Archives of Biochemistry and Biophysics 531 (2013) 24–33

Contents lists available at SciVerse ScienceDirect

Archives of Biochemistry and Biophysics

journal homepage: www.elsevier.com/locate/yabbi

Review

Polypeptide chain collapse and protein folding

Jayant B. Udgaonkar *

National Centre for Biological Sciences, Tata Institute of Fundamental Research, Bangalore 560065, India

ARTICLE INFO

Article history:
Available online 19 October 2012

Keywords:
Unfolded state dynamics
Polypeptide chain collapse
Molten globule
Folding mechanisms
Hydrophobic collapse

ABSTRACT

Polypeptide chain collapse is an integral component of a protein folding reaction. In this review, experimental characterization of the interplay of polypeptide chain collapse, secondary structure formation, consolidation of the hydrophobic core and the development of tertiary interactions, is scrutinized. In particular, the polypeptide chain collapse reaction is examined in the context of the three phenomenological models of protein folding – the hydrophobic collapse model, the framework model and the nucleation condensation model – which describe different ways by which polypeptide chains are able to fold in biologically relevant time-scales.

Introduction

In the unfolded state, a polypeptide chain appears to be describable as a statistical random coil. The radius of gyration (R_g) of an unfolded polypeptide chain that is N residues long, scales as $N^{0.6}$ [1] similar to the scaling ($N^{0.6}$) expected for a homopolymer in good solvent [2]. The low density of a typical random coil polypeptide chain makes it improbable that distant segments of the chain will make contact with each other. In contrast, the folded state of a protein is solid-like in its internal packing and hence, compactness, with its R_g scaling as $N^{0.33}$ [3], similar to the scaling ($N^{0.33}$) expected for a homopolymer in poor solvent [2]. A distinctive physical feature of any protein folding reaction is the greater than 3-fold decrease in volume that accompanies it. There are several unresolved questions about the role of this change in a physical property, size, in the induction of non-covalent chemical structure in the protein. What drives the collapse reaction: side-chain (hydrophobic) interactions or backbone (hydrogen bonding) interactions? Does the polypeptide chain respond like a homopolymer or a random heteropolymer upon a change in solvent conditions from good to bad, or does the evolutionarily selected sequence of the protein chain behave differently? Is polypeptide chain collapse a gradual or a barrier-limited process? What is the interplay between polypeptide chain collapse and secondary structure formation, and between collapse and the formation of the rigid hydrophobic core that is characteristic of globular proteins? Does initial chain collapse accelerate or retard subsequent structure formation? Does polypeptide chain collapse lead to a heterogeneous globule with no specific structure, or directly to a molten globule with some specific structure and a native-like fold? This review examines the results of experimental studies of the collapse reaction in the context of phenomenological models of protein folding.

The unfolded state

Chain contraction and expansion in the unfolded state

Protein folding reactions are best studied experimentally starting from the unfolded state in high denaturant solution. In thermodynamic terms, this state seems to be similar to that of protein unfolded at extremes of pH or high temperature [4–6]. Hence, although protein unfolded by chemical denaturant, pH or temperature may appear different in terms of compactness or residual structure, these differences merely reflect the fact that chain conformations of differing compactness are present in the unfolded state ensemble [7], and that the composition of the unfolded state ensemble may vary gradually when the unfolding conditions are changed. Several studies have indicated that an unfolded protein contracts upon a decrease in denaturant concentration in the range of denaturant concentration in which the protein remains unfolded, in a manner that, while depending on the chemical nature of the denaturant [8], appears to be indicative of the absence of any specific structure. Single molecule measurements have shown that intra-molecular distances within the unfolded states of many proteins contract in a gradual manner upon a decrease in denaturant concentration [9–14], and that polypeptide chain collapse under folding conditions is a diffusive process [15]. Ensemble measurements have also indicated that not only does the unfolded state of barstar swell gradually upon an increase in denaturant concentration, but so does the native state, both through a continuum of forms [16,17]. These studies were equilibrium unfolding studies. More recently, direct kinetic time-resolved, multi-site FRET measurements of the unfolding of monellin have shown that upon

* Fax: +91 80 23636662.
E-mail address: jayant@ncbs.res.in

0003-9861/$ - see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.abb.2012.10.003

penetration of water into the protein, unfolding occurs through a gradual diffusive swelling process whose dynamics can be described by the Rouse model of polymer physics [18].

The unfolded state in refolding conditions

Upon transfer to refolding conditions, before any structure formation occurs during folding, the unfolded state in refolding conditions might be expected to resemble the non-specifically collapsed globule that forms when a homopolymer is transferred from a good to a bad solvent [19,20]. For several proteins, it has been possible to characterize the unfolded states in refolding conditions in equilibrium studies; these studies have shown that this state is indeed collapsed yet devoid of any specific structure [21–25]. It is likely that it is the gain in the entropy of water molecules that are released when hydrophobic residues get at least partially dehydrated, which compensates for the loss in conformational entropy which must accompany the collapse reaction leading to the formation of such a structure-less globule. Moreover, the favorable enthalpy of van der Waals interactions between non-polar residues that associate non-specifically in the structure-less globule, might compensate for the unfavorable enthalpy of dehydration of these residues. Such a collapsed form is expected to be fluid, because entropy would disfavor any unique set of contacts. The unfolded state in refolding conditions resembles in many ways intrinsically disordered proteins whose sequences are less hydrophobic and more charged than proteins that can fold by themselves [26], and whose dimensions also expand upon the addition of denaturant [27,28].

Structure in the unfolded state

Transient fluctuating structures, often stabilized by non-native interactions, whether hydrophobic [9,29,30] or electrostatic [31] in nature, are often found in the unfolded state, and such interactions may make the unfolded form unusually compact. It is generally not known whether the same interactions persist and even develop further when the denaturant is diluted sufficiently for protein folding to commence, but clearly they can potentially act as seeds for the chain collapse reaction that accompanies denaturant dilution [32]. In particular, pre-existing hydrophobic collapse in the unfolded state may be responsible for the ultra-fast folding reactions observed for some proteins [33].

The important factor in the formation of structure in the unfolded state and during initial folding is the dynamics of conformational conversion in the unfolded state. It is commonly assumed that conformation change in the unfolded state occurs in the nanosecond time domain or even faster. Indeed, intra-molecular contacts in short peptides [34] as well in unfolded proteins [15,35] can form in 0.1–1 μs. These intra-molecular contact times have been taken to be indicative of the fastest rates at which proteins can fold, but it must be remembered that folding reactions involve the formation of hundreds or thousands of contacts, and the concerted motions of many or large segments of the polypeptide chain. Indeed, fluorescence correlation spectroscopy measurements [36,37] have shown that large conformational fluctuations in unfolded proteins occur on a 10 to 100-fold slower time scale. Trp-Cys quenching experiments [9,38,39], as well as single molecule measurements [10,40], have also indicated that diffusional motion of the polypeptide chain in the unfolded state may occur on the 100 μs or even slower time scale. Such slow diffusion is perhaps not surprising given that very many attractive intra-molecular interactions would form and have to be broken during the diffusive motion of the chain. The energy landscape for such diffusive motion in the unfolded state would be expected to have very many local minima, and even a larger number of local minima might be expected for the process of diffusive chain collapse, where friction between segments of the chain might slow down the diffusive motion even more [38,41]. If slow diffusive motion is a general property of the unfolded protein chain, it would profoundly affect important aspects of the protein folding reaction, including chain collapse, and directly lead to the operation of multiple folding pathways [42].

Millisecond measurements of collapse and folding

Occurrence of hydrophobic collapse during folding

Until about fifteen years ago, folding studies were restricted to millisecond measurements, and much of our initial understanding, as well as early misconceptions, about the nature of the chain collapse reaction have arisen from these studies [43]. Much of what has been inferred about what happens initially during folding has been from studying the product of the sub-millisecond folding reactions, as it manifests itself at the end of the burst phase (usually of 2–5 ms duration) of stopped-flow mixing. Progress in our understanding has been hampered also by the limitations of the experimental probes that are used to monitor chain collapse, which include small angle X-ray scattering (SAXS)[1] or dynamic light scattering (DLS) measurements, by fluorescence resonance energy transfer (FRET) measurement, or by nuclear magnetic resonance (NMR) spectroscopy. Each probe has its own limitations and strengths, and occasionally, two probes may yield different result. For example, SAXS measurements will occasionally not capture chain collapse when FRET measurements do [44], perhaps because SAXS is strongly biased in favor of larger (uncollapsed) structures, while FRET is biased in favor of smaller (collapsed) structures.

Early studies showed that the product of the millisecond burst phase of the folding of several proteins was competent to bind the hydrophobic dye ANS [45]. This observation was interpreted as indicating that the hydrophobic core had formed in the burst phase, which was still solvent-accessible, but a more conservative interpretation is that only hydrophobic clusters or surfaces large enough to be capable of binding ANS, have formed. But since even these can form only when there is collapse of at least some segments of the polypeptide chain, the observation led to the idea that a hydrophobic collapse occurs initially during folding.

