phase reaction is seen to be compact, indicating that the unfolded polypeptide chain collapses during the initial sub-ms folding reaction[23,73–75]. These burst-phase changes are most commonly interpreted to represent the formation of an early folding intermediate[4], but they could conceivably also arise because of non-specific, solvent-induced conformational

Table 1. Physical measures of protein folding

Probe	Structural feature reported	Reference
Fluorescence	Environment of the intrinsic fluorophores (mainly Trp and Tyr) in a protein, or of protein-bound external fluorophores	27, 172, 217
FRET	Intramolecular distances	40, 87, 171
SAXS	Overall size and shape of the protein in different conformational states	24, 218
Circular dichroism	Secondary structure (far-UV CD) and packing of aromatic residues (near-UV CD)	68, 181
NMR	Environment of individual residues; dynamics in different conformations	94, 124
Absorbance	Environment of the absorbing chromophore	169
HX-MS/HX-NMR	Protection from hydrogen exchange due to specific structure formation	94, 187, 188
Raman spectroscopy	Environment of different residues in different conformations of a protein	66, 86
Infra-red spectroscopy	Secondary structure	219, 220
Force spectroscopy	Forces which glue the three-dimensional structure of a protein	101, 102

rearrangements in the unfolded state[76]. In order to analyse properly the earliest folding events, it is necessary to measure directly the kinetics of folding processes occurring in the sub-ms time domain. High-resolution spectroscopic probes need to be used in conjunction with fast mixing/relaxation methods that can initiate folding reactions in <1 ms. Methodologies currently being employed for the study of sub-ms protein-folding reactions include fast temperature-jump methods, laser-flash photolysis, continuous-flow mixing, NMR methods and single-molecule techniques.

Temperature (T)-jump method

This method has been used widely[55,77,78]. A dead-time as short as 5 μs can be achieved using T-jump instruments based on resistive heating[79]. With the laser T-jump technique, a dead-time as short as 2 ps is achievable[77,79]. The starting state is usually cold-denatured protein. In spite of a very low dead-time, the main caveat of the T-jump method is that complete unfolding of many proteins may not be achieved in the cold denaturated protein[80,81].

Laser-flash photolysis method

In this method, a pulse of light is used to cleave a bond in a photo-activable reagent in order to produce reactive intermediates. Flash photolysis of a CO ligand attached to cytochrome *c* has been used extensively to study folding with ns time resolution[82,83]. The main disadvantage of this method is that it cannot be used with all proteins because of the non-availability of suitable photo-dissociable ligands. It is also not straightforward to distinguish conformational changes which arise due to the refolding reaction from those which take place due to ligand dissociation. A more recent approach has been to enable very rapid pH jumps, and to study the consequent conformational changes[84].

Continuous flow mixing

The main principle behind the working of a continuous flow mixer is the generation of turbulence in the mixing of two liquids[85]. The theoretical limit for the shortest dead-time that can be achieved with this technique[85] is close to 10 μs. The sub-ms kinetics of several proteins has been measured[27,86–90] using a continuous-flow capillary mixer with a mixing dead-time of 20–120 μs. Recently, by introducing turbulence before the point where the two liquids collide, a dead-time of ~11 μs has been achieved[91]. Although not superior to the methods discussed above in terms of the dead-time, the continuous flow method can be used to study the refolding reactions of all proteins which can be unfolded by chemical denaturants, or by pH.

NMR methods

Dynamic NMR methods are useful for the study of fast protein-folding reactions which take place in the sub-ms time domain. NMR resonances are sensitive to exchange processes which occur on the 10 ms–10 μs timescale[92]; both the shape and position of the NMR peaks change due to exchange between conformationally distinct forms. For example, the folding reaction of the *N*-terminal fragment of λ repressor ($\lambda_{6\text{-}85}$), which folds completely in < 1 ms, was successfully monitored using this technique[93]. Motions within the native or the unfolded conformations which occur on the ps-ns timescales can be studied by the measurement of NMR relaxation times[94].

Single molecule methods

The heterogeneity inherent in the unfolded state and in early folding intermediates makes it difficult to characterize different sub-populations and transitions using steady-state methods. Single-molecule methods allow the identification of sub-populations which are otherwise undetectable due to population averaging in bulk experiments. Single-molecule methods utilizing fluorescence resonance energy transfer (FRET)[95–97] and fluorescence correlation spectroscopy (FCS) methods[98,99] as well as atomic force microscopy[100–102] have provided useful insights into the nature of the early events in protein-folding reactions. Although the time regime of < 1 ms is still not accessible in kinetic

experiments utilizing single-molecule methods, these methods provide information on the sub-ms dynamics of unfolded proteins, which are important determinants of the rates of the fastest initial steps of folding reactions[96]. The development of new methods which allow monitoring of the same molecule over an extended period of time[103,104], is expected to make single-molecule experiments more informative.

The nature of the unfolded state may dictate early events in folding

A proper understanding of the earliest events in protein folding not only involves a characterization of the events during folding, but also requires a detailed description of the structure and dynamics of the starting species, i.e. the unfolded state. Understanding the conformational heterogeneity present in the unfolded state is also important because a large number of conformations implies that a large number of microscopic pathways may be operative in protein-folding reactions[105,106].

Topology of the unfolded state

Unlike the native state, the unfolded state is a heterogeneous ensemble of different unstructured forms[72,97,107–109]. The unfolded state in high denaturant concentrations is usually treated as an unperturbed random coil[110,111]. However, a major drawback of the unperturbed random-coil model is the omission of the excluded volume effect. This would mean that impossible conformations like two or more non-bonded atoms in the chain occupying the same space at the same time are not excluded[112].

While the unfolded states of some proteins do not appear to possess residual structure[111,113,114], there is increasing evidence for the presence of residual structure in the unfolded states of other proteins[115–119]. Thus, the validity of a statistical random-coil model to describe the starting species of a protein-folding reaction may appear inappropriate[40,120,121]. On the other hand, it appears that the presence of residual structure may have little or no effect on the random-coil statistics of an unfolded polypeptide chain[122].

The unfolded state may possess a native topology even under highly denaturing conditions[118,123]. For example, in the case of lysozyme, a single tryptophan residue was shown to stabilize a network of hydrophobic clusters in the unfolded state[123]. Recently, a pre-existing, hydrophobically collapsed conformation with both native-like and non-native interactions was shown[124] to exist in the unfolded state of a Trp-cage miniprotein, TC5b. Several proteins show conformational preferences even at high denaturant concentrations[117,118], including local hydrophobic clustering[115,119,125], and helical structures[125–127]. Such residual structures could serve as the seed for the development of the native fold, once the unfolded protein is transferred to refolding conditions.

Dynamics within the unfolded state ensemble

The unfolded state is a highly dynamic state[128], because different conformers in the heterogeneous unfolded state ensemble differ only marginally in their stabilities. Understanding the dynamic behaviour of the unfolded state is important for an assessment of the conformational entropy at the beginning of a protein-folding reaction, as well as for the determination of the fastest rate with which different conformations can interconvert within the unfolded state population.

There is much data available from NMR experiments on conformational fluctuations in the unfolded state[94]. In the case of apomyoglobin, for example, fluctuations on the μs–ms timescale were observed within some regions of the unfolded polypeptide chain with helical propensity[125]. The fact that these local elements of structure are highly dynamic and unstable[129] suggests that rapid interconversion of different conformers occurs within the unfolded-state ensemble.

Studies utilizing methods such as single-molecule FRET[95,98,130] and FCS[99,131] have also provided useful insights into the dynamics of the unfolded state. FCS measurements of the equilibrium structural fluctuations in unfolded apomyoglobin[99] have shown complex relaxation characterized by three time constants of 200, 30 and 3 μs. The 3 and 30 μs relaxations were attributed to motions within the U conformations, and the longer relaxation time (200 μs) was attributed to fluctuations which are common to both U and I. For ribonuclease H (RNase H), a chain reconfiguration time of 20 μs was measured, which is comparable to the 30 μs relaxation observed in the case of apomyoglobin[130]. Intriguingly, transitions between unfolded conformations on the seconds timescale were also seen for RNase H[130]. On the other hand, for an apparent two-state folder, the cold shock protein from *Thermotoga maritima* (CspTm), a chain reconfiguration time of ~50 ns was reported[132], which is much faster than the values reported for apomyoglobin and RNase H. Thus, there seems to be some variation in the measured reconfiguration times of unfolded polypeptide chains. The differences in chain dynamics may be due to a varied degree of ruggedness of the free-energy surface which describes the different unfolded conformations.

Unfolded state in refolding conditions

The heterogeneity and dynamic behaviour of the unfolded state makes it likely that the conformation of the unfolded state is altered when the unfolded chain is transferred to refolding conditions. Random polymers are known to change their conformation when placed in different sol-

vent conditions; they adopt extended conformations in 'good solvents', random flight conformations in 'theta solvents', and collapsed conformations in 'bad solvents'[107]. Intrinsically unstructured proteins are known to exist in collapsed conformations in the absence of denaturants[130]. Polypeptide chains that are kept unfolded by high pH[133,134] truncation[135,136] or reduction in the disulphide bond[137], adopt collapsed conformations when transferred to conditions that favour the native structure. For example, in the case of the pH 12-unfolded form of barstar, the addition of salt results in a collapse of the unfolded protein molecules[138]. An engineered mutant of *Drosophila* Engrailed Homeodomain (EnHD), L16A, exists as a compact denatured form at physiological ionic strengths[139,140]. Thus, unfolded polypeptide chains appear to behave like random polymers in terms of their interaction with different solvents. This suggests that the early folding events may be dominated by non-specific interactions of the polypeptide chain with solvent molecules. Often, compact unfolded conformations of proteins appear to have accumulated elements of structure[140,141]. The interesting question is when and how specific interactions within the polypeptide chain develop during the initial folding reaction of a protein?

Initial collapse in protein-folding reactions

The observation that the unfolded state in native conditions is compact suggests that the folding reaction begins with a solvent-driven collapse of the unfolded polypeptide chain. The occurrence of such a collapse reaction before the main folding reaction of proteins was also observed in several lattice model simulations of protein folding[142,143]. In these simulations, the collapse of the random-coil unfolded form resulted in a reduction of conformational entropy, as there were fewer conformations in the collapsed form than in the completely unfolded form. This highlights the role of the initial collapse reaction in solving the conformational search problem.

