Multiple Routes and Structural Heterogeneity in Protein Folding

Jayant B. Udgaonkar

National Centre for Biological Sciences, Tata Institute of Fundamental Research, Bangalore 560065, India; email: jayant@ncbs.res.in

Annu. Rev. Biophys. 2008. 37:489–510

The *Annual Review of Biophysics* is online at biophys.annualreviews.org

This article's doi: 10.1146/annurev.biophys.37.032807.125920

Copyright © 2008 by Annual Reviews. All rights reserved

1936-122X/08/0609-0489$20.00

Key Words

folding pathways, folding intermediates, competing pathways, protein unfolding

Abstract

Experimental studies show that many proteins fold along sequential pathways defined by folding intermediates. An intermediate may not always be a single population of molecules but may consist of subpopulations that differ in their average structure. These subpopulations are likely to fold via independent pathways. Parallel folding and unfolding pathways appear to arise because of structural heterogeneity. For some proteins, the folding pathways can effectively switch either because different subpopulations of an intermediate get populated under different folding conditions, or because intermediates on otherwise hidden pathways get stabilized, leading to their utilization becoming discernible, or because mutations stabilize different substructures. Therefore, the same protein may fold via different pathways in different folding conditions. Multiple folding pathways make folding robust, and evolution is likely to have selected for this robustness to ensure that a protein will fold under the varying conditions prevalent in different cellular contexts.

Contents

INTRODUCTION

Protein folding is an intriguing example of biological wizardry, in which a polypeptide chain self-assembles into the unique native structure that holds the key to its function. During the folding process, the polypeptide chain diffuses over a multidimensional energy surface, changing shape by condensing, turning, coiling, bending, twisting, and looping to finally produce the finished design. Several fundamental and interlinked questions remain about this self-packing puzzle. Do proteins take shape gradually, in fits and starts, or all at once? Are there flip-flops? What comes first, an outline of the shape or the details? Is there only one folding sequence for each protein?

Folding pathway: a specific sequence of structural events leading to fully folded protein

A vast conformational space is potentially available for the protein chain to sample by meandering over the free energy surface, until it finds the global free energy minimum, the native state. A completely random search of conformations would require an astronomically long time, but proteins typically take no more than a few seconds to fold. The difference in the estimated and biologically relevant timescales is called the Levinthal Paradox. A solution to the Levinthal Paradox becomes apparent when making an analogy between folding to a predetermined structure by random sampling of conformational space and the creation of a predetermined short sentence by a monkey typing randomly on a computer keyboard (126). It would take the monkey nearly 10^{40} keystrokes to type a 28-character sentence correctly by chance, but only a few thousand keystrokes if every correctly typed keystroke is retained in place and only corrections allowed. In the language of protein folding, biologically relevant timescales become possible if partly formed intermediate structures are retained in place. Even though the monkey may not realize it, there are many different ways, each a completely different sequence of correct keystrokes, of arriving at the predetermined sentence. Different folding pathways will portray completely different sequences of folding events. The concept of multiple folding pathways was powerfully made apparent earlier in an analogy comparing protein folding and the assembly of a jigsaw puzzle (36), which can be pieced together from many different starting points in many different ways, all leading to the completely solved puzzle or the native state.

The possibility of multiple folding routes was implicit in one of the earliest models proposed for protein folding, the diffusion-collision model (47), which envisages that folding proceeds through the diffusional collisions of fluctuating native-like microdomains. Such collisions are not required to occur in any particular order, but the microdomains must be sufficiently long-lived (stable) for the

collisions to be productive. Computer simulations show that folding can occur via many different routes as long as, on the average, native interactions are longer-lived than nonnative interactions (61), or if there is a small energy bias against locally incorrect structures (126). Indeed, much of the current work in protein folding is focused on understanding how the laws of physics and chemistry intervene to make certain that folding is not completely random.

This review examines current knowledge and understanding of how multiple pathways operate in protein folding and unfolding, as well as the role of structural heterogeneity, which is becoming increasingly apparent in protein folding reactions, in determining their utilization. Theoretical and computational studies, which have enriched this area of research, have been mentioned briefly, only to place experimental results in the correct context. Heterogeneity and multiple folding pathways arising from disulfide bond formation during folding are not examined (118).

INTERMEDIATES AND FOLDING PATHWAYS

Proteins appear as though they can be assembled from their parts, and there are much data to suggest that structure develops in a hierarchical manner during protein folding (6, 21, 45). The concept of progressive stabilization of structure in steps during folding led to the concept of partially structured folding intermediates defining a sequential pathway of folding, but this notion does not preclude several alternative pathways with many different folding routes. The accumulation of native-like structures in folding intermediates on the sequential folding pathways of many proteins has been demonstrated effectively by pulsed hydrogen exchange (HX) methods used in conjunction with NMR (5, 21, 45). Nevertheless, the roles of folding intermediates remain poorly defined. Are they merely kinetic milestones defining the pathway through which folding molecules are channeled, or do they play a role in directing such channeling? Does their accumulation serve as a pause in the folding process so that any accumulated errors can be corrected in the waiting time? Or do they accumulate because the barrier due to the folding error is large? Intermediates can accelerate folding (19). If a kinetic intermediate forms on the folding pathway after the rate-limiting step, then its equilibrium analog will fold faster than a fully unfolded protein (92); if it precedes the rate-limiting step, it may not (90). Intermediates can therefore play a productive role. Whatever their roles, folding intermediates can serve as useful signposts for identifying separate folding pathways.