Delineation of chain collapse from structure formation

More direct evidence for an initial hydrophobic collapse reaction came from FRET measurement of intra-molecular distance contraction in the burst phase product of the folding of barstar [46]. In the case of barstar, the use of less stabilizing, yet native-like conditions (1 M GdnHCl or 2.4 M urea) allowed the initial non-specific hydrophobic chain collapse to be delineated from secondary structure formation because the latter was not observed. Only when the folding of barstar was studied under more stabilizing conditions (lower denaturant concentration), was the burst phase product found not only to be collapsed but also to possess secondary structure [41,47]. Such delineation of the collapse reaction from subsequent structure formation by comparing folding under marginally stable conditions to that in more strongly stabilizing conditions, was similarly observed in the case of adenylate cyclase [48] and DHFR [49]. It seems that in strongly stabilizing conditions, the initial non-specific collapse reaction is quickly followed by a structure-forming reaction that completes within a millisecond or so of folding, and which leads to the formation of a partially

[1] *Abbreviations used:* SAXS, small angle X-ray scattering; DLS, dynamic light scattering; FRET, fluorescence resonance energy transfer; NMR, nuclear magnetic resonance; CD, circular dichroism.

structured and collapsed intermediate. In less stabilizing conditions, this structured intermediate is not populated; hence, the burst phase product is observed to be the structure-less globule that is the product of initial non-specific chain collapse.

The initial collapse reaction can be delineated from subsequent structure formation even for the "two-state" folders Csp [50,51], chymotrypsin inhibitor [52] and protein L [38], for which all structure is supposed to form concurrently with collapse in the slower rate-limiting step of folding. Usually, however, when folding appears "two-state", because the initial collapsed forms and intermediates are too unstable to be detected in the folding conditions utilized, it is impossible to demonstrate the sequence of structural events leading up to the transition state of folding, which is the least stable structure en route to the native state.

Nevertheless, for many proteins, the millisecond burst phase products of folding were found not only to be collapsed but also to possess circular dichroism (CD)-detectable secondary structure that was, albeit, unstable in affording only marginal protection against amide hydrogen exchange [53]. Although these studies in the millisecond time domain could not temporally resolve secondary structure formation from chain collapse, they seemed to suggest that collapse and secondary structure formation might occur concurrently during folding. The properties of the millisecond burst phase product qualified it to be a kinetic molten globule. Like the corresponding molten globule populated at equilibrium, the kinetic molten globule was compact with native-like secondary structure but with few tertiary interactions [53–55]. The observation of this similarity, and the observation that the secondary structural contents of equilibrium molten globule forms of proteins are correlated with their hydrodynamic volumes [56], contributed to the general belief that the initial chain collapse reaction occurred concurrently with the formation of specific structure during the formation of kinetic molten globules. But sub-millisecond kinetic measurements have now shown that the kinetic molten globule is not formed in the earliest detectable collapse reaction, but in a subsequent structure-forming reaction, also complete within 1–2 ms of the commencement of folding (see below). Hence, the study of equilibrium molten globules [56] cannot be used to conclude that secondary structure forms during the earliest chain collapse reaction.

It seems clear now that that the millisecond burst phase product is a kinetic molten globule and is not the unfolded state in refolding conditions (see above), as had been suggested [57–61]. It is more likely that it is instead the structure-less globule which precedes the kinetic molten globule and which can now be detected at 5–100 μs of folding (see below), which represents the unfolded state in refolding conditions. Thus, it appears that the kinetic molten globule forms in more than one step from the fully unfolded polypeptide chain at high denaturant concentration.

Cooperativity and specificity of the collapse reactions leading to the formation of the kinetic molten globule

Although millisecond mixing could not resolve the kinetics of the initial collapse reaction, detailed study of the kinetic molten globule that is the first observable product of millisecond mixing, yielded important information about the structural transitions in the first millisecond of folding. The magnitude of the burst phase change in signal, as well as its dependence on denaturant concentration, appeared to be probe-dependent for thioredoxin [62] and barstar [47], indicating that this change occurs in more than one step. For some proteins, the properties of the burst phase product appeared to depend in a sigmoidal manner on denaturant concentration. Although such sigmoidal dependences on denaturant concentration have been traditionally interpreted as indicating cooperative transitions, this need not be so: gradual transitions may also show sigmoidal dependences [55,63]. For other proteins, the dependence on denaturant concentration appeared to be linear, suggesting a gradual transition. The extent of chain collapse, as well as of specific structure formation, was found to depend on the conditions of the folding reaction. The structure of the kinetic molten globule was found to be malleable, with different cosolvents able to tune it in different ways [47,64,65]. It appears that the kinetic molten globule formed at 1 ms of folding is heterogeneous in structure, with different sub-populations differentially populated in different folding conditions [64]; hence the kinetic molten globule might appear to be differently structured in different folding conditions. Indeed, this may be true even for intermediate states that are populated late during folding [66].

Inspection, using multi-site FRET, of how different intra-molecular distances in the kinetic molten globule that is the burst phase product formed during the folding of barstar and the PI3K SH3 domain, contract with a change in solvent conditions, yielded two important results regarding chain collapse: (1) For barstar, each distance was found to contract to a different fractional extent. Moreover, the non-sigmoidal dependences on denaturant concentration of the extents of contraction were different for the eleven intra-molecular distances measured (Fig. 1) [67,68]. This result suggested that different segments of the chain contract independently of other segments, as expected for a polymer chain undergoing gradual diffusive collapse. On the other hand, for the PI3K SH3 domain, three intra-molecular distances in the burst phase product of its folding were measured, and these were found to contract synchronously with a decrease in denaturant concentration [69]. (2) Some intra-molecular distances in the kinetic molten globule were found to contract with the same dependence on denaturant concentration, as was found for the contraction of these distances in the fully unfolded state. Indeed, for these distances, the contraction in the kinetic molten globule was co-linear and continuous with that in the unfolded state (Fig. 2). It appeared therefore that some segments of the sequence contract in the manner expected for a fully unfolded state, upon a change in solvent conditions from good to bad. On the other hand, some intra-molecular distances were found to contract to a greater extent at low denaturant concentration than what was predicted from the dependence of the contraction of the unfolded state on denaturant concentration (Fig. 2). Thus, it seemed that some segments of the sequence contract in a specific manner: contraction was more than that predicted for the unfolded state because of the formation of specific structure by some segments of the protein sequence [68]. It was also observed that the greater the extent of specific structure detected by FRET in this manner in the kinetic molten globule, the greater too was the extent of secondary structure detected by CD [41]. The FRET studies therefore allowed the identification of both specific and non-specific components of polypeptide chain collapse during the folding of barstar. Non-specific and specific components of the compaction of the protein chain have also been similarly detected in studies of the initial events that occur during the folding of adenylate cyclase [70].

It therefore appears that the kinetic molten globule does have at least some segments of specific structure, so long as the conditions of folding are stabilizing enough. Equilibrium molten globules appear to have native-like topologies (folds) [71], either because of a few specific tertiary interactions that remain undetectable, or because the native-like fold is determined by how secondary structures and side-chains can loosely pack together to form the kernel of a hydrophobic core. The kinetic molten globule is likely to be similar in this respect, and a native-like fold in the kinetic molten globule is expected to make further folding faster. It would appear therefore that the evolutionarily selected sequence adopts only a single fold in the molten globule state, as it does in the native state, because only the unique fold allows all stabilizing

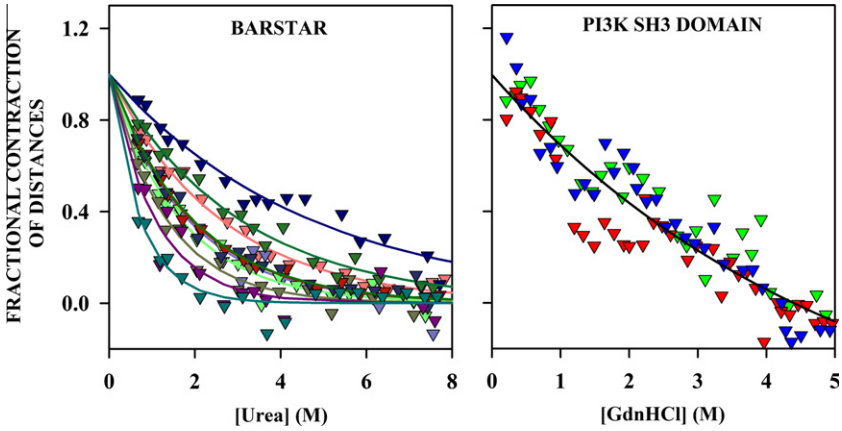

Fig. 1. Asynchronous and synchronous chain collapse upon denaturant dilution in the kinetic molten globule populated at 1 ms of refolding. In the case of barstar, all eleven intra-molecular distances (each shown by a different color/line) contract asynchronously upon denaturant dilution [68]. In the case of the PI3K SH3 domain, the three measured intra-molecular distances contract synchronously [69].