Early experimental evidence regarding the occurrence of an initial collapse reaction was based on the observation of a burst-phase loss in the signals of different spectroscopic probes in ms measurements[23,134,144–149]. The observed burst-phase change was attributed to the formation of an early collapsed intermediate[150]. The presence of specific structure in these initial intermediate forms was inferred from observations of burst-phase changes in circular dichroism (CD) and amide proton protection at a few ms of the folding reaction[150]. The identification of molten globule (MG) intermediates in equilibrium experiments[151,152], and the similarity of the kinetically detected burst-phase folding intermediates to these MG-forms, led several studies to conclude that the MG-forms of proteins are equilibrium analogues of the initial kinetic intermediates[145,151,153].

When the hydrophobic dye, 8-anilino-1-naphthalene sulphonic acid (ANS) was used as a probe to detect structural transitions during protein-folding reactions, a significant enhancement of the ANS fluorescence was seen to occur in the < 1 ms time domain[23,73,154]. This suggested that solvent-exposed hydrophobic patches develop during the sub-ms refolding reactions of these proteins[23,154]. The involvement of hydrophobic interactions in the initial collapse reaction was also inferred from mutational analyses of the burst-phase signal amplitude[147,155]. The burst-phase amplitude decreases upon substitution of a hydrophobic amino acid residue by alanine[147,155]. These studies suggested a possible role of hydrophobic interactions during sub-ms folding reactions. Consolidation of the hydrophobic core appears to occur only during the later stages of the folding reaction, after the initial collapse reaction[156,157]. This suggests that the initial hydrophobic collapse reaction is likely to result in the formation of loose hydrophobic clusters in the collapsed form, which remains largely hydrated during the initial few ms of folding[38,158,159]. Interestingly, the initial hydrophobic collapse reaction precedes any structure formation in the case of barstar[23], when folding is carried out in marginally stabilizing conditions.

Surprisingly, a few proteins do not appear to undergo an initial sub-ms collapse reaction at all[160]. Stopped-flow SAXS measurements showed that both common-type acyl-phosphatase and a mammalian ubiquitin (Ub) variant F45W do not undergo any reduction in chain dimensions when transferred from a high denaturant concentration to folding conditions. It is unlikely, as suggested[160], that pure water can be as good a solvent for the unfolded state as a concentrated denaturant solution. One explanation could be that the collapsed forms are unstable high-energy conformations on the folding pathways of these 'two-state' folding proteins and hence not populated. Another possibility is that the coupling between collapse and structure formation is so strong in such sequences that the two events happen on overlapping timescales. In some evolutionarily optimized amino acid sequences (as are the sequences of the apparently two-state folding proteins which fold in less than a ms), the probability of successful formation of native contacts may be so high that structure formation can happen at a timescale which is practically inseparable from the timescale of chain collapse. At present, it is not clear which features of the amino acid sequence or tertiary structure are critical for the occurrence of an initial collapse reaction.

On the other hand, a fast chain collapse appears to precede structure formation in the case of the *Bacillus caldolyticus* cold shock protein (Bc-Csp), apparently a 'two-state' folder[161]. It is possible that chain collapse will be observed to precede structure formation for other 'two-state' folders as well, provided suitable probes are used. Since the absence of an initial collapse reaction has been seen only in SAXS measurements[29,160], and an initial

collapse reaction is seen in FRET measurements[161,162] of the folding of 'two-state' folders, it will be important to determine if SAXS is as sensitive as FRET in detecting a collapse of the polypeptide chain.

The folding reactions of many non two-state folding proteins are slow. One reason for this could be that in these cases there is a higher probability for wrong contacts to form because the amino acid sequence is not maximally optimized for fast folding. Consequently, folding molecules with wrong interactions may encounter an energy barrier that prevents them from folding quickly to any stable structure. In such cases, folding may be slowed down to such an extent that it happens over a much slower timescale than the chain collapse reaction.

Cooperativity of the initial collapse reaction

Does a free energy barrier separate the collapsed form from the completely unfolded form? The existence of a free-energy barrier between the collapsed and unfolded conformations (Figure 4 *a*) would imply that the collapse reaction is a cooperative all-or-none transition, in which different interactions in the collapsed conformation develop in a synchronous manner. By contrast, a barrierless unimodal transformation of the unfolded state to a collapsed state (Figure 4 *b*) is expected to result in a highly non-cooperative and gradual transition. It is difficult to experimentally distinguish between these two possibilities.

In studies with ms time resolution, a sigmoidal dependence of the burst phase amplitude on the concentration of the denaturant is usually interpreted as the signature of a cooperative two-state transition between the unfolded state and the collapsed intermediate form[73,145,149,158]. A sigmoidal curve is usually considered as the hallmark of a cooperative structural transition in analogy with an all-or-none phase transition. But a reaction involving a continuum of conformations, in which the conformers are related to each other by a linear free-energy relationship, can also give rise to a sigmoidal dependence of the burst-phase amplitude on the denaturant concentration[107,163]. Moreover, the denaturant dependences of the burst-phase amplitude monitored by different spectroscopic probes are seen to be non-coincident for some proteins[38,40,70,144,164]. Such observations suggest that the formation of burst-phase products can be a non-cooperative transition. In fact, in many cases, the dependence of the burst-phase amplitude on denaturant concentration appears to be as gradual as the non-cooperative melt of isolated helices[136,137,144,165,166].

Several equilibrium studies have indicated that the collapse transition is likely to be a higher-order process. In the case of α-lactalbumin and barstar, highly non-cooperative unfolding transitions were observed for the equilibrium MG-models of the collapsed form[138,157,167]. Single-molecule studies with several proteins, including RNase H[130], Csp*Tm*[95,132,168] and protein L[98] have also suggested that the transition which leads to the formation of the collapsed state is a gradual one. Interestingly, the thermal unfolding of the compact (and structured) denatured form of EnHD variant L16A appeared to be gradual and non-cooperative[139]. Such a transition is not expected for a folding intermediate which is separated from the completely unfolded form by a significant energy barrier.

The extent of contraction of several intramolecular distances in the burst-phase product formed during the initial folding reaction of barstar was measured as a function of denaturant concentration[38,40], and was seen to be non-sigmoidal and gradual. Different intramolecular distances appeared to contract in an asynchronous manner. The results from these studies suggest that sub-ms folding reactions may be continuous non-cooperative processes. Importantly, it also appears from these studies that the products of the sub-ms folding reaction of barstar are compact forms with specific structure[40]. It is not clear as to how a highly non-cooperative process leads to the formation of products with specific structure.

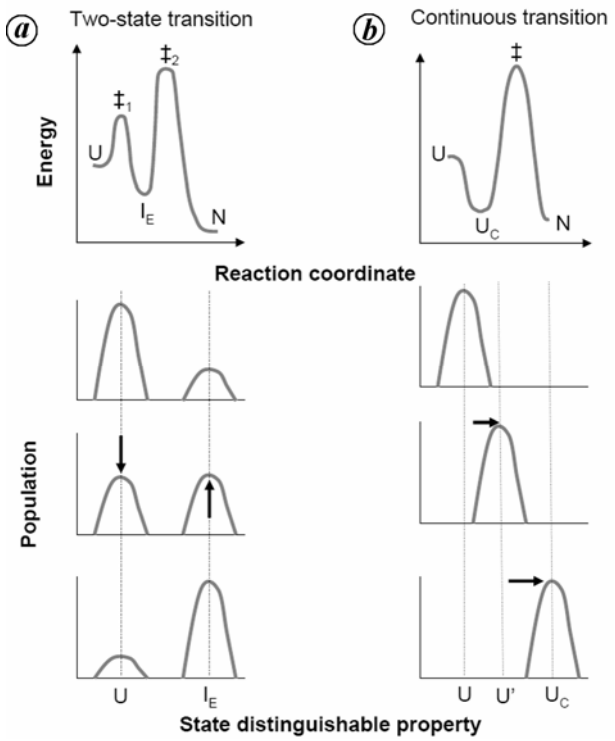

Figure 4. Nature of the initial collapse transition. (***a***) Cooperative transition between the completely unfolded form (*U*) and the collapsed intermediate form (I_E). Both the states co-exist with a sizeable free energy barrier separating the two. The two populations are in constant exchange and their equilibrium ratio is determined by the extent of stability conferred by the reaction conditions. (***b***) Continuous transition involves a continuum of near-degenerate forms of similar densities, and consequently, no energy barrier exists between *U* and the collapsed denatured form (U_C). The population of molecules shows a unimodal shift from *U* to U_C, with change in the folding conditions.

In direct measurements of sub-ms folding reactions by rapid mixing methods, exponential or multi-exponential kinetics are usually seen[27,87–88,169–171]. The observation of exponential sub-ms folding kinetics has been interpreted in terms of a cooperative collapse transition[88,169–172]. There is a strong belief that exponential kinetics arises due to the crossing of a free-energy barrier, and that the product of such a barrier-crossing reaction is likely to be a specific structured intermediate[25,169]. But gradual transitions can also show kinetics which is virtually indistinguishable from single exponential kinetics[54,55,163,173,174]. In a downhill folding funnel, multiple parallel routes can also result in exponential kinetics because of different degrees of parallelization of the routes at different time points of folding[106]. Thus, the observation of single exponential kinetics need not necessarily mean a barrier-crossing event. In the absence of a free energy barrier, the kinetics can be non-exponential[20,25,175], and probe-dependent[52,53]. But, non-exponential kinetics as well as probe-dependent kinetics can also arise due to multiple pathways[176] or due to a distribution of coupled entropic barriers[177]. It has only now become clear that a barrier-limited reaction cannot be distinguished from a barrierless process, solely on the basis of the reaction showing exponential or non-exponential kinetics.

Intriguingly, for some proteins, the fast sub-ms rate constant appears to be essentially insensitive to denaturant concentration[27,87,88,90,171]. For one such protein, barstar[90], this observation has been interpreted to signify the absence of a dominant barrier during the initial sub-ms folding reaction. For another such protein, cytochrome *c*, it was concluded that there is a free-energy barrier between the unfolded state and the collapsed intermediate form[87]. But this conclusion was based on the observation of exponential folding kinetics, and of an activation enthalpy of ~ 3.5 kcal/mol, after correction for the temperature dependence of the viscosity term in Kramer's equation[178]. The difficulty in concluding that a reaction is barrier-limited on the basis of the kinetics being exponential or not, has been discussed above. The validity of the conclusion that the collapse reaction of cytochrome *c* is barrier-limited is also doubtful, because even unstructured fragments of cytochrome *c* (which remain unstructured upon collapse) show exponential sub-ms collapse kinetics with similar rate constants and a comparable activation enthalpy[179].

An additional difficulty in determining the cooperativity of the initial chain collapse reaction is the observation of multi-step sub-ms kinetics in several cases[24,27,87,180]. For example, in the case of cytochrome *c*, secondary structure develops over a slower timescale (~ 2000 s^{-1}), after a faster chain collapse phase (~15,000 s^{-1})[24,87,181]. It is not easy to establish if such multi-phasic sub-ms kinetics arises due to the presence of multiple steps or due to the existence of multiple routes. Nevertheless, the products of the folding reaction populated at 1 ms of folding are likely to be a mixture of conformations. The structural heterogeneity present during a sub-ms folding reaction makes it difficult to infer the exact nature of conformational transformations.