There has been confusion about whether kinetic intermediates are obligatory for folding to occur, because they cannot be detected during the folding of many proteins (40). For these proteins, folding therefore appears to be a two-state process, and the study of their folding reactions becomes difficult because of the lack of partially folded intermediates to characterize. Fortunately, NMR methods now allow sparsely populated intermediates to be identified and characterized, even for a protein otherwise thought to be a two-state folder (50). The discovery of nonlinear free energy relationships in protein unfolding was the first indication of the existence of unstable intermediates (43, 123), which can also account for similar nonlinear free energy relationships observed for the unfolding of many apparently two-state folding proteins (96). Not surprisingly then, unstable intermediates can be stabilized sufficiently relative to the native state, by mutation (69) or by a change in folding conditions (116), to accumulate in the course of folding transitions. Hence, two-state folders can be converted into three-state folders by the stabilization of high-energy intermediates (31), and three-state folders can be converted into apparently two-state folders by the destabilization of folding intermediates (108). Of course, it is not easy to determine in such cases whether the intermediate being stabilized or destabilized is on the same pathway

Multiple pathways: different possible sequences of structural events, all leading to fully folded protein

Folding intermediate: a meta-stable partially folded form with substantial entropy because of poor residue packing; hence an ensemble of microstates

Hydrogen exchange (HX): exchange of the amide hydrogen with solvent protons or deuterons

TS: transition state

utilized before the perturbation in the folding conditions. The change in folding conditions or mutation may result in the stabilization of an intermediate and transition state (TS) on a hidden pathway, leading to the utilization of this pathway (91).

Gradual Folding and Unfolding

It now appears that folding reactions may occur in not one but many consecutive steps. The native states of several proteins can exist in equilibrium with partially unfolded and progressively destabilized forms: It seems almost natural to arrange these sparsely populated intermediate forms in a kinetic sequence of events occurring after the rate-limiting step of folding (21). Protein folding transitions may even be so highly uncooperative so as to be gradual structural transitions. Much of the evidence for this comes from equilibrium folding studies, where the use of high-resolution, site-specific structural probes is possible. For example, time-resolved fluorescence as well as NMR experiments have shown that structure is lost incrementally during the denaturant-induced equilibrium unfolding of barstar (55, 57). NMR and UV-resonance Raman spectroscopy studies have also indicated that the pH- or heat-induced unfolding of several other proteins is gradual and spatially decoupled (3, 38, 94, 106). Single-molecule fluorescence measurements of unfolding also point to gradual changes in specific polypeptide chain dimensions (54, 71).

If gradual folding does occur during the earliest steps of folding, or during the entire folding reaction of ultrafast-folding proteins, because of diffusive motion of the polypeptide chain, then the question is really whether such diffusive motion is restricted along one pathway. Some kinetic evidence for gradual structural transitions during folding is available. The initial collapse reaction during the folding of barstar appears to be a gradual structural transition (103, 104). In some cases the heterogeneity observed in the folding kinetics, when measured by multiple probes, has been interpreted as a signature of downhill folding. So too has been the observation of nonexponential kinetics (32, 93). But the assertion that probe-dependent and nonexponential kinetics must necessarily represent only downhill folding has been contested (34). Understanding the origin of nonexponential kinetics is important because an alternative explanation for such kinetics is that folding occurs via parallel pathways (19, 95).

Energy Landscape Theories and the Funnel View of Folding

The application of many different theoretical and computational methods, particularly the methods of statistical mechanics, to the study of protein folding reactions has had a major impact on how experimental scientists think about folding. A principal result is that folding occurs along a large multitude of folding routes traversing a multidimensional energy landscape in which an energy gradient favors the folded state (16, 17, 77). Statistical mechanical models of folding have led to the funnel representation of folding occurring via many parallel events (16). The funnel view of protein folding envisages an individual folding trajectory for each polypeptide chain. A simple quantitative model for folding kinetics captures many of the features of the funnel models: Folding proceeds through a continuum of intermediates, and there is no single sequential route (125).

Large sets of the folding trajectories, with common features, can be averaged into folding pathways. Each such macroscopic pathway would be distinguished by a specific progression of structural transitions between specific intermediates (80). The application of statistical mechanics has led to a more precise theoretical description of the concepts of folding pathways. Intermediates are not discrete; they are ensembles of structures (81) and the transition from one ensemble of structures to the next on the folding pathway happens on

parallel routes. Not surprisingly, multiple microscopic routes and macroscopic pathways are seen in many different simulations of folding and unfolding (13, 101, 105).