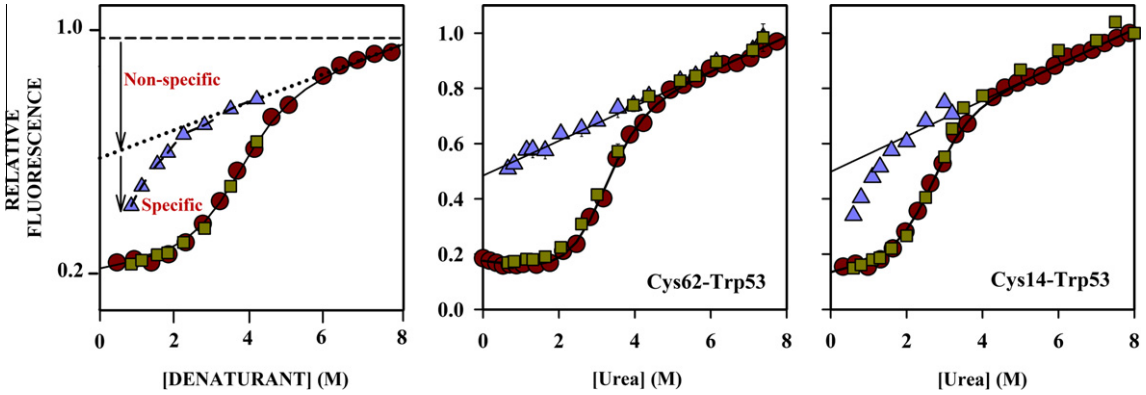

Fig. 2. Specific and non-specific components of chain collapse are observable in the kinetic molten globule, I_E, populated at 1 ms of folding of barstar. Intra-molecular distances were measured by FRET: the fluorescence change is a measure of (1-E), where E is the FRET efficiency, a measure of intra-molecular distance. The figure shows that intra-molecular distances contract in I_E upon dilution of urea. The purple triangles show how the dependence of the signal for I_E depends on denaturant concentration, compared to that for the completely unfolded state at high denaturant concentration (brown circles). Different intra-molecular distances contract to different fractional extents. Some distances (e.g. the Cys62–Trp53 distance) contract to the extent expected for non-specific solvent-induced collapse, while others (e.g. the Cys14-Trp53 distance) contract more than that predicted by solvent-change. The figure is based on data from [67,68].

interactions and structures to be consistent with each other [72] even at this early stage in folding. It should, however, be remembered that the specific structure in the kinetic molten globule need not be only native-like. Non-native secondary structure has been detected in the kinetic molten globules populated at a few milliseconds of folding for several proteins [73–75].

Sub-millisecond measurements of collapse and folding

Multiple phases of folding in the first millisecond

The advent of microsecond mixing methods [76,77], some with mixing dead-times as short as 4 μs, has yielded important information about what happens in the first 1 ms of folding. (1) The folding of many proteins appears to begin by chain collapse in the sub-millisecond time domain [38,41,49,78–80]. (2) For many proteins, at least two steps can be kinetically resolved in the first millisecond [35,41,79–81]. The fastest step, complete within 4 μs for protein L [38], 20 μs for cytochrome c, lysozyme and apomyoglobin [80], within 35 μs for DHFR [79], within 70 μs for ribonuclease A [81] and within 120 μs for barstar [41], appears to represent non-specific collapse of at least some segments of the polypeptide chain. The slower step in the 150–500 μs domain, appears to lead to the formation of a kinetic molten globule. (3) Secondary structure appears to form to a negligible or minor extent during the 150–500 μs kinetic phase in the case of cytochrome c [82] and monellin [83], but has been shown to form to a significant extent in the case of apomyoglobin [84,85] and ribonuclease A [86]. (4) In the case of barstar [41] and ribonuclease A [86], the second phase which leads to secondary structure formation in strongly stabilizing conditions is also accompanied by the formation of hydrophobic surfaces to which ANS can bind (Fig. 3). In the case of barstar, ANS does not bind to the product of the faster phase complete within 120 μs. (5) The R_g of the kinetic molten globules that are the products of sub-millisecond folding of several proteins scales as $N^{0.35}$ [87]. This scaling is similar to the scaling ($N^{0.33}$) expected for a homopolymer in poor solvent. In summary, the sub-millisecond mixing measurements indicate that non-specific collapse occurs within the dead-time (4–150 μs) of mixing, and that specific structure forms later within the first millisecond.

The kinetics of each sub-millisecond folding phase was found to be exponential in nature. In the case of cytochrome c, large unstructured fragments of the protein were found to undergo rapid collapse with exponential kinetics with characteristic times very similar to those seen for the exponential kinetics of initial chain collapse of the intact protein [88], suggesting that the exponential kinetics was not due to the formation of specific structure during

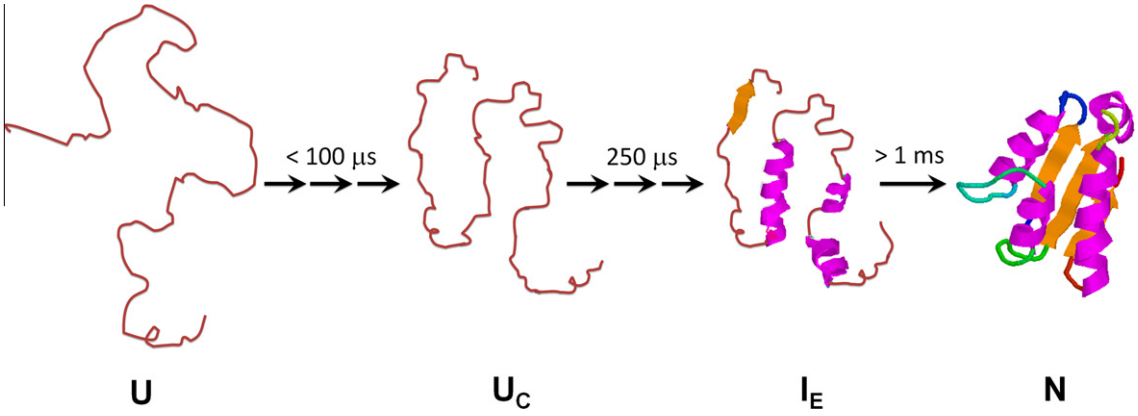

Fig. 3. Chain collapse during first millisecond of folding of barstar. A non-specific collapse occurs within the first 100 μs leading to the formation of a structure-less globule (pre-molten globule) U_C. This collapse reaction is observed by FRET, and appears to be induced by solvent-change. Marginally more chain collapse observable by FRET, as well as the development of hydrophobic clusters observable by ANS binding, occur over the next millisecond, leading to the formation of the kinetic molten globule intermediate, I_E. I_E possesses specific structure, including secondary structure, in strongly stabilizing conditions, but not in less stable conditions. I_E folds further on the 10–100 ms time scale, with further contraction of intra-molecular distances. The figure is based on data from [41].

the collapse reaction, as otherwise believed (see above). In the case of ribonuclease A, a folding-incompetent variant was found to collapse to a lesser extent than does the folding-competent protein, as measured by ANS binding. Nevertheless, both forms of the protein showed the same extent of change in far-UV CD within 300 μs of the commencement of folding [86]. This observation suggests that the very initial CD changes may not directly report on the formation of secondary structure. It is well known that aromatic residues can contribute substantially to the far-UV CD spectrum of a protein [89–92], and it seems likely that the CD changes observed during the first 300 μs may be the consequence of non-specific burial of aromatic residues. In this context, it should also be noted that the observation of ANS binding during chain collapse cannot also be taken as a signature of specific collapse, as was done [86], because ANS can bind non-specifically to hydrophobic patches [93,94]. Clearly, the observation of exponential kinetics cannot be used to conclude that specific structure has formed, at least for these two proteins.

Kinetics of the initial non-specific chain collapse reaction

Microsecond mixing methods are unable to capture the kinetics of the initial non-specific chain collapse reaction during folding, which usually occurs within the dead time (4–120 μs) of microsecond mixing (see above). Experimental estimates of intra-molecular contact formation in unfolded polypeptides appear to converge to a time of 1 μs for contact to be made by points separated by 100 residues in the polypeptide chain [95], but collapse must be driven by the formation of many contacts, and it is not clear how fast all the intra-molecular contacts form during chain collapse. On the other hand, temperature jump studies suggest that chain collapse may occur much faster, within 100 ns in a denatured small protein [96], and a single molecule study has suggested that collapse can occur this fast even during folding [15]. Such a rate of collapse is even faster than the rate of secondary structure formation by short peptides [97]. A recent single molecule fluorescence study of the transition path time, the time taken for a polypeptide chain to fold by diffusing over the transition state barrier once it becomes committed to do so in the course of its diffusive meandering, suggests that this transition may take only about 10 μs [98]. It is, however, not obvious as to what extent of chain collapse and specific structure formation has occurred prior to, and has led to the moment of commitment to fully fold.