If the initial collapse reaction is indeed all-or-none, coexistence of both the compact and extended unfolded forms should be observed in experiments[107,182]. In the case of cytochrome *c*, time-resolved FRET-monitored stopped-flow refolding experiments showed that the distance distribution has two major components at 1 ms of refolding, corresponding to the completely unfolded form, and a compact form with an almost native-like intramolecular distance[183]. But it could not be ascertained whether the two forms are in exchange with each other. A subsequent experiment using multiple FRET-pairs, suggested that the initial sub-ms collapse transition involves other intermediate forms as well, and the sub-ms folding reaction is therefore more likely to be non-cooperative[37]. Site-specific variations were also seen in direct measurements of the sub-ms folding phase of cyt *c*′ by time-resolved FRET[184].

Thus, a simple 'two-state' model may not appropriately describe the complex sub-ms folding reactions of proteins. There can be two interpretations for the origin of complexity in the sub-ms collapse reactions. The collapse can be regarded as a 'two-state' transition with static disorder, in which case the transition would involve multiple pathways and several small activation barriers[41,185]. Alternatively, it can be regarded as a higher-order reaction with dynamic disorder, which would imply that the reaction involves several steps[41,185]. Direct kinetic measurements of the sub-ms collapse reaction with multiple site-specific probes are needed to provide insight into the exact mechanism of the initial collapse transition.

Initial collapse: specific or non-specific?

A specific collapse reaction would lead to partial formation of structure that is also present in the final, fully folded form of the protein. On the other hand, a non-specific collapse reaction would lead to a compact form not containing any native-like structure. The question whether a specific intermediate forms on the folding pathway at the end of the initial sub-ms folding reaction of a protein, has been a contentious one[4,76]. The major evidence supporting the formation of a specific burst-phase intermediate (specific collapse) has been the observation of a ms burst-phase change in a measured spectroscopic property (e.g. fluorescence or CD)[70,146,147,158], observation of exponential kinetics in directly measured sub-ms folding reactions[87,170], non-exponential dependences of folding rates on denaturant concentration[146,186], protection from HX pulse-labelling[187,188], and a discrepancy in the values of ΔG and m obtained from kinetic and equilibrium experiments[189]. But burst-phase changes may also originate as the consequence of a solvent-driven

response of the unfolded polypeptide chain when it is transferred from highly denaturing conditions to those that favour folding[76]. Such a behaviour would be analogous to that of a random polymer showing different responses to different solvent conditions[107]. A polymer-like collapse reaction is expected to be non-specific and the products of such reactions are likely to be randomly collapsed species without any specific structure[4,150].

There is some evidence in favour of the initial collapse reaction being non-specific in nature. In the case of ribonuclease A (RNase A), an unstructured disulphide-broken analogue of the protein (rRNase A) can reproduce the CD burst-phase change seen with the disulphide-intact folding protein[137]. When transferred to lower denaturant concentrations, unstructured fragments of cytochrome *c* show a burst-phase jump in the fluorescence signal identical to that seen with the intact protein[136]. Several experimental artifacts may also suggest the presence of an early folding intermediate, even when it is actually not present[76]. These include dead-time artifacts, poor signal-to-noise ratio, aggregation, low HX-protection or, even the use of different instruments to cover kinetics in different time domains[76]. It has been suggested that stable folding intermediates are usually not populated before the formation of the main transition state, and when they do, it is only because optional barriers arise on the folding pathway of proteins[160].

The principal reason why the initial collapse reaction is considered to be specific is the observation of specific structure in the product of the burst-phase reaction[23,144–149]. In the case of Csp*Tm*, a significant amount of β-sheet structure accumulates in the product of the sub-ms folding reaction[168]. The early intermediate I_E, which is known to populate the folding pathway of barstar at a few ms of the refolding reaction, possesses secondary structure, the extent of which varies with the stabilizing effects of the solvent conditions[70,158]. In the case of *Escherichia coli* dihydrofolate reductase, the compact intermediate which accumulates at a few ms of the folding reaction, possesses a significant amount of secondary structure[28]. The denatured state of the EnHD variant, L16A, also possesses a specific structure[140], and it was suggested that it is a folding intermediate poised to fold into the native state. The unfolded state of disulphide-intact RNase A behaves differently from its non-folding analogue, i.e. disulphide-reduced RNase A (rRNase A)[190]. The folding RNase A chain, at 22 ms after the commencement of the folding reaction, is more compact and distinct in shape in comparison to the rRNase A chain under identical conditions. Moreover, RNase A binds to ANS during the sub-ms folding reaction, but rRNase does not. In the case of RNase H, by inducing protein folding/unfolding by mechanical means, it was suggested that the early intermediate form[149], which forms in <10 ms, is a discrete thermodynamic state[101]. In the case of monellin, the shape of the product of the sub-ms folding reaction is different from that of the completely unfolded state[190].

In the case of barstar, measurement of 11 intramolecular distances by FRET, in the early intermediate, I_E, at a few ms of the folding reaction, shows that some distances contract only to an extent expected for the unfolded polypeptide chain[38,40]. However, the contraction of several other distances, exceeds the extent of contraction expected from only a solvent-driven response of the unfolded polypeptide chain[40]. From the results of these studies, it appears that the initial sub-ms collapse reaction involves two components, a non-specific component, which represents a solvent-induced contraction of the unfolded polypeptide chain, and a specific component, which originates from the formation of specific structure in the product of the sub-ms folding reaction[38,40]. Interestingly, the dimensions of two unstructured fragments of barstar do not change with a change in the denaturant concentration[40].

In an earlier study with barstar, the product of the sub-ms folding reaction was shown to possess specific secondary structure under strongly stabilizing conditions[70], whereas in marginally stabilizing conditions there was no detectable structure in the product of sub-ms folding. Importantly, the structure present in the product of the sub-ms folding reaction is specific to the specific folding conditions employed[70]. This study suggests that, upon a change in the folding conditions, one or more of the structural components in the initial intermediate ensemble can be stabilized preferentially, and that, under different folding conditions different folding pathways become operative. Surprisingly, this issue has generally not been addressed for other proteins.

Collapse and structure formation

Do polypeptide chain collapse and secondary structure formation occur concurrently, or do the two events happen on different timescales? This has been an important question. In native-like conditions, the unfolded states of proteins can be both compact and structured[4,124,140]. In the equilibrium molten-globule forms of several proteins, like those of apomyoglobin[191] and bovine α-lactalbumin[192], significant native-like secondary structure is present. The amount of secondary structure increases progressively as the polypeptide chain becomes more compact on the equilibrium folding pathway of apomyoglobin[191]. In the case of barstar, a salt-induced collapse of the high-pH unfolded form is followed by accumulation of secondary structure[138,157]. Thus, it appears from these studies of equilibrium models of chain collapse, that increased chain compaction may facilitate the development of secondary structure during the folding reaction.

Simulations show that increasing compactness results in the accumulation of different elements of secondary structure[193]. The elements of secondary structure accumu-

late in the compact denatured forms possibly because these structures are entropically favoured among the different possible conformations in the population of compact molecules[193]. There is evidence, both from experiments and simulations, that helix formation is governed by a conformational diffusion search in the ensemble of coil conformations[67,194]. The intrinsic helix–coil equilibrium constant for a six-residue stretch with four rotatable bonds is 1.05, which corresponds to 0.20 kT in energy[193]. This kind of energy barrier can be easily traversed through a diffusive search. It is possible that the packing interactions and tertiary contacts developed during the later stages of the folding reaction stabilize the helical structures, so that a helix becomes favourable over the coil by 1.7 kT.

Until recently, it was not possible to discern the actual sequence of events during sub-ms folding reactions, because they could not be resolved temporally. Nevertheless, it was shown in the case of barstar, that a fast hydrophobic collapse precedes the formation of secondary structure[23], for folding in marginally stabilizing conditions. In the case of proteins for which sub-ms folding has been resolved temporally, such as RNase A and BBL, a fast non-specific collapse is seen to precede structure formation[26,27]. In the case of cytochrome *c* and monellin also, it was seen that a significant structure develops only after a fast collapse reaction[24,195,196]. In the case of apomyoglobin, it was not possible to resolve temporally collapse and secondary structure formation within the 300 μs dead-time of the mixing instrument[180]; consequently, it could not be ascertained if the collapse reaction precedes structure formation in this case.

It is still unclear whether fast chain collapse preceding secondary-structure formation is a general feature of protein-folding reactions. As discussed above, for a few apparent two-state folders, chain collapse and structure formation appear to occur in a concerted manner, but for CspB, a fast chain collapse is seen to precede structure formation[158]. In this context, it is notable that the formation of secondary structure has not been studied in the sub-ms time domain for most proteins[28,70,168,180].

Are there multiple pathways for the chain collapse reaction?

The unfolded state is an extremely heterogeneous ensemble of conformations[72,107–109]. Do different conformations in the unfolded-state ensemble trace their own unique folding trajectories, or do they all converge into a single folding-competent conformation before the folding reaction begins? In the latter case, different conformations in the unfolded-state ensemble would have to convert into a folding-competent conformation. In such a case, the presence of a large number of folding-incompetent unfolded conformations may result in a kinetic bottleneck leading to a significant slowing down of the folding rate[197], even when there is fast exchange among them. This suggests that multiple folding trajectories must be present during the earliest phases of the protein-folding reactions. Multiple microscopic routes and pathways are seen in protein-folding simulations[69,198,199]. Structural heterogeneity seen even during later stages of protein-folding reactions, is indeed suggestive of multiple folding pathways[37,71,102,200–203].

The product of the sub-ms folding reaction appears to be structurally heterogeneous[37,70,164,203] for several proteins, and the heterogeneity must arise because of multiple folding routes. In a recent multi-site FRET study of the folding of barstar, it was seen that the urea dependences of the extent of contraction of different intramolecular distances in the initially collapsed form are highly uncorrelated[38,40]. This suggests that the initial chain-collapse reaction of barstar might occur via multiple routes.

In order to reach the unique native fold, structural heterogeneity must reduce during the protein-folding reaction. It was seen that for the slow folding reaction of barstar, studied using multi-site time-resolved FRET measurements, conformational heterogeneity in the late folding intermediate, I_L, reduces with increasing stability[71]. This finding is consistent with statistical mechanical models which predict a progressive reduction in structural heterogeneity during folding reactions[20].