STRUCTURAL HETEROGENEITY IN PROTEIN FOLDING REACTIONS

Identifying Structural Heterogeneity

Whereas it is comparatively easy to identify structural heterogeneity in simulation studies, it has proven difficult to identify experimentally structural heterogeneity during protein folding. Folding intermediates accumulate only transiently and are consequently difficult to characterize. The important question is whether a folding intermediate represents only one macrostate and hence only one ensemble-averaged structure, or whether it represents several subpopulations of conformations coexisting with one another. If there are several subpopulations, there may be independent folding pathways, hence the importance of identifying structural heterogeneity in folding intermediates.

In most experimental studies, only one or two spectroscopic probes have been used to monitor folding, and the probes typically used, such as circular dichroism, fluorescence, and NMR, report on ensemble-averaged properties of all molecules present. A single probe cannot distinguish between subpopulations of molecules present together; therefore, its use can be misleading (101). When many probes have been used, for example, in the study of the slow folding reaction of barstar (8, 109), the heterogeneity of protein folding reactions can become apparent. Probes that can identify heterogeneity in a site-specific manner include multisite, time-resolved fluorescence resonance energy transfer (TRFRET) measurements of intramolecular distances, and proteolysis enabled, mass spectrometric measurements of structure affording protection to HX.

Multisite, TRFRET methods not only allow a distance-based and hence structure-based distinction to be made between different conformations when several conformations are present, but also allow each state to be quantified when the fluorescence decays are analyzed by the model-independent maximum entropy method (MEM) (55). The fluorescence lifetime distributions so obtained not only enable the distinction of subpopulations in an ensemble, but also provide a snapshot of the conformational heterogeneity in each subpopulation. In the first reported use of TRFRET coupled to MEM analysis, it could be shown that structure is lost in a continuously incremental manner during denaturant-induced equilibrium unfolding of barstar (55).

The pulsed-labeling HX-MS methodology (49) is the other method of choice for detecting coexisting subpopulations of different conformations. The conformations in each subpopulation afford different degrees of protection against HX. Their differential labeling by HX and hence different masses can be distinguished by electrospray ionization mass spectrometry. HX-MS experiments have shown that the folding of several proteins, including lysozyme (66), apomyoglobin (73), cytochrome *c* (122), interleukin-1β (37), and the α-subunit of tryptophan synthase (120), proceeds via the transient accumulation of an intermediate that is both obligatory and productive.

Unlike pulsed HX-MS, the pulsed HX-NMR method (5, 21, 45, 49) can easily provide residue-level structural information on folding intermediates. But it can determine only indirectly what fraction of molecules are structured at any time of folding and whether different fractions of molecules possess structures in different and/or similar regions. If a single sequential pathway is present, protection against HX at any individual amide site should develop in a single step during folding. If, instead, there is more than one kinetic phase, it suggests that there are subpopulations of molecules possessing unfolded structure at that site that fold at different rates (113).

FRET: fluorescence resonance energy transfer

Multisite, time-resolved FRET (TRFRET): a method in which FRET efficiency is determined from the measurement of fluorescence lifetime decays, and which is used to determine different intramolecular distances

MS: mass spectrometry

Pulsed HX-MS: a mass spectrometric-based method of distinguishing multiple coexisting subpopulations of molecules that differ in the number of amide hydrogens, protected by different extents of structure, from exchange labeling

Structural Heterogeneity in the Unfolded State

Subpopulations of kinetically distinct populations exist in the unfolded state. For some proteins, there are subpopulations distinguished by prolyl (117) or nonprolyl *cis* peptide bonds (75). Such subpopulations interconvert at rates much slower than those of most folding reactions, and hence each subpopulation folds independently of the other. For other proteins, subpopulations may differ by whether a specific nonnative interaction is present, and these subpopulations may also fold independently of each other if the interaction is broken slowly compared with folding (20). Yet for other proteins, there are subpopulations that differ in the presence or absence of residual structures, whether native-like or not (45, 67, 97, 102). Nonnative residual structures can hold up folding, whereas native residual structures can possibly speed up folding. If there are multiple residual structures that are kinetically competent to fold, parallel folding pathways may result.

When such subpopulations are absent, it is assumed that unfolded-state molecules are kinetically homogeneous, with all molecules interconverting rapidly compared with the rate of the folding reaction. But measurements of conformational fluctuations in proteins unfolded in chemical denaturants indicate that they occur in the 0.05- to 20-μs time domain (12, 35, 54, 71). Acid unfolded proteins often show slower conformational fluctuations (14), even in the millisecond time domain (65), which may disappear upon addition of chemical denaturant (98). The earliest steps in folding for many proteins occur in the 1- to 10-μs time domain (18), as do entire folding reactions for other proteins (24, 68). Hence, with respect to the fastest folding reactions, the unfolded state may be kinetically heterogeneous, and this kinetic heterogeneity can lead to the initial, fastest folding reactions proceeding in parallel.