Although polypeptide chains are heteropolymers for which very many intra-molecular interactions are possible, the collapse reaction they undergo initially during folding appears to be strikingly similar to that undergone by homopolymers upon a change in solvent from good to bad. Such similarity would be expected only if very few specific interactions form when a polypeptide chain collapses at the commencement of folding. Unfortunately, it is not possible to theoretically predict the characteristic time of polymer collapse [99]. Theoretical and computational studies of the kinetics of homopolymer chain collapse have suggested that homopolymers collapse in a continuous manner [100–102], or with multiphase kinetics [20,103–105], just as polypeptide chains appear to do (see above). The similarity of polypeptide chain collapse during folding to homopolymer collapse suggests that although the former must be driven by the formation of many contacts (see above), these contacts must be non-specific in nature.

Kinetics of the collapse reaction leading to the formation of the kinetic molten globule

An important question is whether the sub-millisecond folding reaction leading to the formation of the kinetic molten globule is an all-or-none, barrier-limited transition or a gradual transition. Because barrier-limited processes show exponential kinetics, the observation of exponential kinetics for the collapse reactions was interpreted in earlier studies to indicate that chain collapse was a barrier-limited process, and that it was barrier-limited because it led to the formation of specific structure [106,107]. But although barrier-limited processes invariably show exponential kinetics, the converse is not necessarily true (see above). Moreover, gradual processes may also show exponential kinetics [108]. Hence, the mere observation of exponential kinetics does not imply that the collapse reaction is barrier-limited. An early theoretical study [109] had suggested that even entire folding and unfolding reactions may be gradual in nature, a suggestion borne out both by experiment [16,17,19,110–112] and simulation [113]. If chain collapse and the formation of the kinetic molten globule occur gradually and continuously through near-degenerate states of intermediate densities, it would be slowed down by hydrodynamic coupling of the polypeptide chain with the solvent. Different magnitudes of hydrodynamic coupling with solvent as well as internal friction would lead to different characteristic times of collapse for different polypeptide chains.

In the case of barstar, the second slower sub-millisecond kinetic phase of folding, measured using ANS binding, was interpreted to be a gradual process even when it led to the formation of specific structure in the kinetic molten globule (Fig. 3), on the basis of its rate being independent of denaturant concentration and of temperature [41]. For other proteins too, equilibrium models of the kinetic molten globules that are the final product of sub-millisecond folding, were also found to unfold very non-cooperatively or gradually [114–117]. The probable reason why kinetic molten globules, or their equilibrium molten globule analogs, may display gradual unfolding transitions is the absence of a consolidated core, and the lack of defined tertiary interactions.

In this context it is important to note that the rates of the sub-millisecond folding reactions of other proteins also appear to have very little if any dependence on denaturant concentration [106,107,118,119], as do the folding reactions of some ultra-fast folding proteins whose folding is slowed down by marginal folding barriers and is hence complete within a few microseconds [120]. Although these reactions were assumed to be barrier-limited processes because of the observation of exponential kinetics, they may well be gradual in nature [112]. It should also be noted that it may be possible to tune the initial collapse reactions during folding from being gradual to being barrier-limited by mutation or by a change in folding conditions, just as it is possible to do for the folding reactions of ultra-fast folding proteins [111,121,122] whose entire folding reactions are complete within a few microseconds.

An equilibrium model of chain collapse

Notwithstanding the immense utility and necessity of kinetic experiments for directly characterizing the initial chain collapse reaction, it is still difficult to utilize high resolution probes with microsecond time resolution. Equilibrium studies of protein folding have been instrumental in identifying collapsed partially structured non-native forms of proteins, but it is not straightforward in establishing whether these conformations populate folding pathways [54]. In the case of barstar, it was found that the protein could be fully unfolded at high pH (D form), and that folding could be induced in progressively structured intermediates by the addition of salt (Fig. 4) [123]. Upon the addition of low concentrations of salt sufficient to screen electrostatic repulsion, the protein was found to collapse into a structure-less globule (P form), fully hydrated in its interior, with no secondary or tertiary structure, and with a volume somewhat larger than that of the native state. Upon further addition of salt, secondary and tertiary structure were found to form concurrently with the extrusion of water from the core. Rigidification of the hydrophobic core, as measured by time-resolved fluorescence anisotropy decay kinetics of a core tryptophan residue [124], was found to occur at a stage (I_{CC}) when specific structure has only partially formed. Finally water is fully extruded from the core and a dry molten globule (B form) forms. On the basis of kinetic experiments, in which folding was started from each of these different forms of the protein, it could be shown that P, I_{CC} and B are on the direct pathway of folding of the D form to the N state. This equilibrium model of collapse clearly delineated a non-specific collapse reaction from subsequent specific structure formation. It should be noted that this equilibrium model is based on the study of collapse and structure formation in the absence of any denaturants, which could potentially modulate, in a non-specific manner, both the collapse and structure formation reactions. It is likely that pre-molten globule forms with progressively increasing structural content, as identified and characterized in this equilibrium model of folding, precede the kinetic molten globule that is populated at about a millisecond of folding on the folding pathways of many proteins (see above).

Physical and chemical forces that drive chain collapse and structure formation

Driving chain collapse

Any polymer chain is expected to undergo a collapse reaction when intra-chain interactions dominate over chain-solvent interactions. What are the intra-chain interactions that drive a protein chain to undergo non-specific collapse? It is generally thought that chain collapse is driven by hydrophobic interactions leading to the clustering of non-polar residues [45,46,125]. Such hydrophobic collapse is expected to be rapid because van der Waal's interactions are non-specific, and hence, expected to form very rapidly. But chain collapse may not be driven entirely by hydrophobic interactions, as suggested in studies of the temperature-induced collapse of unfolded proteins [126]. This and other studies suggest that the intra-chain interactions that drive chain collapse may instead be hydrogen bonding interactions in the polypeptide backbone, which act in a non-specific long-range manner [127–131] or in a specific short-range manner [126,132]. It is known that hydrogen bonding interactions strengthen upon dilution of urea to initiate folding, while hydrophobic interactions remain unaffected [133,134]. In the case of barstar, the ANS fluorescence-monitored formation of hydrophobic clusters which occurred in the slower of the two sub-millisecond phases of folding (see above) was preceded by a faster phase which may be driven by hydrogen bonding interactions. It should, however, be noted that it would be difficult to achieve as well as to sustain any specific non-local hydrogen bonding interactions in the initial structure-less globule or in the kinetic molten globule formed at about 1 ms of refolding, both of which are fluid in that they possess very few if any specific tertiary interactions.

If stabilized by hydrophobic interactions, the structure-less globule formed as a result of non-specific collapse is expected to be a heterogeneous ensemble with fluctuating and continuously changing intra-molecular interactions because van der Waals interactions are non-specific in nature. If the structure-less globule is only loosely compact and highly dynamic, there will be relatively little entropic cost to its formation. In accordance with theoretical studies [135], the extent of polypeptide chain collapse is also known to be affected by electrostatic interactions [14,31,123], just as the dimensions of intrinsically disordered proteins are [136]. The interactions that influence the stability of the structure-less globule are expected to be predominantly non-native interactions, given that it is formed in a gradual diffusive process that is non-specific in nature, and it is likely that at least some of these non-native interactions persist further into the folding process.

Driving specific structure formation

The kinetic consequence of initial chain collapse in facilitating folding is not obvious. Chain compaction greatly reduces the conformational space available for the chain to sample during folding; the entropy barrier to structure formation during folding is lowered. In this way, folding could be speeded up. On the other hand, if chain compaction were to lead to too dense a collapsed form, then it could be expected that the diffusive exploration of conformational space would be hindered because of internal friction [137]. It is likely, because the structural rearrangements of folding have to occur in the condensed form as a result of initial chain compaction, that folding occurs through many local minima (intermediates) on the energy landscape. The polypeptide chain is likely to become less flexible upon contraction as intra-chain interactions dominate over chain-solvent interactions [34]. Consequently folding would be expected to slow down. The advantage of non-spe-