Folding along multiple tracks explains why the earliest events, including the chain collapse reactions are so fast. Observed folding rates are seen to be proportional to the number of microscopic routes available to a protein[46]. Interestingly, proteins with higher α-helical content appear to fold faster than α–β or β proteins. This is also supported by experimental observations that a majority of ultra-fast folding proteins are α-helix-rich[31,46]. A possible explanation is that α-helical proteins have more folding routes because helix nucleation can occur at multiple sites, whereas β-structures usually nucleate from a single site[46].

How fast is the initial chain-collapse reaction?

The rate at which an unfolded polypeptide chain collapses cannot be faster than the fastest rate of intramolecular contact formation[83]. The formation of a ~60 residue loop in the case of cytochrome *c* was seen[83,204] to take place with a time constant of ~3 μs. Based on this study, an upper limit of 10^6 s^{-1} was proposed for the rate of the initial polypeptide collapse reaction[83,204]. This is a useful estimate, but it is unlikely to be a universal limit for all polypeptide sequences. For example, it has been observed in some studies that the kinetics of intrachain loop formation is dependent both on the sequence[61] and the position of the loop-forming residues[205]. Moreover, it

should be remembered that many hundreds of intramolecular contacts need to form during the folding of even a small protein.

The observed timescale of the collapse reaction for different proteins has a range spanning from a few tens of ns to several seconds[26,29,87]. The timescale for the occurrence of a fast non-specific collapse was experimentally found to be ~60 ns for BBL[26]. Similar results were obtained in single-molecule studies with Csp*Tm*, where a chain reconfiguration time of 50 ns was observed in the unfolded state[132]. The ultra-fast collapse transitions which happen on the sub-μs timescale may represent the non-specific response of an unfolded polypeptide chain upon a change to refolding conditions. For RNase H, however, a chain reconfiguration time of 20 μs was observed in single-molecule studies, which is close to the timescale of the sub-ms folding reaction observed in ensemble measurements[130]. The differences seen in the values of the rates of chain collapse for different proteins possibly originate from the difference in the roughness of the energy surface a collapsing polypeptide chain has to traverse. The ruggedness of the energy surface of a folding reaction is likely to be a complex function of the chain composition and chain length. In direct measurements of sub-ms refolding kinetics, major kinetic phases with rate constants of 65 μs (cytochrome *c*), 600 μs (protein G), 100 μs (acyl-CoA binding protein), 80 μs (RNase A) and 150 μs (bacterial immunity protein, Im7) were observed[27,87,88,171]. Because these timescales are relatively long compared to those of the elementary events in protein folding, it has been argued that the sub-ms kinetics measured for these proteins represents specific events, and that the presence of a large barrier slows down the initial folding reactions[169].

For a few apparently two-state folding proteins, chain collapse and structure formation occur concomitantly and in such cases, the timescale of the collapse reaction is usually much longer[19,29]. It is possible that in these proteins the collapse reaction is slowed down because it is coupled to specific structure formation, the extent of which may vary from protein to protein. Another possibility is that a fast chain collapse remains undetected in the case of many such proteins because of the nature of the probes utilized. For example, in the case of Bc-Csp, a fast chain collapse precedes the main folding transition when probed by FRET[161,162]; whereas in conventional fluorescence-monitored folding experiments and thermodynamic analyses, the folding reaction of the protein appeared to be two-state[206–208]. Surprisingly, the kinetics of the collapse reaction of Bc-Csp could not be time-resolved even with a 10 ns dead-time[162].

It is not clear at present why an initial collapse reaction is not observed with the ultra-fast folders. For EnHD, the denatured state is substantially structured[139,209], and this residual structure could possibly be the reason for the fast folding of EnHD[48]. Similar arguments have been made to explain the ultra-fast folding of BBL[48]. TC5b is a small (20 amino acids) protein, which despite a relatively high contact order atypical for a protein of this length, shows ultra-fast folding[210]. Photo-CIDNP (chemically induced dynamic nuclear polarization) NMR measurements, have shown that the unfolded state of TC5b has residual structure because of a pre-existing hydrophobic collapse[124]. This observation provides a possible explanation for why many of the ultra-fast folding proteins do not show an initial hydrophobic collapse preceding the main folding transition, upon initiation of refolding.

The observed differences in the timescales of the collapse reaction seen with different proteins appear to be difficult to explain. It has been suggested that a difference in topological frustration or roughness of the energy surface can give rise to differences in the folding rates[211]. The folding rates of proteins are known to correlate well with the native-state topology[212–214]. In a recent analysis, a striking correlation between some structural parameters (absolute contact order, chain length and the number of non-native contact clusters), and the folding rates of both apparent two-state folders and the non two-state folders was observed[215,216]. Thus, it is also possible that the difference seen for the rate of the chain-collapse reaction of different proteins might be a function of the topology and structure of the protein.

Concluding remarks

The study of initial events during protein folding is crucial to obtain an understanding of protein-folding reactions. It appears now that polypeptide chain collapse plays a dominant role in shaping the early events. The nature and extent of chain collapse may determine the extent of the specific structure that is formed. Since chain collapse is determined by solvent conditions, the nature of the product of folding at 1ms is expected to depend on the folding conditions. If this is indeed true, then the subsequent folding reaction may follow different routes in different folding conditions, because the starting collapsed folding intermediate is different for different folding conditions.

Understanding the nature of the barriers that slow down the fast folding reactions that complete within the first millisecond has become an important issue. A single dominant barrier would dictate that the compact ensemble at 1 ms coexists with the extended unfolded-state ensemble. Alternatively, it now seems possible that the initial collapse reaction may be so highly non-cooperative so as to be a gradual, continuous transition. The latter possibility raises the exciting possibility that it may become possible to examine in real time how proteins begin to fold.

1. Zwanzig, R., Szabo, A. and Bagchi, B., Levinthal's paradox. *Proc. Natl. Acad. Sci. USA*, 1992, **89**, 20–22.

2. Levinthal, C., Are there pathways for protein folding? *J. Chem. Phys.*, 1968, **65**, 44–45.
3. Ramachandran, G. N. and Sasisekharan, V., Conformation of polypeptides and proteins. *Adv. Protein Chem.*, 1968, **23**, 283–438.
4. Roder, H., Maki, K. and Cheng, H., Early events in protein folding explored by rapid mixing methods. *Chem. Rev.*, 2006, **106**, 1836–1861.
5. Dyer, R. B., Ultrafast and downhill protein folding. *Curr. Opin. Struct. Biol.*, 2007, **17**, 38–47.
6. Fersht, A., *Structure and Mechanism in Protein Science*, W. H. Freeman & Company, 1998, p. 631.
7. Ptitsyn, O. B., Stages in the mechanism of self-organization of protein molecules. *Dokl. Akad. Nauk. SSSR*, 1973, **210**, 1213–1215.
8. Kim, P. S. and Baldwin, R. L., Specific intermediates in the folding reactions of small proteins and the mechanism of protein folding. *Annu. Rev. Biochem.*, 1982, **51**, 459–489.
9. Udgaonkar, J. B. and Baldwin, R. L., NMR evidence for an early framework intermediate on the folding pathway of ribonuclease A. *Nature*, 1988, **335**, 694–699.
10. Karplus, M. and Weaver, D. L., Protein-folding dynamics. *Nature*, 1976, **260**, 404–406.
11. Bashford, D., Cohen, F. E., Karplus, M., Kuntz, I. D. and Weaver, D. L., Diffusion-collision model for the folding kinetics of myoglobin. *Proteins*, 1988, **4**, 211–227.
12. Karplus, M. and Weaver, D. L., Protein folding dynamics: the diffusion-collision model and experimental data. *Protein Sci.*, 1994, **3**, 650–668.
13. Robson, B. and Pain, R. H., Analysis of the code relating sequence to conformation in proteins: possible implications for the mechanism of formation of helical regions. *J. Mol. Biol.*, 1971, **58**, 237–259.
14. Dill, K. A., Theory for the folding and stability of globular proteins. *Biochemistry*, 1985, **24**, 1501–1509.
15. Gutin, A. M., Abkevich, V. I. and Shakhnovich, E. I., Is burst hydrophobic collapse necessary for protein folding? *Biochemistry*, 1995, **34**, 3066–3076.
16. Wetlaufer, D. B., Nucleation, rapid folding and globular intrachain regions in proteins. *Proc. Natl. Acad. Sci. USA*, 1973, **70**, 697–701.
17. Fersht, A. R., Optimization of rates of protein folding: the nucleation–condensation mechanism and its implications. *Proc. Natl. Acad. Sci. USA*, 1995, **92**, 10869–10873.
18. Daggett, V. and Fersht, A. R., Is there a unifying mechanism for protein folding? *Trends Biochem. Sci.*, 2003, **28**, 18–25.
19. Jackson, S. E., How do small single-domain proteins fold? *Fold Des.*, 1998, **3**, R81–R91.
20. Bryngelson, J. D., Onuchic, J. N., Socci, N. D. and Wolynes, P. G., Funnels, pathways, and the energy landscape of protein folding: a synthesis. *Proteins*, 1995, **21**, 167–195.
21. Dill, K. A. and Chan, H. S., From Levinthal to pathways to funnels. *Nature Struct. Biol.*, 1997, **4**, 10–19.
22. De Prat Gay, G., Ruiz-Sanz, J., Neira, J. L., Itzhaki, L. S. and Fersht, A. R., Folding of a nascent polypeptide chain *in vitro*: cooperative formation of structure in a protein module. *Proc. Natl. Acad. Sci. USA*, 1995, **92**, 3683–3686.
23. Agashe, V. R., Shastry, M. C. and Udgaonkar, J. B., Initial hydrophobic collapse in the folding of barstar. *Nature*, 1995, **377**, 754–757.
24. Akiyama, S., Takahashi, S., Kimura, T., Ishimori, K., Morishima, I., Nishikawa, Y. and Fujisawa, T., Conformational landscape of cytochrome *c* folding studied by microsecond-resolved small-angle X-ray scattering. *Proc. Natl. Acad. Sci. USA*, 2002, **99**, 1329–1334.
25. Ferguson, N. and Fersht, A. R., Early events in protein folding. *Curr. Opin. Struct. Biol.*, 2003, **13**, 75–81.
26. Sadqi, M., Lapidus, L. J. and Munoz, V., How fast is protein hydrophobic collapse? *Proc. Natl. Acad. Sci. USA*, 2003, **100**, 12117–12122.
27. Welker, E., Maki, K., Shastry, M. C., Juminaga, D., Bhat, R., Scheraga, H. A. and Roder, H., Ultrarapid mixing experiments shed new light on the characteristics of the initial conformational ensemble during the folding of ribonuclease A. *Proc. Natl. Acad. Sci. USA*, 2004, **101**, 17681–17686.
28. Arai, M., Kondrashkina, E., Kayatekin, C., Matthews, C. R., Iwakura, M. and 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.
29. Plaxco, K. W., Millett, I. S., Segel, D. J., Doniach, S. and Baker, D., Chain collapse can occur concomitantly with the rate-limiting step in protein folding. *Nature Struct. Biol.*, 1999, **6**, 554–556.
30. Gruebele, M., Protein folding: the free energy surface. *Curr. Opin. Struct. Biol.*, 2002, **12**, 161–168.
31. Kubelka, J., Hofrichter, J. and Eaton, W. A., The protein folding 'speed limit'. *Curr. Opin. Struct. Biol.*, 2004, **14**, 76–88.
32. Shastry, M. C. and Udgaonkar, J. B., The folding mechanism of barstar: evidence for multiple pathways and multiple intermediates. *J. Mol. Biol.*, 1995, **247**, 1013–1027.
33. Colon, W. and Roder, H., Kinetic intermediates in the formation of the cytochrome *c* molten globule. *Nature Struct. Biol.*, 1996, **3**, 1019–1025.
34. Patra, A. K. and Udgaonkar, J. B., Characterization of the folding and unfolding reactions of single-chain monellin: evidence for multiple intermediates and competing pathways. *Biochemistry*, 2007, **46**, 11727–11743.
35. Bhuyan, A. K. and Udgaonkar, J. B., Observation of multistate kinetics during the slow folding and unfolding of barstar. *Biochemistry*, 1999, **38**, 9158–9168.
36. Garcia-Mira, M. M., Sadqi, M., Fischer, N., Sanchez-Ruiz, J. M. and Munoz, V., Experimental identification of downhill protein folding. *Science*, 2002, **298**, 2191–2195.
37. Pletneva, E. V., Gray, H. B. and Winkler, J. R., Snapshots of cytochrome *c* folding. *Proc. Natl. Acad. Sci. USA*, 2005, **102**, 18397–18402.
38. Sinha, K. K. and Udgaonkar, J. B., Dependence of the size of the initially collapsed form during the refolding of barstar on denaturant concentration: evidence for a continuous transition. *J. Mol. Biol.*, 2005, **353**, 704–718.
39. Sadqi, M., Fushman, D. and Munoz, V., Atom-by-atom analysis of global downhill protein folding. *Nature*, 2006, **442**, 317–321.
40. Sinha, K. K. and Udgaonkar, J. B., Dissecting the non-specific and specific components of the initial folding reaction of barstar by multi-site FRET measurements. *J. Mol. Biol.*, 2007, **370**, 385–405.
41. Karplus, M., Aspects of protein reaction dynamics: deviations from simple behaviour. *J. Phys. Chem.*, 2000, **B104**, 11–27.
42. Eaton, W. A., Searching for 'downhill scenarios' in protein folding. *Proc. Natl. Acad. Sci. USA*, 1999, **96**, 5897–5899.
43. Munoz, V., Conformational dynamics and ensembles in protein folding. *Annu. Rev. Biophys. Biomol. Struct.*, 2007, **36**, 395–412.
44. Yang, W. Y. and Gruebele, M., Folding at the speed limit. *Nature*, 2003, **423**, 193–197.
45. Kubelka, J., Chiu, T. K., Davies, D. R., Eaton, W. A. and Hofrichter, J., Sub-microsecond protein folding. *J. Mol. Biol.*, 2006, **359**, 546–553.
46. Ghosh, K., Ozkan, S. B. and Dill, K. A., The ultimate speed limit to protein folding is conformational searching. *J. Am. Chem. Soc.*, 2007, **129**, 11920–11927.
47. Arora, P., Oas, T. G. and Myers, J. K., Fast and faster: a designed variant of the B-domain of protein A folds in 3 microsec. *Protein Sci.*, 2004, **13**, 847–853.
48. Ferguson, N. *et al.*, Ultra-fast barrier-limited folding in the peripheral subunit-binding domain family. *J. Mol. Biol.*, 2005, **353**, 427–446.