Even if the unfolded state is kinetically homogeneous with very rapid interconversion (a pre-equilibrium) between unfolded conformations, it seems that the initial ultrafast folding reactions must occur via multiple pathways. If only one or a few folding tracks from the unfolded state exist, it implies that only a very small fraction of the astronomically large number of conformations is competent to fold. Then it is expected that the multitude of fast conformational changes in the unfolded state would lead to the initial folding rates being slowed down drastically to rates lower than those observed for the fastest folding reactions (12, 19).

Structural Heterogeneity in Early-Collapsed Forms

The earliest intermediates in which structural heterogeneity has been detected have been the collapsed forms seen to accumulate at a few milliseconds of folding. This heterogeneity has manifested itself not as many different structural forms, but instead as a few coexisting subpopulations. In the case of cytochrome *c*, the early (millisecond) intermediate (99) has subpopulations of molecules with and without a misligated heme, but the specific secondary structure present is different in the two subpopulations (2). In the case of ribonuclease A, the early (millisecond) intermediate contains both structured and unstructured subpopulations (39). For apomyoglobin, the structural origin of the heterogeneity could be identified (72, 73). For barstar, the structural composition of the early (millisecond) intermediate depends on the conditions in which folding is carried out (86, 87). Various subpopulations in the intermediate are stabilized differentially in different folding conditions. Structural heterogeneity in the early (millisecond) intermediates has also been detected in the case of lysozyme and thioredoxin (7, 27).

Understanding the structural heterogeneity in the early- (millisecond) collapsed intermediates is important because for several proteins, kinetic partitioning into different folding pathways happens at this stage

(15, 83, 88, 100, 119). Structural fluctuations in the equilibrium analogs of such early intermediates are slower than those in fully unfolded proteins (14) but much faster than the subsequent folding step. It appears that there must be specific structures present in the subpopulations of the intermediate, as identified in the case of apomyoglobin (72), that direct folding along different routes.

Multisite TRFRET experiments have shown that multiple populations of distinct conformations are present at a few milliseconds during the folding of cytochrome *c* (85) and differ in the extent to which different intramolecular distances have contracted. Molecules with one or more distances extended could be observed throughout the folding process, but all distances do not appear to remain extended in any molecule, i.e., completely unfolded molecules disappear early during folding. Similar results were obtained when folding was carried out under conditions in which misligation to the heme was allowed and not allowed to occur. Hence, such an optional barrier does not appear responsible for the multiple intermediate conformations and the consequent multiple refolding pathways that are suggested by this structural heterogeneity.

Structural Heterogeneity in Transition States

ϕ-value analysis, in which energetic interactions of a suitably mutated side chain in the TS are compared to those in the native state, is used routinely to map the fate of individual side chains in the TS of folding (23, 24). ϕ-values usually range between 0 (unfolded-like TS) and 1 (native-like TS). The meaning of ϕ-values between 0 and 1, which have been commonly observed for most of the proteins studied to date (30), remains uncertain. Such partial ϕ-values are invariably interpreted in terms of partial structure formation in the TS on a single sequential pathway. Partial ϕ-values can also arise if the TS is an ensemble of multiple structural forms, which are presumably formed on parallel pathways (22). ϕ-values less than 0 and more than 1 are also observed, although not too infrequently, and such ϕ-values may arise when folding molecules switch from one folding channel to another (79).

It is as difficult to demonstrate experimentally that the TS comprises distinct structural ensembles as it is to show that there is only one ensemble. Several experimental results support the latter possibility (22), but other results are suggestive of a structurally heterogeneous TS. ϕ-value analysis of circularly permuted variants of a SH3 domain (114) and of S6 (60) indicated that differently structured TSs, and hence different folding pathways, can lead to the same structure from the same protein sequence. These results are striking because proteins in the same family with high sequence identity appear to have a similarly structurally restricted TS (63), although not always (124). More distantly related members of the same family appear to follow structurally distinct folding pathways (74). When the TS appears to have diffuse structure, mutational changes in the sequence, or changes in the order of secondary structure in the sequence, can lead to different polarized regions of the diffuse structure serving as the TS. This would lead to the TS for the folding of differently perturbed sequences being dissimilar and hence their folding pathways being distinct (76).

ϕ-value analysis: a method for determining the presence of a stabilizing interaction in the TS by determining whether a perturbing mutation has the same effect on the energy of the TS as it has on that of the native protein

Structural Heterogeneity in Folding Intermediates

The late-folding intermediate ensemble, I_L, of barstar was the first intermediate to have its structural heterogeneity characterized by the multisite, TRFRET approach coupled with MEM analysis (110). Four different intramolecular distances within the structural components of I_L were measured at the same time of folding. The structure of I_L was examined under conditions that confer different stabilities, by changing the concentration of urea in which the protein was folded