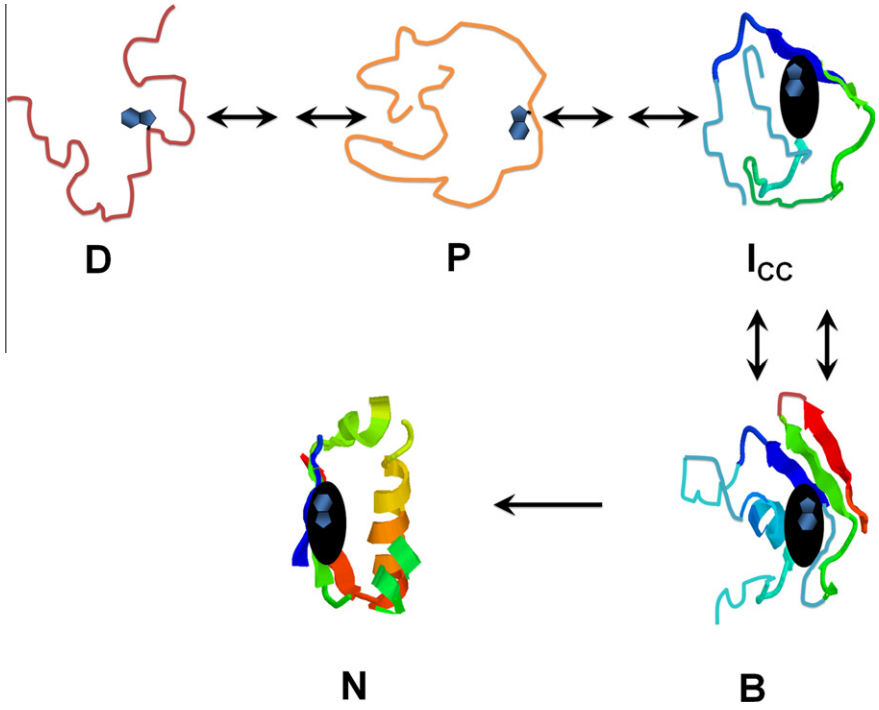

Fig. 4. Equilibrium model for the folding of high pH-unfolded barstar. The high pH denatured form, the D form, has the same dimensions as the urea or GdnHCl unfolded protein. Upon addition of salt, it undergoes a global collapse to form the structure-less globule P, whose volume is about 30% more than the native state. The P form is devoid of secondary or any other specific structure. Upon further addition of salt, specific structure develops in a highly non-cooperative manner. The hydrophobic core consolidates and rigidifies in I_{CC}, when secondary structure has only partially developed. Upon yet more addition of salt, water is ejected from the core, and additional secondary structure forms leading to the dry molten globule form B. The transition from P to B appears to be essentially a continuous structural transition. The P, I_{CC} and B forms are folding intermediates that populate the direct folding pathway from the D form to the N state. The figure is based on data from [6,124]. The indole ring shown is the side-chain of a tryptophan residue whose dynamics were measured and used to report when the core (black oval) becomes rigid during the overall folding process.

cific collapse preceding structure formation is not obvious. Theoretical studies have shown that sequences that fold the fastest are those for which collapse occurs simultaneously with the formation of secondary structure [138]. It is easy to envisage that this is because fewer non-native interactions form when collapse and structure formation occur concurrently, and so fewer non-native interactions have to be broken for further folding to occur. Obviously, the requirement for initial chain collapse during folding is poorly understood [139]. It seems logical to envisage that a very rapid initial chain collapse of the polypeptide chain during folding, in which hydrophobic residues get sequestered away from solvent, would ensure that aggregation events arising from non-specific inter-chain hydrophobic interactions are kept to a minimum. It is likely that it is the pattern of polar and non-polar side-chains on the evolutionarily selected polypeptide chain which ensures that such sequestration of non-polar residues does indeed takes place during the initial collapse reaction.

The role of initial chain collapse in facilitating further structure formation needs to be understood better. When the chain becomes sufficiently compact, steric constraints will favor chain conformations that are themselves compact as are secondary structural elements [140]. Also, when water gets extruded from the interior as a consequence of chain compaction, chain-solvent hydrogen bonding will be replaced by chain-chain hydrogen bonding because of the high energetic cost of desolvation, and chain-chain hydrogen bonding is efficiently accomplished by the formation of secondary structures. Not surprisingly, water extrusion and secondary structure formation were found to occur concomitantly in studies of salt-induced chain folding of high pH-unfolded barstar, after an initial non-specific collapse of the polypeptide chain [6].

Native proteins seem to be required to possess rigid, tightly packed cores. Tight core packing appears to require that at least some secondary structure first forms, well after the initial chain collapse reaction [141]. For core consolidation to occur, the core residues have to be first brought into proximity, and this would be enabled by the diffusive and exploratory nature of the initial chain collapse reaction. An all-or-none collapse reaction would be unlikely to accomplish this. Core consolidation itself is likely to be a cooperative process because many residue side-chains have to fit collectively into place [124], and secondary structural units would then pack tightly against each other after at least some concerted orientational adjustment, resulting in a coarse native-like topology. The tightly packed core would then dictate how the remaining secondary structure is formed and arranged.

Chain collapse in the context of other models of folding

The framework model

The framework model for how proteins fold, according to which the formation of secondary structure precedes that of tertiary structure, has been influential in protein folding studies [142–144]. In this hierarchical model, folding begins at several points along the sequence, leading to the formation of transient structures stabilized by local interactions [143]. Mutual stabilization of the local (secondary) structures occurs upon collision. Hence, the framework model posits that compaction of the polypeptide chain during folding occurs when fluctuating elements of secondary structure coalesce together by diffusion, a process described quantitatively by the diffusion collision model [145,146]. The framework model gives precedence to secondary structure formation over chain collapse as the initial event in folding, in contrast to the hydrophobic collapse model in which chain collapse ultimately drives secondary structure formation. Although proteins may look

like they have been assembled from their secondary structural elements (even though the order of their assembly cannot be predicted), there now seems to be strong evidence (see above) that polypeptide chain collapse precedes any significant secondary structure formation.

It appears that only when the polypeptide chain hydrophobicity is intrinsically low [139], or when folding is carried out at very low temperature where hydrophobic interactions are weakened [147], can secondary structure formation precede chain collapse. For the ultra-fast folding proteins, whose folding is complete in the microsecond time domain, it has not been possible, experimentally, to delineate chain collapse from secondary structure formation. Computer simulations of twelve ultra-fast folding proteins have suggested that secondary structural elements form early during folding, at a time when only a small number of non-local contacts have formed [122]. But for these proteins, secondary structural elements were found to form transiently in the unfolded state itself, and folding appeared to initiate in regions that formed native structure in the unfolded state [122]. For many ultra-fast folding proteins that are helical in nature, it is possible that folding is very fast because helix formation can be nucleated at many points, and the consequent operation of multiple folding pathways may be the reason why these proteins fold very rapidly [121].

Nucleation models.

The first nucleation models proposed [148,149] envisaged folding to be initiated by nucleation. In these models, a folding nucleus is formed through random search by small continuous regions of sequence through local interactions, and such nucleation could occur independently in separate parts of the polypeptide chain. It was proposed that such a nucleus grew by the addition of chain segments that are adjacent to it in the amino acid sequence. Formation of the nucleus is slow, and subsequent structure propagation is rapid. The existence of the folding nucleus significantly increases the rate of formation of native structure. It was suggested that a nucleation site could be a specific pocket stabilized by hydrophobic interactions [150]. In such nucleation models, global non-specific chain collapse is not expected to occur initially during folding, but during the structure propagation step.

The alternative nucleation model is based on the idea that folding nuclei are not local, but that they are formed by diffusive motion of the segments of the polypeptide chain, which brings together specific non-adjacent segments of sequence to form a condensed conformation [151]. Once such a nucleus forms, the chain rapidly condenses further to form the compact native state. The folding nucleus is formed by a defined set of native contacts between residues scattered along the length of the polypeptide chain. It is expected to be present in the transition state of folding, along with optional contacts that stabilize it. An early computer simulation study could identify a specific nucleus that formed early as the transition state of folding [152], and many subsequent simulations have identified folding nuclei [3,153]. The transition state is now referred to as an extended nucleus. The nucleation process would be similar to the nucleation-growth mechanism of first order (all-or-none) phase transitions, with the nucleus being thermodynamically the most unstable state on the reaction coordinate. For a folding reaction, nucleation would be thermodynamically unfavorable because of the entropic cost of forming the condensed nucleus. In this nucleation-growth or nucleation condensation model [153,154], substantial chain collapse occurs prior to the formation of a folding nucleus [9,39,51,52,155,156].

The nucleation-condensation model for protein folding is used to describe the folding of proteins that appear to fold by a two-state mechanism [157]. Indeed, it is strictly applicable to proteins that display such an all-or-none folding mechanism. For such proteins, the folding nucleus would be present in the transition state of folding, and no intermediates would be observable after the nucleation event because folding subsequently occurs in a downhill manner. Nevertheless, the nucleation growth model has been applied more generally to proteins that fold *via* an intermediate [158], with nucleation leading to the formation of the intermediate.

One approach to determine the residues that comprise the folding nucleus is to identify non-functional residues that have remained conserved in functionally related proteins whose sequences have diverged during evolution. In the case of the cytochromes and globins, potential folding nuclei have been identified as the structures being formed by conserved non-functional residues that are in contact with each other [159]. The residues of the folding nuclei identified in this manner were found to be at the interface of α-helices, and the formation of such nuclei would require substantial collapse of the polypeptide chain.

The other approach is to characterize the structure of the transition state by protein engineering, using Φ-value analysis [155,160,161]. The Φ value is a measure of the extent to which a mutation destabilizes a transition state relative to the native state. Although Φ values measure energetic quantities, they are interpreted to give structural information on the transition state of folding. A Φ value of 1 is taken to mean that an interaction is as well formed in the transition state as it is in the native state of the wt protein. A Φ value of 0 is taken to mean that the interaction is not at all formed in the transition state. A partial Φ value is nearly always interpreted to represent partial formation of an interaction, although it could alternatively signify that multiple pathways of folding are operative. The folding nucleus is taken to comprise residues that show high Φ values, and partial Φ values are assumed to contribute to the extended nucleus.