49. Religa, T. L., Johnson, C. M., Vu, D. M., Brewer, S. H., Dyer, R. B. and Fersht, A. R., The helix-turn-helix motif as an ultrafast independently folding domain: the pathway of folding of Engrailed homeodomain. *Proc. Natl. Acad. Sci. USA*, 2007, **104**, 9272–9277.
50. Li, P., Oliva, F. Y., Naganathan, A. N. and Munoz, V., Dynamics of one-state protein folding. *Proc. Natl. Acad. Sci. USA*, 2009, **106**, 103–108.
51. Leeson, D. T., Gai, F., Rodriguez, H. M., Gregoret, L. M. and Dyer, R. B., Protein folding and unfolding on a complex energy landscape. *Proc. Natl. Acad. Sci. USA*, 2000, **97**, 2527–2532.
52. Gruebele, M., Downhill protein folding: evolution meets physics. *C. R. Biol.*, 2005, **328**, 701–712.
53. Ma, H. and Gruebele, M., Kinetics are probe-dependent during downhill folding of an engineered lambda6-85 protein. *Proc. Natl. Acad. Sci. USA*, 2005, **102**, 2283–2287.
54. Gruebele, M., Comments on probe-dependent and independent relaxation kinetics: unreliable signatures of downhill folding. *Proteins*, 2007, **70**, 1099–1102.
55. Hagen, S. J., Probe-dependent and nonexponential relaxation kinetics: unreliable signatures of downhill protein folding. *Proteins*, 2007, **68**, 205–217.
56. Yang, W. Y. and Gruebele, M., Folding lambda-repressor at its speed limit. *Biophys. J.*, 2004, **87**, 596–608.
57. Cho, S. S., Weinkam, P. and Wolynes, P. G., Origins of barriers and barrierless folding in BBL. *Proc. Natl. Acad. Sci. USA*, 2008, **105**, 118–123.
58. Huang, F., Sato, S., Sharpe, T. D., Ying, L. and Fersht, A. R., Distinguishing between cooperative and unimodal downhill protein folding. *Proc. Natl. Acad. Sci. USA*, 2007, **104**, 123–127.
59. Cellmer, T., Henry, E. R., Kubelka, J., Hofrichter, J. and Eaton, W. A., Relaxation rate for an ultrfast folding protein is independent of chemical denaturant concentration. *J. Am. Chem. Soc.*, 2007, **129**, 14564–14565.
60. Godoy-Ruiz, R., Henry, E. R., Kubelka, J., Hofrichter, J., Munoz, V., Sanchez-Ruiz, J. M. and Eaton, W. A., Estimating free-energy barrier heights for an ultrafast folding protein from calorimetric and kinetic data. *J. Phys. Chem. B*, 2008, **112**, 5938–5949.
61. Fierz, B., Satzger, H., Root, C., Gilch, P., Zinth, W. and Kiefhaber, T., Loop formation in unfolded polypeptide chains on the picoseconds to microseconds time scale. *Proc. Natl. Acad. Sci. USA*, 2007, **104**, 2163–2168.
62. Lapidus, L. J., Eaton, W. A. and Hofrichter, J., Measuring the rate of intramolecular contact formation in polypeptides. *Proc. Natl. Acad. Sci. USA*, 2000, **97**, 7220–7225.
63. Hagen, S. J., Carswell, C. W. and Sjolander, E. M., Rate of intrachain contact formation in an unfolded protein: temperature and denaturant effects. *J. Mol. Biol.*, 2001, **305**, 1161–1171.
64. Singh, V. R. and Lapidus, L. J., The intrinsic stiffness of polyglutamine peptides. *J. Phys. Chem. B*, 2008, **112**, 13172–13176.
65. Munoz, V., Thompson, P. A., Hofrichter, J. and Eaton, W. A., Folding dynamics and mechanism of beta-hairpin formation. *Nature*, 1997, **390**, 196–199.
66. Lednev, I. K., Karnoup, A. S., Sparrow, M. C. and Asher, S. A., Alpha-helix peptide folding and unfolding activation barriers: a nanosecond UV resonance Raman study. *J. Am. Chem. Soc.*, 1999, **121**, 8074–8086.
67. Huang, C. Y., Getahun, Z., Zhu, Y., Klemke, J. W., DeGrado, W. F. and Gai, F., Helix formation via conformation diffusion search. *Proc. Natl. Acad. Sci. USA*, 2002, **99**, 2788–2793.
68. Kimura, T., Takahashi, S., Akiyama, S., Uzawa, T., Ishimori, K. and Morishima, I., Direct observation of the multistep helix formation of poly-L-glutamic acids. *J. Am. Chem. Soc.*, 2002, **124**, 11596–11597.
69. Shimada, J. and Shakhnovich, E. I., The ensemble folding kinetics of protein G from an all-atom Monte Carlo simulation. *Proc. Natl. Acad. Sci. USA*, 2002, **99**, 11175–11180.
70. Pradeep, L. and Udgaonkar, J. B., Osmolytes induce structure in an early intermediate on the folding pathway of barstar. *J. Biol. Chem.*, 2004, **279**, 40303–40313.
71. Sridevi, K., Lakshmikanth, G. S., Krishnamoorthy, G. and Udgaonkar, J. B., Increasing stability reduces conformational heterogeneity in a protein folding intermediate ensemble. *J. Mol. Biol.*, 2004, **337**, 699–711.
72. Pletneva, E. V., Gray, H. B. and Winkler, J. R., Many faces of the unfolded state: conformational heterogeneity in denatured yeast cytochrome C. *J. Mol. Biol.*, 2005, **345**, 855–867.
73. Mann, C. J. and Matthews, C. R., Structure and stability of an early folding intermediate of *Escherichia coli* trp aporepressor measured by far-UV stopped-flow circular dichroism and 8-anilino-1-naphthalene sulfonate binding. *Biochemistry*, 1993, **32**, 5282–5290.
74. Eliezer, D., Jennings, P. A., Wright, P. E., Doniach, S., Hodgson, K. O. and Tsuruta, H., The radius of gyration of an apomyoglobin folding intermediate. *Science*, 1995, **270**, 487–488.
75. Chen, L., Wildegger, G., Kiefhaber, T., Hodgson, K. O. and Doniach, S., Kinetics of lysozyme refolding: structural characterization of a non-specifically collapsed state using time-resolved X-ray scattering. *J. Mol. Biol.*, 1998, **276**, 225–237.
76. Krantz, B. A., Mayne, L., Rumbley, J., Englander, S. W. and Sosnick, T. R., Fast and slow intermediate accumulation and the initial barrier mechanism in protein folding. *J. Mol. Biol.*, 2002, **324**, 359–371.
77. Phillips, C. M., Mizutani, Y. and Hochstrasser, R. M., Ultrafast thermally induced unfolding of RNase A. *Proc. Natl. Acad. Sci. USA*, 1995, **92**, 7292–7296.
78. Nolting, B., Temperature-jump induced fast refolding of cold-unfolded protein. *Biochem. Biophys. Res. Commun.*, 1996, **227**, 903–908.
79. Gruebele, M., *Fast Relaxation Methods*, Wiley-VCH: Weiheim, 2005, 1st edn, vol. 1, p. 572.
80. Wong, K. B., Freund, S. M. and Fersht, A. R., Cold denaturation of barstar: ^{1}H, ^{15}N and ^{13}C NMR assignment and characterisation of residual structure. *J. Mol. Biol.*, 1996, **259**, 805–818.
81. Nolting, B., Golbik, R., Soler-Gonzalez, A. S. and Fersht, A. R., Circular dichroism of denatured barstar suggests residual structure. *Biochemistry*, 1997, **36**, 9899–9905.
82. Jones, C. M. *et al.*, Fast events in protein folding initiated by nanosecond laser photolysis. *Proc. Natl. Acad. Sci. USA*, 1993, **90**, 11860–11864.
83. Hagen, S. J., Hofrichter, J., Szabo, A. and Eaton, W. A., Diffusion-limited contact formation in unfolded cytochrome *c*: estimating the maximum rate of protein folding. *Proc. Natl. Acad. Sci. USA*, 1996, **93**, 11615–11617.
84. Saxena, A. M., Udgaonkar, J. B. and Krishnamoorthy, G., Protein dynamics control proton transfer from bulk solvent to protein interior: a case study with a green fluorescent protein. *Protein Sci.*, 2005, **14**, 1787–1799.
85. Regenfuss, P., Clegg, R. M., Fulwyler, M. J., Barrantes, F. J. and Jovin, T. M., Mixing liquids in microseconds. *Rev. Sci. Instrum.*, 1985, **56**, 283–290.
86. Takahashi, S., Yeh, S. R., Das, T. K., Chan, C. K., Gottfried, D. S. and Rousseau, D. L., Folding of cytochrome *c* initiated by submillisecond mixing. *Nature Struct. Biol.*, 1997, **4**, 44–50.
87. Shastry, M. C. and Roder, H., Evidence for barrier-limited protein folding kinetics on the microsecond time scale. *Nature Struct. Biol.*, 1998, **5**, 385–392.
88. Park, S. H., Shastry, M. C. and Roder, H., Folding dynamics of the B1 domain of protein G explored by ultrarapid mixing. *Nature Struct. Biol.*, 1999, **6**, 943–947.
89. Lapidus, L. J., Yao, S., McGarrity, K. S., Hertzog, D. E., Tubman, E. and Bakajin, O., Protein hydrophobic collapse and early folding steps observed in a microfluidic mixer. *Biophys. J.*, 2007, **93**, 218–224.