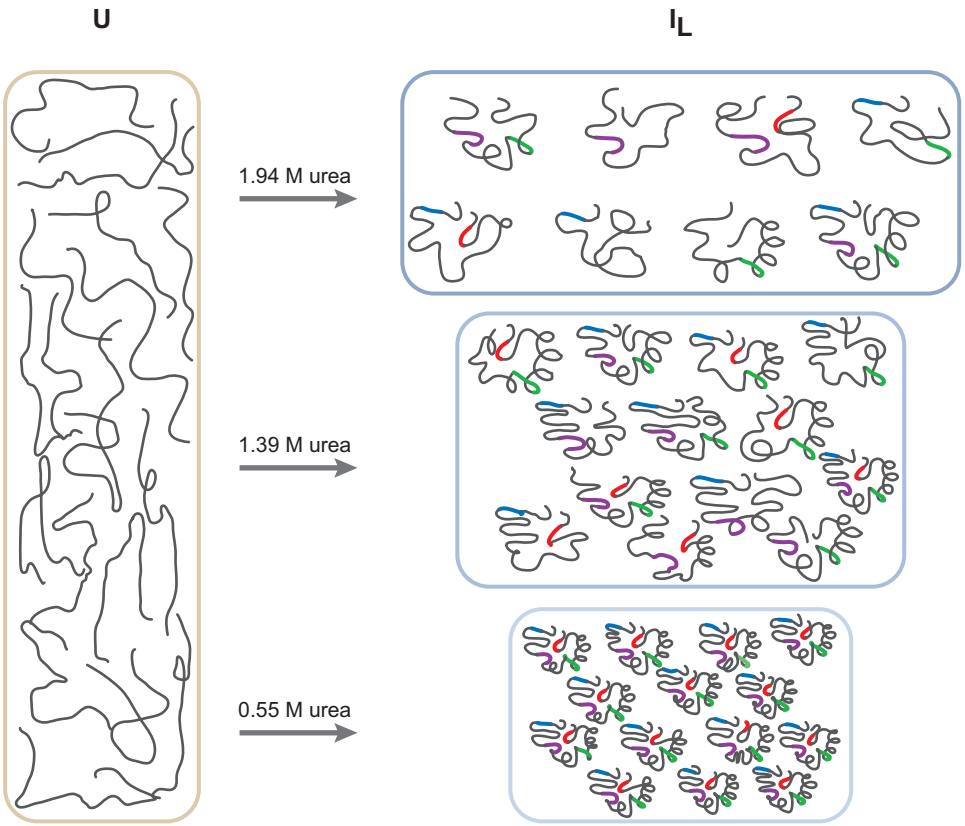

Figure 1

An experimental folding funnel. The late-folding intermediate, I_L, of barstar appears to be structurally heterogeneous. Four intramolecular distances were measured within the population of I_L in different concentrations of urea (110). A native-like distance is shown as a colored line. When the protein is folded in strongly stabilizing conditions, most of the protein molecules have all four distances similar to those in the native state. In marginally stabilizing conditions, protein molecules have one, two, three, or four distances that are native-like. The results indicate that available conformational space, as represented by structural heterogeneity in I_L, becomes restricted in more stable conditions. This figure is adapted from Reference 110.

(**Figure 1**). As stability is decreased, the population of molecules with unfolded-like distances increases: More distances in more of the molecules become unfolded-like. Thus, the results demonstrated that even a late intermediate such as I_L consists of subpopulations of different structural forms, and that the conformational entropy of the intermediate decreases as structures become more stable. This finding suggests that not only is there a reduction in the extent to which the intermediate ensemble is populated, but there is also a change in the structural composition of the intermediate ensemble. Under some conditions the more structured members of the intermediate ensemble are preferentially populated; under other conditions the less structured members are preferentially populated. This fundamental result implies that the folding pathway observed for a given protein will appear different under different conditions.

Pulsed HX-MS in conjunction with low pH proteolysis could identify site-specific differences in structure and hence structural

heterogeneity during the folding of cytochrome *c*. Coexisting subpopulations of intermediates, which differed in their structure, could be identified, and three different folding pathways were identified (122). What was particularly intriguing about the results was that intermediates structured in one region appeared to coexist with intermediates unstructured in the same region but structured elsewhere at the same time of folding. Similar results were obtained using the TRFRET approach with yeast cytochrome *c*. In the case of the α-subunit of tryptophan synthase, coexisting populations of intermediates differing in HX-protective structure could be identified (120), in accordance with earlier studies that had indicated multiple intermediates and parallel folding pathways (115).

In the case of ribonuclease A, a pulsed HX-NMR study had not only shown the accumulation of an obligatory folding intermediate, but also suggested that the structure of the intermediate consolidates as folding proceeds (113). The protection against HX at any amide site, afforded by the intermediate, was found to increase continuously from the time it is first observed to the time it is fully stabilized. This result indicates that structural changes leading to the stabilization of the intermediate occur continuously during this time. Gradual structural changes may not be uncommon during the folding of proteins.