Unexpectedly, residues identified by Φ value analysis to participate in the folding nuclei of several proteins were found not to exhibit preferential evolutionary conservation [162]. Another study showed that only 3 out of the 125 very accurately determined Φ values reported in the literature for different proteins were larger than 0.8 [163], bringing into question whether specific nuclei do indeed form. More recently, the interpretation of a Φ value representing the strength of a specific tertiary interaction has been questioned. When multiple mutations were made at a site [164], it was observed that the destabilization of the transition state was linearly proportional to the destabilization of the native state, and that even highly destabilizing mutations had a negligible effect on transition state structure. This observation suggested that the Φ value is a general property of the structural site, and may not be indicative of any side-chain interactions such as those that would define a folding nucleus [165]. Another analysis suggested that Φ values are correlated more with local interactions than with specific tertiary interactions that would define a folding nucleus, and that the transition state resembles a molten globule [166]. It is possible that the native-like topology of the transition state [167] is determined by the same factors that allow molten globules to achieve their native tertiary folds (see above). It has been suggested that it may not be possible to specify the existence of a specific folding nucleus [168] and hence temporally resolve the collapse reaction by the use of this elegant protein engineering methodology [161].

Summary

This review has evaluated our current understanding of polypeptide chain collapse, an essential component of protein folding reactions. In particular, it appears now that non-specific chain collapse may indeed be the first event in protein folding. The step-

wise and gradual nature of polypeptide chain collapse reactions has become apparent in folding studies of more proteins. Much remains, however, to be understood of the nature of the intra-molecular interactions that drive chain collapse. In particular, it has become even more important to identify the nature and extent of transient fluctuating structure in the unfolded state, and to determine what component of this structure plays a dominant role in chain collapse and subsequent structure formation. It is still not clear as to what extent the initial structure-less globule and the later kinetic molten globule possess native-like folds, and how they achieve their native-like folds. The physical basis of how chain collapse itself drives further structure formation needs further experimental study. While there has been much progress, important questions remain unanswered.

Acknowledgments

I thank members of my laboratory, as well as Shachi Gosavi, for discussion and comments on the manuscript, and Jogender Singh and Amrita Dasgupta for assistance in preparing this review. The author is a recipient of a JC Bose National Research Fellowship from the Government of India. Work in his laboratory has been funded by the Tata Institute of Fundamental Research, by the Department of Biotechnology and by the Department of Science and Technology, Government of India.

References

[1] J.E. Kohn, I.S. Milett, J. Jacob, B. Azgrovic, Y.M. Dillon, et al., Proc. Natl. Acad. Sci. USA 101 (2004) 12491–12496.
[2] P.J. Flory, Statistical Mechanics of Chain Molecules, Wiley, New York, 1969.
[3] D. Thirumalai, E.P. O'Brien, G. Morrison, C. Hyeon, Annu. Rev. Biophys. 39 (2010) 159–183.
[4] W. Pfeil, P.L. Privalov, Biophys. Chem. 4 (1976) 41–50.
[5] V.R. Agashe, J.B. Udgaonkar, Biochemistry 34 (1995) 3286–3299.
[6] B.R. Rami, J.B. Udgaonkar, Biochemistry 41 (2002) 1710–1716.
[7] K.A. Dill, D. Shortle, Annu. Rev. Biochem. 60 (1991) 795–825.
[8] A.M. Saxena, J.B. Udgaonkar, G. Krishnamoorthy, J. Mol. Biol. 359 (2006) 174–189.
[9] V.A. Voelz, M. Jäger, S. Yao, Y. Chen, L. Zhu, S.A. Waldauer, G.R. Bowman, M. Friedrichs, O. Bakajin, L.J. Lapidus, S. Weiss, V.S. Pande, J. Am. Chem. Soc. 134 (2012) 12565–12577.
[10] E.V. Kuzmenkina, C.D. Heyes, G.U. Nienhaus, J. Mol. Biol. 357 (2006) 313–324.
[11] B. Schuler, E.A. Lipman, W.A. Eaton, Nature 419 (2002) 743–747.
[12] E. Sherman, G. Haran, Proc. Natl. Acad. Sci. USA 103 (2006) 11539–11543.
[13] A. Hoffmann, A. Kane, D. Nettels, D.E. Hertzog, P. Baumgärtel, J. Lengefeld, G. Reichardt, D.A. Horsley, R. Seckler, O. Bakajin, B. Schuler, Proc. Natl. Acad. Sci. USA 104 (2007) 105–110.
[14] H. Hofmann, R.P. Golbik, M. Ott, C.G. Hubner, R. Ulbrich-Hofmann, J. Mol. Biol. 376 (2008) 597–605.
[15] D. Nettels, I.V. Gopich, A. Hoffmann, B. Schuler, Proc. Natl. Acad. Sci. USA 104 (2007) 2655–2660.
[16] R. Swaminathan, J.B. Udgaonkar, N. Periasamy, G. Krishnamoorthy, Biochemistry 35 (1996) 9150–9157.
[17] G.S. Lakshmikanth, K. Sridevi, G. Krishnamoorhty, J.B. Udgaonkar, Nat. Struct. Biol. 8 (2001) 799–804.
[18] S.K. Jha, D. Dhar, G. Krishnamoorthy, J.B. Udgaonkar, Proc. Natl. Acad. Sci. USA 106 (2009) 11113–11118.
[19] K.A. Dill, Biochemistry 29 (1990) 7133–7155.
[20] D. Thirumalai, J. Phys. 5 (1995) 1457–1467.
[21] D. Hamada, M. Hoshino, M. Kataoka, A.L. Fink, Y. Goto, Biochemistry 32 (1993) 10351–10358.
[22] Y.K. Mok, C.M. Kay, L.E. Kay, J. Forman-Kay, J. Mol. Biol. 289 (1999) 619–638.
[23] B. Anil, Y. Li, J.-H. Cho, D.P. Raleigh, Biochemistry 45 (2006) 10110–10116.
[24] M.J. Cliff, C.J. Craven, J.P. Marston, A.M. Hounlow, A.R. Clarke, J.P. Waltho, J. Mol. Biol. 385 (2009) 266–277.
[25] C.L. Pashley, G.J. Morgan, A.P. Kalverda, G.S. Thomson, C. Klenthous, S.E. Radford, J. Mol. Biol. 416 (2012) 300–318.