90. Sinha, K. K. and Udgaonkar, J. B., Barrierless evolution of structure during the sub-ms refolding reaction of a small protein. *Proc. Natl. Acad. Sci. USA*, 2008, **105**, 7998–8003.
91. Matsumoto, S., Yane, A., Nakashima, S., Hashida, M., Fujita, M., Goto, Y. and Takahashi, S., A rapid flow mixer with 11-ms mixing time microfabricated by a pulsed-laser ablation technique: observation of a barrier-limited collapse in cytochrome c folding. *J. Am. Chem. Soc.*, 2007, **129**, 3840–3841.
92. Sandstrom, J., *Dynamic NMR Spectroscopy*, Academic, New York, 1982.
93. Huang, G. S. and Oas, T. G., Structure and stability of monomeric lambda repressor: NMR evidence for two-state folding. *Biochemistry*, 1995, **34**, 3884–3892.
94. Juneja, J. and Udgaonkar, J. B., NMR studies of protein folding. *Curr. Sci.*, 2003, **84**, 157–172.
95. Schuler, B., Lipman, E. A. and Eaton, W. A., Probing the free-energy surface for protein folding with single-molecule fluorescence spectroscopy. *Nature*, 2002, **419**, 743–747.
96. Lipman, E. A., Schuler, B., Bakajin, O. and Eaton, W. A., Single-molecule measurement of protein folding kinetics. *Science*, 2003, **301**, 1233–1235.
97. Hamadani, K. M. and Weiss, S., Non-equilibrium single molecule protein folding in a co-axial mixer. *Biophys. J.*, 2008, **95**, 352–365.
98. Sherman, E. and Haran, G., Coil-globule transition in the denatured state of a small protein. *Proc. Natl. Acad. Sci. USA*, 2006, **103**, 11539–11543.
99. Chen, H., Rhoades, E., Butler, J. S., Loh, S. N. and Webb, W. W., Dynamics of equilibrium structural fluctuations of apomyoglobin measured by fluorescence correlation spectroscopy. *Proc. Natl. Acad. Sci. USA*, 2007, **104**, 10459–10464.
100. Fernandez, J. M. and Li, H., Force-clamp spectroscopy monitors the folding trajectory of a single protein. *Science*, 2004, **303**, 1674–1678.
101. Cecconi, C., Shank, E. A., Bustamante, C. and Marqusee, S., Direct observation of the three-state folding of a single protein molecule. *Science*, 2005, **309**, 2057–2060.
102. Walther, K. A., Grater, F., Dougan, L., Badilla, C. L., Berne, B. J. and Fernandez, J. M., Signatures of hydrophobic collapse in extended proteins captured with force spectroscopy. *Proc. Natl. Acad. Sci. USA*, 2007, **104**, 7916–7921.
103. Cohen, A. E. and Moerner, W. E., Suppressing Brownian motion of individual biomolecules in solution. *Proc. Natl. Acad. Sci. USA*, 2006, **103**, 4362–4365.
104. Kinoshita, M., Kamagata, K., Maeda, A., Goto, Y., Komatsuzaki, T. and Takahashi, S., Development of a technique for the investigation of folding dynamics of single proteins for extended time periods. *Proc. Natl. Acad. Sci. USA*, 2007, **104**, 10453–10458.
105. Dill, K. A. and Chan, H. S., From Levinthal to pathways to funnels. *Nature Struct. Biol.*, 1997, **4**, 10–19.
106. Schonbrun, J. and Dill, K. A., Fast protein folding kinetics. *Proc. Natl. Acad. Sci. USA*, 2003, **100**, 12678–12682.
107. Dill, K. A. and Shortle, D., Denatured states of proteins. *Annu. Rev. Biochem.*, 1991, **60**, 795–825.
108. Lakshmikanth, G. S., Sridevi, K., Krishnamoorthy, G. and Udgaonkar, J. B., Structure is lost incrementally during the unfolding of barstar. *Nature Struct. Biol.*, 2001, **8**, 799–804.
109. Navon, A., Ittah, V., Landsman, P., Scheraga, H. A. and Haas, E., Distributions of intramolecular distances in the reduced and denatured states of bovine pancreatic ribonuclease A. Folding initiation structures in the C-terminal portions of the reduced protein. *Biochemistry*, 2001, **40**, 105–118.
110. Tanford, C., Kawahara, K. and Lapanje, S., Proteins in 6-M guanidine hydrochloride. Demonstration of random coil behaviour. *J. Biol. Chem.*, 1966, **241**, 1921–1923.
111. Kohn, J. E. *et al.*, Random-coil behaviour and the dimensions of chemically unfolded proteins. *Proc. Natl. Acad. Sci. USA*, 2004, **101**, 12491–12496.
112. Pappu, R. V., Srinivasan, R. and Rose, G. D., The Flory isolated-pair hypothesis is not valid for polypeptide chains: implications for protein folding. *Proc. Natl. Acad. Sci. USA*, 2000, **97**, 12565–12570.
113. Meekhof, A. E. and Freund, S. M., Probing residual structure and backbone dynamics on the milli- to picosecond timescale in a urea-denatured fibronectin type III domain. *J. Mol. Biol.*, 1999, **286**, 579–592.
114. Bhavesh, N. S., Juneja, J., Udgaonkar, J. B. and Hosur, R. V., Native and nonnative conformational preferences in the urea-unfolded state of barstar. *Protein Sci.*, 2004, **13**, 3085–3091.
115. Neri, D., Billeter, M., Wider, G. and Wuthrich, K., NMR determination of residual structure in a urea-denatured protein, the 434-repressor. *Science*, 1992, **257**, 1559–1563.
116. Logan, T. M., Theriault, Y. and Fesik, S. W., Structural characterization of the FK506 binding protein unfolded in urea and guanidine hydrochloride. *J. Mol. Biol.*, 1994, **236**, 637–648.
117. Hodsdon, M. E. and Frieden, C., Intestinal fatty acid binding protein: the folding mechanism as determined by NMR studies. *Biochemistry*, 2001, **40**, 732–742.
118. Shortle, D. and Ackerman, M. S., Persistence of native-like topology in a denatured protein in 8 M urea. *Science*, 2001, **293**, 487–489.
119. Schwarzinger, S., Wright, P. E. and Dyson, H. J., Molecular hinges in protein folding: the urea-denatured state of apomyoglobin. *Biochemistry*, 2002, **41**, 12681–12686.
120. McCarney, E. R., Werner, J. H., Bernstein, S. L., Ruczinski, I., Makarov, D. E., Goodwin, P. M. and Plaxco, K. W., Site-specific dimensions across a highly denatured protein; a single molecule study. *J. Mol. Biol.*, 2005, **352**, 672–682.
121. Saxena, A. M., Udgaonkar, J. B. and Krishnamoorthy, G., Characterization of intra-molecular distances and site-specific dynamics in chemically unfolded barstar: evidence for denaturant-dependent non-random structure. *J. Mol. Biol.*, 2006, **359**, 174–189.
122. Fitzkee, N. C. and Rose, G. D., Reassessing random-coil statistics in unfolded proteins. *Proc. Natl. Acad. Sci. USA*, 2004, **101**, 12497–12502.
123. Klein-Seetharaman, J. *et al.*, Long-range interactions within a nonnative protein. *Science*, 2002, **295**, 17119–17122.
124. Mok, K. H., Kuhn, L. T., Goez, M., Day, I. J., Lin, J. C., Andersen, N. H. and Hore, P. J., A pre-existing hydrophobic collapse in the unfolded state of an ultrafast folding protein. *Nature*, 2007, **447**, 106–109.
125. Yao, J., Chung, J., Eliezer, D., Wright, P. E. and Dyson, H. J., NMR structural and dynamic characterization of the acid-unfolded state of apomyoglobin provides insights into the early events in protein folding. *Biochemistry*, 2001, **40**, 3561–3571.
126. Wong, K. B., Clarke, J., Bond, C. J., Neira, J. L., Freund, S. M., Fersht, A. R. and Daggett, V., Towards a complete description of the structural and dynamic properties of the denatured state of barnase and the role of residual structure in folding. *J. Mol. Biol.*, 2000, **296**, 1257–1282.
127. Mohana-Borges, R., Goto, N. K., Kroon, G. J., Dyson, H. J. and Wright, P. E., Structural characterization of unfolded states of apomyoglobin using residual dipolar couplings. *J. Mol. Biol.*, 2004, **340**, 1131–1142.
128. Dyson, H. J. and Wright, P. E., Insights into protein folding from NMR. *Annu. Rev. Phys. Chem.*, 1996, **47**, 369–395.
129. Kazmirski, S. L., Wong, K. B., Freund, S. M., Tan, Y. J., Fersht, A. R. and Daggett, V., Protein folding from a highly disordered denatured state: the folding pathway of chymotrypsin inhibitor 2 at atomic resolution. *Proc. Natl. Acad. Sci. USA*, 2001, **98**, 4349–4354.