In pulsed HX-NMR studies with ribonuclease A (113), lysozyme (88), and apomyoglobin (72), by varying the pH as well as the duration of the labeling pulse, the multiphasic nature of the observed kinetics was shown to not be the result of partial labeling of an intermediate. In other words, one subpopulation of protein molecules had not transformed into a folding intermediate when another subpopulation had transformed completely. For cytochrome *c*, the coexistence of three subpopulations of folding molecules, when folding was carried out under conditions in which heme misligation did not occur, could be surmised from a pulsed HX-NMR study (20) and confirmed by a pulsed HX-MS study (122).

One explanation suggested for the presence of different subpopulations of folding molecules, seen for many proteins (5), is that folding occurs on at least two pathways, one slow and one fast (113). Another explanation suggested that a transient barrier prevents the second population from folding and that once the barrier was surmounted, this second subpopulation enters the same sequential folding pathway (113). On the basis of many elegant studies of the folding of cytochrome *c*, (21, 107), it was subsequently suggested that the presence of an optional barrier to the folding of a subpopulation of protein molecules might be a general feature of many protein folding reactions (52).

Finally, ^{19}F NMR is also useful for detecting heterogeneity in protein folding reactions. In studies of the folding of a slowly folding mutant form of the intestinal fatty acid binding protein, multiple conformations could be detected early during folding and an intermediate could be shown to experience multiple conformations (58).

Reduction of Heterogeneity and the Models of Folding

Given that the unfolded state is structurally heterogeneous, and that the native state is a unique state, when is heterogeneity reduced during folding? Is it reduced in many consecutive steps, or is it reduced in one initial step? Different models for folding (**Figure 2**) envisage different scenarios. In nucleation models (23), a folding nucleus, extended or otherwise, forms initially, perhaps in the unfolded protein itself, and the nucleus acts as a scaffold for the rest of the structure to build upon. In such models, therefore, much of the reduction in structural heterogeneity is expected to occur as a first step. In other hierarchical models, such as the framework model (6), the reduction in structural heterogeneity is expected to occur throughout the folding process. In hydrophobic collapse models (1), a nonspecific hydrophobic collapse is expected to occur as the first step, and it is expected

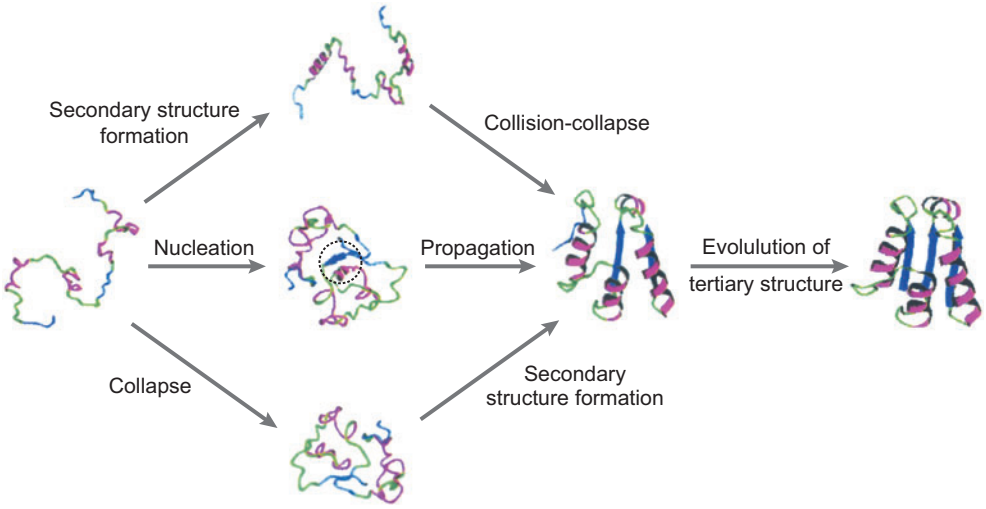

Figure 2

Pathways of protein folding. In the framework model, precedence is given to the formation of secondary structural units. In the hydrophobic collapse model, precedence is given to an initial chain collapse. In the nucleation-condensation model, an extended nucleus is formed early during folding. Molten globule-like intermediates accumulate during the folding of many proteins. For some proteins, particularly those following nucleation mechanisms, a molten globule intermediate does not usually accumulate. Although different folding pathways are usually discussed in the context of different proteins, can a single protein utilize fundamentally different folding pathways in different folding conditions? For example at very low temperatures, at which hydrophobic interactions are weakened, a protein could conceivably switch from a hydrophobic collapse to a framework mechanism.

that such a collapse will lead to a heterogeneous, loosely compact state with many different types of nonnative interactions between different segments of the polypeptide chain. In such a model, therefore, some of the reduction in heterogeneity may happen later during folding.

KINETIC STUDIES

During the folding of even a small protein, many hundreds of noncovalent interactions form, starting from an unfolded state that is extremely heterogeneous. It is therefore not surprising that folding kinetics are often complex, and it is not easy to determine experimentally whether the steps that give rise to this complexity occur on sequential or parallel folding pathways, or a combination thereof (115). Exponential kinetics, such as those usually seen in folding and unfolding studies, arise from individual protein molecules behaving in a random manner. But single exponential kinetics do not necessarily imply a single route (19). Single exponential kinetics are seen when there is one route from one homogeneous starting state, or when there are many parallel routes. Multiple exponential kinetics for the formation of N is are seen if there are multiple unfolded states that interconvert slowly compared to the folding rate. When complex kinetics are observed, it is important to use minimal models to analyze them; otherwise, models can take a life of their own (41).