[26] V.N. Uversky, J.R. Gillespie, A.L. Fink, Proteins 41 (2000) 415–427.
[27] V.N. Uversky, Eur. J. Biochem. 269 (2002) 2–12.
[28] N. Jain, M. Bhattacharya, S. Mukhopadhyay, Biophys. J. 101 (2011) 1720–1729.
[29] J. Klein-Seetharaman, M. Oikawa, S.B. Grimshaw, J. Wirmer, E. Duchardt, T. Ueda, T. Imoto, L.J. Smith, C.M. Dobson, H. Schwalbe, Science 295 (2002) 1719–1722.
[30] D.J. Felitsky, M.A. Lietzow, H.J. Dyson, P.E. Wright, Proc. Natl. Acad. Sci. USA 105 (2008) 6278–6283.
[31] L. Pradeep, J.B. Udgaonkar, Biochemistry 43 (2004) 11393–11402.
[32] J. Juneja, J.B. Udgaonkar, Curr. Sci. 84 (2003) 157–172.
[33] K.H. Mok, L.T. Kuhn, M. Goez, I.J. Day, J.C. Lin, N.H. Andersen, P.J. Hore, Nature 447 (2007) 106–109.
[34] F. Krieger, B. Fierz, O. Bieri, M. Drewello, T. Kiefhaber, J. Mol. Biol. 332 (2003) 265–274.
[35] M. Buscaglia, B. Schuler, L.J. Lapidus, W.A. Eaton, J. Hofrichter, J. Mol. Biol. 332 (2003) 9–12.
[36] K. Chattopadhyay, E.L. Elson, C. Frieden, Proc. Natl. Acad. Sci. USA 102 (2005) 2385–2389.
[37] H. Chen, E. Rhoades, J.S. Butler, S.N. Loh, W.W. Webb, Proc. Natl. Acad. Sci. USA 104 (2007) 10459–10464.
[38] S.A. Waldauer, O. Bakajin, T. Ball, Y. Chen, S.J. DeCamp, M. Kopka, M. Jager, V.R. Singh, W.J. Wedemeyer, S. Weiss, S. Yao, L.J. Lapidus, HFSP J. 2 (2008) 388–395.
[39] S.A. Waldauer, O. Bakajin, L.J. Lapidus, Proc. Natl. Acad. Sci. USA 107 (2010) 13713–13717.
[40] M. Kinoshita, K. Kamagata, A. Maeda, Y. Goto, T. Komatsuzaki, S. Takahashi, Proc. Natl. Acad. Sci. USA 104 (2007) 10453–10458.
[41] K.K. Sinha, J.B. Udgaonkar, Proc. Natl. Acad. Sci. USA 105 (2008) 7998–8003.
[42] J.B. Udgaonkar, Annu. Rev. Biophys. 37 (2008) 489–510.
[43] K.K. Sinha, J.B. Udgaonkar, Curr. Sci. 96 (2009) 1053–1070.
[44] T.Y. Yoo, S.P. Meisburger, J. Hinhaw, L. Pollack, G. Haran, T.R. Sosnick, K. Plaxco, J. Mol. Biol. 418 (2012) 226–236.
[45] G.V. Semisotnov, N.A. Rodionova, O.I. Razgulyaev, V.N. Uversky, A.F. Gripas, R.I. Gilmanshin, Biopolymers 31 (1991) 119–128.
[46] V.R. Agashe, M.C.R. Shastry, J.B. Udgaonkar, Nature 377 (1995) 754–757.
[47] L. Pradeep, J.B. Udgaonkar, J. Biol. Chem. 279 (2004) 40303–40313.
[48] V. Ratner, D. Amir, E. Kahana, E. Haas, J. Mol. Biol. 352 (2005) 683–699.
[49] M. Arai, E. Kondrashkina, C. Kayatekin, C.R. Matthews, M. Iwakura, O. Bilsel, J. Mol. Biol. 368 (2007) 219–229.
[50] E.A. Lipman, B. Schuler, O. Bakajin, W.A. Eaton, Science 301 (2003) 1233–1235.
[51] C. Magg, F.X. Schmid, J. Mol. Biol. 335 (2004) 1309–1323.
[52] K.M. Hamadani, S. Weiss, Biophys. J. 95 (2008) 352–365.
[53] P. Jennings, P.E. Wright, Science 262 (1993) 892–896.
[54] M. Arai, K. Kuwajima, Adv. Protein Chem. 53 (2000) 209–282.
[55] A.V. Finkelstein, O.B. Ptitsyn, Protein Physics, Academic Press, London, 2002.
[56] V.N. Uversky, A.L. Fink, FEBS Lett. 515 (2002) 79–83.
[57] T.R. Sosnick, M.D. Shtilerman, L. Mayne, S.W. Englander, Proc. Natl. Acad. Sci. USA 94 (1997) 8545–8550.
[58] L. Chen, G. Wildegger, T. Kiefhaber, K.O. Hodgson, S. Doniach, J. Mol. Biol. 276 (1998) 225–237.
[59] A.K. Bhuyan, J.B. Udgaonkar, Relevance of burst phase changes in optical signals of polypeptides during protein folding, in: M. Vijayan, N. Yathindra, A.S. Kolaskar (Eds.), Perspectives in Structural Biology, Universities Press, Hyderabad, 1999, p. 293.
[60] J. Jacob, B. Krantz, R.S. Dothager, P. Thiyagarajan, T.R. Sosnick, J. Mol. Biol. 338 (2004) 369–382.
[61] B.A. Krantz, L. Mayne, J. Rubmbley, S.W. Englander, T.R. Sosnick, J. Mol. Biol. 324 (2002) 359–371.
[62] R.E. Georgescu, J.H. Lee, M.E. Goldberg, M.L. Tasayko, A.F. Chaffotte, Biochemistry 37 (1998) 10286–10297.
[63] M.J. Parker, S. Marqusee, J. Mol. Biol. 293 (1999) 1195–1210.
[64] L. Pradeep, J.B. Udgaonkar, J. Mol. Biol. 324 (2002) 331–347.
[65] A. Dasgupta, J.B. Udgaonkar, Biochemistry 51 (2012) 4723–4734.
[66] K. Sridevi, G. Lakshmikanth, G. Krishnamoorthy, J.B. Udgaonkar, J. Mol. Biol. 337 (2004) 699–711.
[67] K.K. Sinha, J.B. Udgaonkar, J. Mol. Biol. 353 (2005) 704–718.
[68] K.K. Sinha, J.B. Udgaonkar, J. Mol. Biol. 370 (2007) 385–405.
[69] A. Dasgupta, J.B. Udgaonkar, J. Mol. Biol. 403 (2010) 430–445.
[70] T. Orevi, E. Ben Ishay, M. Pirchi, M.H. Jacob, D. Amir, E. Haas, J. Mol. Biol. 385 (2009) 1230–1242.
[71] Z.Y. Peng, P.S. Kim, Biochemistry 33 (1994) 2136–2141.
[72] N. Go, Annu. Rev. Biophys. Bioeng. 12 (1983) 183–210.
[73] D. Hamada, S. Segawa, Y. Goto, Nat. Struct. Biol. 3 (1996) 868–873.
[74] K. Kuwata, M.C.R. Shastry, H. Cheng, M. Hoshino, C.A. Batt, Y. Goto, H. Roder, Nat. Struct. Biol. 8 (2001) 151–155.
[75] J. Li, M. Shinjo, Y. Matsumura, M. Morita, D. Baker, M. Ikeguchi, H. Kihara, Biochemistry 46 (2007) 5072–5082.
[76] C.K. Chan, H. Yi, S. Takahashi, D.L. Rousseau, W.A. Eaton, J. Hofrichter, Proc. Natl. Acad. Sci. USA 94 (1997) 1779–1784.
[77] H. Roder, K. Maki, H. Cheng, Chem. Rev. 106 (2006) 1836–1861.
[78] M. Arai, K. Ito, T. Inobe, M. Nakao, K. Maki, K. Kamagata, H. Kihara, Y. Amenmiya, K. Kuwajima, J. Mol. Biol. 321 (2002) 121–132.
[79] M. Arai, M. Iwakura, C.R. Matthews, O. Bilsel, J. Mol. Biol. 410 (2011) 329–342.
[80] L.J. Lapidus, S. Yao, K.S. McGarrity, D.E. Hertzog, E. Tubman, O. Bakajin, Biophys. J. 93 (2007) 218–224.
[81] E. Welker, K. Maki, M.C.R. Shastry, D. Juminaga, R. Bhat, H.A. Scheraga, H. Roder, Proc. Natl. Acad. Sci. USA 101 (2004) 17681–17686.
[82] S. Akiyama, S. Takahashi, K. Ishimori, I. Morishima, Nat. Struct. Biol. 7 (2000) 514–520.
[83] T. Kimura, T. Uzawa, K. Ishimori, I. Morishima, S. Takahashi, T. Konno, S. Akiyama, T. Fujisawa, Proc. Natl. Acad. Sci. USA 102 (2005) 2748–2753.
[84] T. Uzawa, S. Akiyama, T. Kimura, S. Takahashi, K. Ishimori, I. Morishima, T. Fujisawa, Proc. Natl. Acad. Sci. USA 101 (2004) 1171–1176.