130. Kuzmenkina, E. V., Heyes, C. D. and Nienhaus, G. U., Single-molecule Forster resonance energy transfer study of protein dynamics under denaturing conditions. *Proc. Natl. Acad. Sci. USA*, 2005, **102**, 15471–15476.
131. Chattopadhyay, K., Saffarian, S., Elson, E. L. and Frieden, C., Measurement of microsecond dynamic motion in the intestinal fatty acid binding protein by using fluorescence correlation spectroscopy. *Proc. Natl. Acad. Sci. USA*, 2002, **99**, 14171–14176.
132. Nettels, D., Gopich, I. V., Hoffmann, A. and Schuler, B., Ultrafast dynamics of protein collapse from single-molecule photon statistics. *Proc. Natl. Acad. Sci. USA*, 2007, **104**, 2655–2660.
133. Morar, A. S., Olteanu, A., Young, G. B. and Pielak, G. J., Solvent-induced collapse of alpha-synuclein and acid-denatured cytochrome *c*. *Protein Sci.*, 2001, **10**, 2195–2199.
134. Bhuyan, A. K. and Udgaonkar, J. B., *Relevance of Burst Phase Changes in Optical Signals of Polypeptides During Protein Folding*, Universities Press, Hyderabad, 1999.
135. Sosnick, T. R., Mayne, L. and Englander, S. W., Molecular collapse: the rate-limiting step in two-state cytochrome *c* folding. *Proteins*, 1996, **24**, 413–426.
136. Sosnick, T. R., Shtilerman, M. D., Mayne, L. and Englander, S. W., Ultrafast signals in protein folding and the polypeptide contracted state. *Proc. Natl. Acad. Sci. USA*, 1997, **94**, 8545–8550.
137. Qi, P. X., Sosnick, T. R. and Englander, S. W., The burst phase in ribonuclease A folding and solvent dependence of the unfolded state. *Nature Struct. Biol.*, 1998, **5**, 882–884.
138. Rami, B. R. and Udgaonkar, J. B., Mechanism of formation of a productive molten globule form of barstar. *Biochemistry*, 2002, **41**, 1710–1716.
139. Mayor, U., Grossmann, J. G., Foster, N. W., Freund, S. M. and Fersht, A. R., The denatured state of Engrailed Homeodomain under denaturing and native conditions. *J. Mol. Biol.*, 2003, **333**, 977–991.
140. Religa, T. L., Markson, J. S., Mayor, U., Freund, S. M. and Fersht, A. R., Solution structure of a protein denatured state and folding intermediate. *Nature*, 2005, **437**, 1053–1056.
141. Chugha, P. and Oas, T. G., Backbone dynamics of the monomeric l repressor denatured state ensemble under nondenaturing conditions. *Biochemistry*, 2007, **46**, 1141–1151.
142. Sali, A., Shakhnovich, E. and Karplus, M., How does a protein fold? *Nature*, 1994, **369**, 248–251.
143. Dobson, C. M., Sali, A. and Karplus, M., Protein folding: A perspective from theory and experiment. *Angew. Chem. Int. Ed.*, 1998, **37**, 868–893.
144. Elove, G. A., Chaffotte, A. F., Roder, H. and Goldberg, M. E., Early steps in cytochrome *c* folding probed by time-resolved circular dichroism and fluorescence spectroscopy. *Biochemistry*, 1992, **31**, 6876–6883.
145. Jennings, P. A. and Wright, P. E., Formation of a molten globule intermediate early in the kinetic folding pathway of apomyoglobin. *Science*, 1993, **262**, 892–896.
146. Houry, W. A., Rothwarf, D. M. and Scheraga, H. A., Circular dichroism evidence for the presence of burst-phase intermediates on the conformational folding pathway of ribonuclease A. *Biochemistry*, 1996, **35**, 10125–10133.
147. Khorasanizadeh, S., Peters, I. D. and Roder, H., Evidence for a three-state model of protein folding from kinetic analysis of ubiquitin variants with altered core residues. *Nature Struct. Biol.*, 1996, **3**, 193–205.
148. Park, S. H., O'Neil, K. T. and Roder, H., An early intermediate in the folding reaction of the B1 domain of protein G contains a native-like core. *Biochemistry*, 1997, **36**, 14277–14283.
149. Raschke, T. M. and Marqusee, S., The kinetic folding intermediate of ribonuclease H resembles the acid molten globule and partially unfolded molecules detected under native conditions. *Nature Struct. Biol.*, 1997, **4**, 298–304.
150. Roder, H., Colon, W., Kinetic role of early intermediates in protein folding. *Curr. Opin. Struct. Biol.*, 1997, **7**, 15–28.
151. Ptitsyn, O. B., Structures of folding intermediates. *Curr. Opin. Struct. Biol.*, 1995, **5**, 74–78.
152. Kuwajima, K., The molten globule state of alpha-lactalbumin. *FASEB J.*, 1996, **10**, 102–109.
153. Arai, M. and Kuwajima, K., Rapid formation of a molten globule intermediate in refolding of alpha-lactalbumin. *Fold. Des.*, 1996, **1**, 275–287.
154. Engelhard, M. and Evans, P. A., Kinetics of interaction of partially folded proteins with a hydrophobic dye: evidence that molten globule character is maximal in early folding intermediates. *Protein Sci.*, 1995, **4**, 1553–1562.
155. Colon, W., Elove, G. A., Wakem, L. P., Sherman, F. and Roder, H., Side chain packing of the N- and C-terminal helices plays a critical role in the kinetics of cytochrome *c* folding. *Biochemistry*, 1996, **35**, 5538–5549.
156. Sridevi, K., Juneja, J., Bhuyan, A. K., Krishnamoorthy, G. and Udgaonkar, J. B., The slow folding reaction of barstar: the core tryptophan region attains tight packing before substantial secondary and tertiary structure formation and final compaction of the polypeptide chain. *J. Mol. Biol.*, 2000, **302**, 479–495.
157. Rami, B. R., Krishnamoorthy, G. and Udgaonkar, J. B., Dynamics of the core tryptophan during the formation of a productive molten globule intermediate of barstar. *Biochemistry*, 2003, **42**, 7986–8000.
158. Pradeep, L. and Udgaonkar, J. B., Differential salt-induced stabilization of structure in the initial folding intermediate ensemble of barstar. *J. Mol. Biol.*, 2002, **324**, 331–347.
159. Jha, S. K. and Udgaonkar, J. B., Exploring the cooperativity of the fast folding reaction of a small protein using pulsed thiol labeling and mass spectrometry. *J. Biol. Chem.*, 2007, **282**, 37479–37491.
160. Jacob, J., Krantz, B., Dothager, R. S., Thiyagarajan, P. and Sosnick, T. R., Early collapse is not an obligate step in protein folding. *J. Mol. Biol.*, 2004, **338**, 369–382.
161. Magg, C. and Schmid, F. X., Rapid collapse precedes the fast two-state folding of the cold shock protein. *J. Mol. Biol.*, 2004, **335**, 1309–1323.
162. Magg, C., Kubelka, J., Holtermann, G., Haas, E. and Schmid, F. X., Specificity of the initial collapse in the folding of the cold shock protein. *J. Mol. Biol.*, 2006, **360**, 1067–1080.
163. Parker, M. J. and Marqusee, S., The cooperativity of burst phase reactions explored. *J. Mol. Biol.*, 1999, **293**, 1195–1210.
164. Georgescu, R. E., Li, J. H., Goldberg, M. E., Tasayco, M. L. and Chaffotte, A. F., Proline isomerization-independent accumulation of an early intermediate and heterogeneity of the folding pathways of a mixed alpha/beta protein, *Escherichia coli* thioredoxin. *Biochemistry*, 1998, **37**, 10286–10297.
165. Scholtz, J. M., Marqusee, S., Baldwin, R. L., York, E. J., Stewart, J. M., Santoro, M. and Bolen, D. W., Calorimetric determination of the enthalpy change for the alpha-helix to coil transition of an alanine peptide in water. *Proc. Natl. Acad. Sci. USA*, 1991, **88**, 2854–2858.
166. Williams, S., Causgrove, T. P., Gilmanshin, R., Fang, K. S., Callender, R. H., Woodruff, W. H. and Dyer, R. B., Fast events in protein folding: helix melting and formation in a small peptide. *Biochemistry*, 1996, **35**, 691–697.
167. Schulman, B. A., Kim, P. S., Dobson, C. M. and Redfield, C., A residue-specific NMR view of the non-cooperative unfolding of a molten globule. *Nature Struct. Biol.*, 1997, **4**, 630–634.
168. Hoffmann, A. *et al.*, Mapping protein collapse with single-molecule fluorescence and kinetic synchrotron radiation circular dichroism spectroscopy. *Proc. Natl. Acad. Sci. USA*, 2007, **104**, 105–110.
169. Roder, H., Maki, K., Cheng, H. and Shastry, M. C., Rapid mixing methods for exploring the kinetics of protein folding. *Methods*, 2004, **34**, 15–27.