Parallel Pathways in Folding

It is instructive to consider the logic of the steps that led to two competing pathways proposed for describing the folding of lysozyme, a two-domain protein. Early work had established the presence of a collapsed intermediate at a few milliseconds of folding, and the observation of two subsequent kinetic

phases, to which the initial collapse reaction is uncoupled, indicated the formation of a folding intermediate, I, during folding (53). Pulsed HX-NMR as well as HX-MS experiments (66, 88) indicated that I contains native-like α-helical structure in one domain. Interrupted unfolding experiments showed that there are no optically silent slow equilibration reactions in the unfolded state (U) that might be responsible for the observed kinetics (119). Interrupted refolding experiments distinguish and quantify native protein and intermediate at different times of folding, on the basis of the differences in their unfolding kinetics. These experiments, as well as inhibitor binding experiments, showed that native protein (N) forms in two kinetic phases (64). Both sequential off-pathway ($I \rightleftharpoons U \rightleftharpoons N$) and on-pathway ($U \rightleftharpoons I \rightleftharpoons N$) mechanisms could be ruled out (119). A triangular mechanism, in which kinetic partitioning into two competing pathways occurs once the initial collapse reaction occurs, and in which the α and β domains fold independently of each other, can account for the data. More recently, the data for lysozyme have been reinterpreted in terms of a single sequential pathway, $U \rightleftharpoons I \rightleftharpoons N$, but with I also forming a misfolded dead-end intermediate, I_X (52). But this mechanism is not minimal in that it is four-state, and only two kinetic phases (not three, as expected for a four-state mechanism) are observed in the kinetic studies.

The folding mechanism of lysozyme becomes more complex both at higher pH and at higher salt concentration (9). Additional pathways become operative. For the apparently two-state folding protein S6, an additional competing folding pathway defined by a collapsed intermediate also becomes operative at high salt concentration, in addition to the direct $U \rightarrow N$ pathway seen at low salt concentration (78).

Competing pathways have also been proposed for the folding of barstar (100). Unfolded barstar exists in two subpopulations, U_F (30%) and U_S (70%), that differ in their configuration of the Tyr47-Pro48 bond. In marginally stabilizing conditions, folding occurs via the $U_S \rightleftharpoons U_F \rightleftharpoons N$ pathway. As folding conditions are made more stabilizing, the $U_S \rightleftharpoons I_E \rightleftharpoons I_N \rightleftharpoons N$ pathway starts competing with the $U_S \rightleftharpoons U_F \rightleftharpoons N$ pathway until eventually all U_S molecules fold via the $U_S \rightleftharpoons I_E \rightleftharpoons I_N \rightleftharpoons N$ pathway. This happens because stabilization of the early- and late-folding intermediates, I_E and I_N, in more native-like conditions makes this pathway operate faster. In even more stabilizing folding conditions, a third pathway for the folding of U_S becomes operative. This happens because I_E consists of at least two subpopulations (I_{M1} and I_{M2}) of molecules, and I_{M2} is stabilized more than I_{M1} in strongly stabilizing conditions. Hence the $U_S \rightleftharpoons I_{M2} \rightleftharpoons N$ pathway is utilized more than the $U_S \rightleftharpoons I_{M1} \rightleftharpoons N$ pathway. This could be discerned because both I_{M1} and I_{M2} bind a hydrophobic dye, and the dye is kicked off the folding protein molecules in two different kinetic steps. These two steps appeared to be on two competing pathways because the relative amplitude of the fast step increases at the expense of the slow step as folding conditions become strongly stabilizing. The folding mechanism of high pH-unfolded barstar is remarkably similar to that of the guanidine hydrochloride (GdnHCl)-unfolded protein (91), with three folding pathways. The relative utilization of the three pathways changes, with a change in the pH at which folding is carried out, presumably because the stabilities of the intermediates and TSs change with pH.

It is remarkable how little attention is paid to slow folding reactions because they are thought to represent local structural changes such as proline isomerization. These reactions, however, are coupled kinetically to the main folding reactions, and hence the observed slow rate constants can also provide information on the microscopic rate constants of the faster preceding reactions (8, 109). Following this approach, more than a dozen probes were used to study the slow folding and unfolding reactions of barstar. By examining the differences in the kinetics observed

U_F: unfolded protein that refolds fast

U_S: unfolded protein that refolds slowly

GdnHCl: guanidine hydrochloride

using many different probes, the presence of multiple pathways for folding as well as unfolding could be established.