[85] T. Uzawa, C. Nishimura, S. Akiyama, K. Ishimori, S. Takahashi, H.D. Dyson, P.E. Wright, Proc. Natl. Acad. Sci. USA 105 (2008) 13859–13864.
[86] T. Kimura, S. Akiyama, T. Uzawa, K. Ishimori, I. Morishima, T. Fujisawa, S. Takahashi, J. Mol. Biol. 350 (2005) 349–362.
[87] T. Uzawa, T. Kimura, K. Ishimori, I. Morishima, T. Matsui, et al., J. Mol. Biol. 357 (2006) 997–1008.
[88] L. Qiu, C. Zachariah, S.J. Hagen, Phys. Rev. Lett. 90 (2003) 168103.
[89] R.W. Woody, Biopolymers 17 (1978) 1451–1467.
[90] K. Kuwajima, E.P. Garvey, V.E. Finn, C.R. Matthews, S. Sugai, Biochemistry 30 (1991) 7693–7703.
[91] S. Vuilleumier, J. Sancho, R. Loewenthal, A.R. Fersht, Biochemistry 32 (1993) 10303–10313.
[92] R.W. Woody, Eur. J. Biophys. 23 (1994) 253–262.
[93] L. Stryer, J. Mol. Biol. 13 (1965) 482–495.
[94] Y. Goto, T. Azuma, K. Hamaguchi, J. Biochem. 85 (1979) 1427–1438.
[95] O. Bieri, J. Wirz, B. Hellrung, M. Schutkowski, M. Drewello, T. Kiefhaber, Proc. Natl. Acad. Sci. USA 96 (1999). 9697-9601.
[96] M. Sadqi, L.J. Lapidus, V. Munoz, Proc. Natl. Acad. Sci. USA 100 (2003) 12117–12122.
[97] W.A. Eaton, V. Munoz, S.J. Hagen, G.S. Jas, L.J. Lapidus, E.R. Henery, J. Hofrichter, Annu. Rev. Biophys. Biomol. Struct. 29 (2000) 327–359.
[98] H.S. Chung, K. McHale, J.M. Louis, W.A. Eaton, Science 335 (2012) 981–984.
[99] Y.A. Kuznetsov, E.G. Timoshenko, K.A. Dawson, J. Chem. Phys. 103 (1995) 4807–4818.
[100] C. Williams, F. Brochard, H.L. Frisch, Annu. Rev. Phys. Chem. 32 (1981) 433–451.
[101] P.G. DeGennes, J. Phys. Lett. 46 (1985) L-639–L-642.
[102] H.S. Chan, K.A. Dill, Annu. Rev. Biophys. Biophys. Chem. 20 (1991) 447–490.
[103] J. Ma, J.E. Straub, E.I. Shakhnovich, J. Chem. Phys. 103 (1995) 2615–2624.
[104] E. Pitard, H. Orland, Europhys. Lett. 41 (1998) 467–472.
[105] G. Ziv, D. Thirumalai, G. Haran, Phys. Chem. Chem. Phys. 11 (2009) 83–93.
[106] M.C.R. Shastry, H. Roder, Nat. Struct. Biol. 5 (1998) 385–392.
[107] S.J. Hagen, W.A. Eaton, J. Mol. Biol. 297 (2000) 781–789.
[108] S.J. Hagen, Proteins Struct. Funct. Genet. 50 (2003) 1–4.
[109] D.C. Poland, H.A. Scheraga, Biopolymers 3 (1965) 401–419.
[110] M.M. Garcia-Mira, M. Sadqi, N. Fischer, J.M. Sanchez-Ruiz, V. Munoz, Science 298 (2002) 1291–1295.
[111] M. Gruebele, C. R. Biol. 328 (2005) 701–712.
[112] A. Naganathan, U. Doshi, V. Munoz, J. Am. Chem. Soc. 129 (2007) 5673–5682.
[113] G.G. Maisarudze, A. Liwo, S. Oldziej, H.A. Scheraga, J. Am. Chem. Soc. 132 (2010) 9444–9452.
[114] A. Shimizu, M. Ikeguchi, S. Sugai, Biochemistry 32 (1993) 13198–13203.
[115] B.A. Schulman, P.S. Kim, C.M. Dobson, C. Redfield, Nat. Struct. Biol. 4 (1997) 630–634.
[116] A.K. Chamberlain, S. Marqusee, Biochemistry 37 (1998) 1736–1742.
[117] U. Mayor, J.G. Grossmann, N.W. Foster, S.M.V. Freund, A.R. Fersht, J. Mol. Biol. 333 (2003) 977–991.
[118] S.H. Park, M.C.R. Shastry, H. Roder, Nat. Struct. Biol. 6 (1999) 943–947.
[119] K. Teilum, K. Maki, B.B. Kragelund, F.M. Poulsen, H. Roder, Proc. Natl. Acad. Sci. USA 99 (2002) 9807–9812.
[120] T. Cellmer, E.R. Henry, J. Kubelka, J. Hofrichter, W.A. Eaton, J. Am. Chem. Soc. 129 (2007) 14564–14565.
[121] K. Ghosh, S.B. Ozkan, K.A. Dill, J. Am. Chem. Soc. 129 (2007) 11920–11927.
[122] K. Lindorff-Larsen, S. Piana, R.O. Dror, D.E. Shaw, Science 334 (2011) 517–520.
[123] B.R. Rami, J.B. Udgaonkar, Biochemistry 40 (2001) 15267–15279.
[124] B.R. Rami, G. Krishnamoorthy, J.B. Udgaonkar, Biochemistry 42 (2003) 7986–8000.
[125] H.J. Dyson, P.E. Wright, H.A. Scheraga, Proc. Natl. Acad. Sci. USA 103 (2006) 13057–13061.
[126] D. Nettels, S. Muller-Spath, Proc. Natl. Acad. Sci. USA 106 (2009) 20740–20745.
[127] S.L. Crick, M. Jayaraman, C. Frieden, R. Wetzel, R.V. Pappu, Proc. Natl. Acad. Sci. USA 103 (2006) 16764–16769.
[128] A. Moglich, K. Joder, T. Kiefhaber, Proc. Natl. Acad. Sci. USA 103 (2006) 12394–12399.
[129] H. Daidone, S. Neuweiler, M. Doose, J.C. Sauer, PLoS Comput. Biol. 6 (2010) e1000645.
[130] H. Gong, L.L. Porter, G.D. Rose, Protein Sci. 20 (2011) 417–427.
[131] D.P. Teufel, C.M. Johnson, J.K. Lum, H. Neuweiler, J. Mol. Biol. 409 (2011) 250–262.
[132] G.D. Rose, P.J. Fleming, J.R. Banavar, A. Maritan, Proc. Natl. Acad. Sci. USA 103 (2006) 16623–16633.
[133] D.W. Bolen, G.D. Rose, Annu. Rev. Biochem. 77 (2008) 339–362.
[134] L.M. Holthauzen, J. Rosgen, D.W. Bolen, Biochemistry 49 (2010) 1310–1318.
[135] D. Stigter, D.O.V. Alonso, K.A. Dill, Proc. Natl. Acad. Sci. USA 88 (1991) 4176–4188.
[136] S. Muller-Spath, A. Soranno, V. Hirschfeld, H. Hofmann, S. Ruegger, L. Reymond, D. Nettels, B. Schuler, Charge interactions can dominate the dimensions of intrinsically disordered proteins, Proc. Natl. Acad. Sci. USA 107 (2010) 14609–14614.
[137] A. Soranno, B. Buchli, D. Nettels, R.R. Cheng; S. Muller-Spath, S.H. Pfeil, A. Hoffmann, E.A. Lipman, D.E. Makarov, B. Schuler, Proc. Natl. Acad. Sci. USA (2012) early edition, Available from: <www.pnas.org/cgi/doi/10.1073/pnas.1117368109>.
[138] D.K. Klimov, D. Thirumalai, Phys. Rev. Lett. 76 (1996) 4070–4073.
[139] A.M. Gutin, V.I. Abkevich, E.I. Shakhnovich, Biochemistry 34 (1995) 3066–3076.
[140] H.S. Chan, K.A. Dill, Proc. Natl. Acad. Sci. USA 87 (1990) 6388–6392.
[141] K. Sridevi, J. Juneja, A.K. Bhuyan, G. Krishnamoorthy, J.B. Udgaonkar, J. Mol. Biol. 302 (2000) 479–495.
[142] O.B. Ptitsyn, Dokl. Akad. Nauk SSSR 210 (1973) 1213–1215.
[143] P.S. Kim, R.L. Baldwin, Annu. Rev. Biochem. 51 (1982) 459–489.
[144] J.B. Udgaonkar, R.L. Baldwin, Nature 335 (1988) 694–699.
[145] M. Karplus, D.C. Weaver, Nature 260 (1976) 404–406.
[146] M. Karplus, D.C. Weaver, Protein Sci. 3 (1994) 650–668.
[147] C. Dumont, Y. Matsumura, S.J. Kim, J. Li, E. Kondrashkina, H. Kihara, M. Gruebele, Protein Sci. 15 (2006) 2596–2604.
[148] C. Levinthal, Biol. Syst. Proc. 67 (1969) 22.
[149] D.B. Wetlaufer, Proc. Natl. Acad. Sci. USA 70 (1973) 697–701.
[150] R.R. Matheson, H.A. Scheraga, Macromolecules 11 (1978) 819–829.
[151] T.Y. Tsong, R.L. Baldwin, P. McPhie, J. Mol. Biol. 63 (1972) 453–475.
[152] V.I. Abkevich, A.M. Gutin, E.I. Shakhnovich, Biochemistry 33 (1994) 10026–10036.
[153] P.F.N. Faisca, J. Phys.: Condens. Matter 21 (2009) 373102.
[154] A.R. Fersht, Proc. Natl. Acad. Sci. USA 92 (1995) 10869–10873.
[155] D.E. Otzen, L.S. Itzhaki, N.F. Elmsry, S.E. Jackson, A.R. Fersht, Proc. Natl. Acad. Sci. USA 91 (1994) 10422–10425.
[156] B. Nolting, D.A. Agard, Proteins 73 (2008) 754–764.
[157] S.E. Jackson, A.R. Fersht, Biochemistry 30 (1991) 10428–10435.
[158] T. Kiefhaber, A. Bachmann, G. Wildegger, C. Wagner, Biochemistry 36 (1997) 5108–5112.
[159] O.B. Ptitsyn, J. Mol. Biol. 278 (1998) 655–666.
[160] A.R. Fersht, A. Matouschek, L. Serrano, J. Mol. Biol. 224 (1992) 771–782.
[161] A.R. Fersht, S. Sato, Proc. Natl. Acad. Sci. USA 101 (2004) 7977–7983.
[162] S.M. Larson, I. Ruczinski, A.R. Davidson, D. Baker, K.W. Paxco, J. Mol. Biol. 316 (2002) 225–233.
[163] D.P. Raleigh, K.W. Plaxco, Protein Pept. Lett. 12 (2005) 117–122.
[164] J.G.B. Northey, K.L. Maxwell, A.R. Davidson, J. Mol. Biol. 320 (2002) 389–402.
[165] R.L. Baldwin, C. Frieden, G.D. Rose, Proteins 78 (2010) 2725–2737.
[166] A.N. Naganathan, V. Munoz, Proc. Natl. Acad. Sci. USA 107 (2010) 8611–8616.
[167] K.W. Plaxco, K.T. Simons, D. Baker, J. Mol. Biol. 277 (1998) 985–994.
[168] I.E. Sanchez, T. Kiefhaber, J. Mol. Biol. 334 (2003) 1077–1085.