170. Capaldi, A. P., Shastry, M. C., Kleanthous, C., Roder, H. and Radford, S. E., Ultrarapid mixing experiments reveal that Im7 folds via an on-pathway intermediate. *Nature Struct. Biol.*, 2001, **8**, 68–72.
171. Teilum, K., Maki, K., Kragelund, B. B., Poulsen, F. M. and Roder, H., Early kinetic intermediate in the folding of acyl-CoA binding protein detected by fluorescence labeling and ultrarapid mixing. *Proc. Natl. Acad. Sci. USA*, 2002, **99**, 9807–9812.
172. Maki, K., Cheng, H., Dolgikh, D. A. and Roder, H., Folding kinetics of staphylococcal nuclease studied by tryptophan engineering and rapid mixing methods. *J. Mol. Biol.*, 2007, **368**, 244–255.
173. Hagen, S. J., Exponential decay kinetics in 'downhill' protein folding. *Proteins*, 2003, **50**, 1–4.
174. Knott, M. and Chan, H. S., Criteria for downhill protein folding: calorimetry, chevron plot, kinetic relaxation, and single-molecule radius of gyration in chain models with subdued degrees of cooperativity. *Proteins*, 2006, **65**, 373–391.
175. Sabelko, J., Ervin, J. and Gruebele, M., Observation of strange kinetics in protein folding. *Proc. Natl. Acad. Sci. USA*, 1999, **96**, 6031–6036.
176. Saigo, S. and Shibayama, N., Highly nonexponential kinetics in the early-phase refolding of proteins at low temperatures. *Biochemistry*, 2003, **42**, 9669–9676.
177. Ye, X., Ionascu, D., Gruia, F., Yu, A., Benabbas, A. and Champion, P. M., Temperature-dependent heme kinetics with nonexponential binding and barrier relaxation in the absence of protein conformational substates. *Proc. Natl. Acad. Sci. USA*, 2007, **104**, 14682–14687.
178. Hagen, S. J. and Eaton, W. A., Two-state expansion and collapse of a polypeptide. *J. Mol. Biol.*, 2000, **301**, 1019–1027.
179. Qiu, L., Zachariah, C. and Hagen, S. J., Fast chain contraction during protein folding: 'foldability' and collapse dynamics. *Phys. Rev. Lett.*, 2003, **90**, 168103.
180. Uzawa, T., Akiyama, S., Kimura, T., Takahashi, S., Ishimori, K., Morishima, I. and Fujisawa, T., Collapse and search dynamics of apomyoglobin folding revealed by submillisecond observations of alpha-helical content and compactness. *Proc. Natl. Acad. Sci. USA*, 2004, **101**, 1171–1176.
181. Akiyama, S., Takahashi, S., Ishimori, K. and Morishima, I., Stepwise formation of alpha-helices during cytochrome *c* folding. *Nature Struct. Biol.*, 2000, **7**, 514–520.
182. Bilsel, O. and Matthews, C. R., Molecular dimensions and their distributions in early folding intermediates. *Curr. Opin. Struct. Biol.*, 2006, **16**, 86–93.
183. Lyubovitsky, J. G., Gray, H. B. and Winkler, J. R., Mapping the cytochrome *c* folding landscape. *J. Am. Chem. Soc.*, 2002, **124**, 5481–5485.
184. Kimura, T., Lee, J. C., Gray, H. B. and Winkler, J. R., Site-specific collapse dynamics guide the formation of the cytochrome *c*′ four-helix bundle. *Proc. Natl. Acad. Sci. USA*, 2007, **104**, 117–122.
185. Zwanzig, R., Rate processes with dynamical disorder. *Acc. Chem. Res.*, 1990, **23**, 148–152.
186. Sanchez, I. E. and Kiefhaber, T., Evidence for sequential barriers and obligatory intermediates in apparent two-state protein folding. *J. Mol. Biol.*, 2003, **325**, 367–376.
187. Udgaonkar, J. B. and Baldwin, R. L., Early folding intermediate of ribonuclease A. *Proc. Natl. Acad. Sci. USA*, 1990, **87**, 8197–8201.
188. Houry, W. A. and Scheraga, H. A., Structure of a hydrophobically collapsed intermediate on the conformational folding pathway of ribonuclease A probed by hydrogen-deuterium exchange. *Biochemistry*, 1996, **35**, 11734–11746.
189. Clarke, J. and Fersht, A. R., An evaluation of the use of hydrogen exchange at equilibrium to probe intermediates on the protein folding pathway. *Fold. Des.*, 1996, **1**, 243–254.
190. Kimura, T., Akiyama, S., Uzawa, T., Ishimori, K., Morishima, I., Fujisawa, T. and Takahashi, S., Specifically collapsed intermediate in the early stage of the folding of ribonuclease A. *J. Mol. Biol.*, 2005, **350**, 349–362.
191. Eliezer, D., Yao, J., Dyson, H. J. and Wright, P. E., Structural and dynamic characterization of partially folded states of apomyoglobin and implications for protein folding. *Nature Struct. Biol.*, 1998, **5**, 148–155.
192. Balbach, J., Forge, V., Lau, W. S., Jones, J. A., van Nuland, N. A. and Dobson, C. M., Detection of residue contacts in a protein folding intermediate. *Proc. Natl. Acad. Sci. USA*, 1997, **94**, 7182–7185.
193. Chan, H. S. and Dill, K. A., Origins of structure in globular proteins. *Proc. Natl. Acad. Sci. USA*, 1990, **87**, 6388–6392.
194. Hummer, G., Garcia, A. E. and Garde, S., Helics nucleation kinetics from molecular simulations in explicit solvent. *Proteins*, 2001, **42**, 77–84.
195. Kimura, T. *et al.*, Specific collapse followed by slow hydrogen-bond formation of beta-sheet in the folding of single-chain monellin. *Proc. Natl. Acad. Sci. USA*, 2005, **102**, 2748–2753.
196. Kimura, T. *et al.*, Dehydration of main-chain amides in the final folding step of single-chain monellin revealed by time-resolved infrared spectroscopy. *Proc. Natl. Acad. Sci. USA*, 2008, **105**, 13391–13396.
197. Ellison, P. A. and Cavagnero, S., Role of unfolded state heterogeneity and en-route ruggedness in protein folding kinetics. *Protein Sci.*, 2006, **15**, 564–582.
198. Duan, Y. and Kollman, P. A., Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution. *Science*, 1998, **282**, 740–744.
199. Chavez, L. L., Gosavi, S., Jennings, P. A. and Onuchic, J. N., Multiple routes lead to the native state in the energy landscape of the beta-trefoil family. *Proc. Natl. Acad. Sci. USA*, 2006, **103**, 10254–10258.
200. Radford, S. E., Dobson, C. M. and Evans, P. A., The folding of hen lysozyme involves partially structured intermediates and multiple pathways. *Nature*, 1992, **358**, 302–307.
201. Viguera, A. R., Serrano, L. and Wilmanns, M., Different folding transition states may result in the same native structure. *Nature Struct. Biol.*, 1996, **3**, 874–880.
202. Lindberg, M. O., Tangrot, J. and Oliveberg, M., Complete change of the protein folding transition state upon circular permutation. *Nature Struct. Biol.*, 2002, **9**, 818–822.
203. Wu, Y., Kondrashkina, E., Kayatekin, C., Matthews, C. R. and Bilsel, O., Microsecond acquisition of heterogeneous structure in the folding of a TIM barrel protein. *Proc. Natl. Acad. Sci. USA*, 2008, **105**, 13367–13372.
204. Abel, C. J., Goldbeck, R. A., Latypov, R. F., Roder, H. and Kliger, D. S., Conformational equilibration time of unfolded protein chains and the folding speed limit. *Biochemistry*, 2007, **46**, 4090–4099.
205. Fierz, B. and Kiefhaber, T., End-to-end vs interior loop formation kinetics in unfolded polypeptide chains. *J. Am. Chem. Soc.*, 2007, **129**, 672–679.
206. Perl, D., Welker, C., Schindler, T., Schroder, K., Marahiel, M. A., Jaenicke, R. and Schmid, F. X., Conservation of rapid two-state folding in mesophilic, thermophilic and hyperthermophilic cold shock proteins. *Nature Struct. Biol.*, 1998, **5**, 229–235.
207. Schindler, T., Herrler, M., Marahiel, M. A. and Schmid, F. X., Extremely rapid protein folding in the absence of intermediates. *Nature Struct. Biol.*, 1995, **2**, 663–673.
208. Schindler, T. and Schmid, F. X., Thermodynamic properties of an extremely rapid protein folding reaction. *Biochemistry*, 1996, **35**, 16833–16842.
209. Mayor, U. *et al.*, The complete folding pathway of a protein from nanoseconds to microseconds. *Nature*, 2003, **421**, 863–867.

210. Qiu, L., Pabit, S. A., Roitberg, A. E. and Hagen, S. J., Smaller and faster: the 20-residue Trp-cage protein folds in 4 micros. *J. Am. Chem. Soc.*, 2002, **124**, 12952–12953.
211. Chavez, L. L., Onuchic, J. N. and Clementi, C., Quantifying the roughness on the free energy landscape: entropic bottlenecks and protein folding rates. *J. Am. Chem. Soc.*, 2004, **126**, 8426–8432.
212. Plaxco, K. W., Simons, K. T. and Baker, D., Contact order, transition state placement and the refolding rates of single domain proteins. *J. Mol. Biol.*, 1998, **277**, 985–994.
213. Baker, D., A surprising simplicity to protein folding. *Nature*, 2000, **405**, 39–42.
214. Gillespie, B. and Plaxco, K. W., Using protein folding rates to test protein folding theories. *Annu. Rev. Biochem.*, 2004, **73**, 837–859.
215. Kamagata, K., Arai, M. and Kuwajima, K., Unification of the folding mechanisms of non-two-state and two-state proteins. *J. Mol. Biol.*, 2004, **339**, 951–965.
216. Kamagata, K. and Kuwajima, K., Surprisingly high correlation between early and late stages in non-two-state protein folding. *J. Mol. Biol.*, 2006, **357**, 1647–1654.
217. Kuwata, K., Shastry, R., Cheng, H., Hoshino, M., Batt, C. A., Goto, Y. and Roder, H., Structural and kinetic characterization of early folding events in beta-lactoglobulin. *Nature Struct. Biol.*, 2001, **8**, 151–155.
218. Uzawa, T. *et al.*, 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.
219. Ballew, R. M., Sabelko, J. and Gruebele, M., Direct observation of fast protein folding: the initial collapse of apomyoglobin. *Proc. Natl. Acad. Sci. USA*, 1996, **93**, 5759–5764.
220. Eaton, W. A., Munoz, V., Hagen, S. J., Jas, G. S., Lapidus, L. J., Henry, E. R. and Hofrichter, J., Fast kinetics and mechanisms in protein folding. *Annu. Rev. Biophys. Biomol. Struct.*, 2000, **29**, 327–359.
221. Chan, H. S. and Dill, K. A., Protein folding in the landscape perspective: chevron plots and non-Arrhenius kinetics. *Proteins*, 1998, **30**, 2–33.

ACKNOWLEDGEMENTS. We thank members of our laboratory for discussions. Work in our laboratory is funded by the Tata Institute of Fundamental Research as well as by the Department of Biotechnology and the Department of Science and Technology, Government of India.

Received 30 May 2008; revised accepted 9 March 2009