The importance of interrupted folding experiments in demonstrating the existence of multiple folding pathways for many proteins cannot be overemphasized. For some proteins, multiple pathways originate because of prolyl peptide bond isomerization in the unfolded state (7, 10, 27, 39, 48, 115, 120), whereas for other proteins, multiple pathways originate because of nonprolyl peptide bond isomerization in the unfolded state (75, 82). For several proteins, neither type of peptide bond isomerization is responsible for creating multiple parallel pathways (28, 42, 46). For some proteins, the multiple pathways originate at later stages in folding, and not only in the unfolded state (10, 83), indicating that prolyl peptide bond isomerization is not responsible.

Optional errors and multiple folding pathways. An unresolved question in protein folding studies is whether two subpopulations of unfolded protein molecules such as U_F and U_S, which are in a proline isomerization-limited slow equilibrium with each other, nevertheless have structurally equivalent folding pathways. U_S molecules begin folding with the nonnative configuration of an X-Pro bond, but it is possible that the same sequence of events occurs on the U_S pathway and on the U_F pathway. In the case of barstar, U_S and U_F molecules begin folding with the same apparent rate constant (100); the U_S folding pathway is slower only because it has a final, slow step that is absent from the U_F folding pathway. In contrast, the U_S and U_F molecules of ribonuclease A begin folding at vastly different rates (26). In the case of barstar there are fundamental differences between the pathways of U_F and U_S: 8-anilino-1-naphthalene sulfonic acid (ANS) binding does not occur during the folding of U_F whereas it does on both pathways originating from U_S. For ribonuclease A, pulsed HX-NMR studies show different extents of structure formation in the earliest intermediates on the U_{VF} folding pathway, which has all native-like X-Pro bonds, and the U_S folding pathway, which has one non-native-like X-Pro bond (39, 113).

In the case of cytochrome *c*, a folding barrier arises owing to a pH-dependent misligation of the heme by His33. It is important to determine whether the subpopulation of protein molecules whose folding is slowed down by such a transient barrier utilizes the same folding pathway as does the subpopulation of protein molecules for which the barrier is absent. The folding of GdnHCl-unfolded cytochrome *c* is described by the pathway $U \leftrightharpoons I_1 \rightleftharpoons I_2 \rightleftharpoons N$, and the observed rates of formation of I_1, I_2, and N do not change, even when the fraction of molecules with heme misligation is varied by changing the pH (2). Misligation of the heme affects the secondary structure of I_2 (2). Pulsed HX-MS studies indicated that only protein molecules with the heme correctly ligated form structured N- and C-terminal helices in I_2 (122). The subpopulation of protein molecules with no error first form structured N- and C-terminal helices and later form structure in another sequence segment. On the other hand, most of the protein molecules in the subpopulation of molecules with an error in heme ligation appear to form structured N- and C-terminal helices concurrently with structure in the other sequence segment. Hence, it appears that the subpopulations of molecules with and without the misligated heme fold on fundamentally different pathways.

Folding pathways in urea and GdnHCl. Urea and GdnHCl unfold proteins because they preferentially interact with the unfolded over the folded state. The preferential free energy of the interaction of GdnHCl is usually more than twofold higher than that of urea. Because of this and because GdnHCl, but not urea, is charged, it is likely that there might be differences in the folding pathways of urea- and GdnHCl-unfolded proteins. Surprisingly, the folding pathways of very

few urea-unfolded and GdnHCl-unfolded proteins have been compared directly. Barstar is one such protein (111) for which the folding of U_F, populated fully by transient unfolding, was compared in urea and GdnHCl solutions. The degree to which nonpolar surface is buried in the TS compared with that in N is 2.3-fold less for folding in GdnHCl than for folding in urea. This means that the TS for the major folding step has vastly different structures in the two denaturants, with the structure in urea far more N-like than the structure in GdnHCl. Nevertheless, the TSs in the two denaturants are equal in energy. If two TSs are equal in energy, yet different in structure, it is likely that they are on different folding pathways. Hence, a protein folding in urea may use a pathway fundamentally different from what it uses for folding in GdnHCl.

The initial folding reaction. It is not too difficult to achieve the temporal resolution required to study the earliest folding reactions (18, 68), but it is still difficult to couple fast temporal resolution with residue-level structural resolution. The only probes used so far in ultrafast measurements of folding have been those that report on gross structure, and very few studies have used more than one gross structural probe. Nevertheless, there are indications that parallel folding reactions do occur on the 10-μs timescale (29, 32, 33).

In a recent study, 11 different intramolecular distances were measured by FRET in I_E, which accumulates at a few milliseconds during the folding of barstar (104). I_E is the product of an initial hydrophobic collapse followed by structural changes in the submillisecond time domain, and the structure of I_E depends on the folding conditions (1, 86, 87). The multisite FRET-enabled measurement of 11 intramolecular distances (**Figure 3**) showed that in the transition from U to I_E structure formation is not synchronized across different regions of the protein. The high degree of site-specific heterogeneity strongly suggests that I_E is formed by diffusive motions of the polypeptide chain along multiple routes, but direct submillisecond measurements of these site-specific changes are still needed to confirm this result.

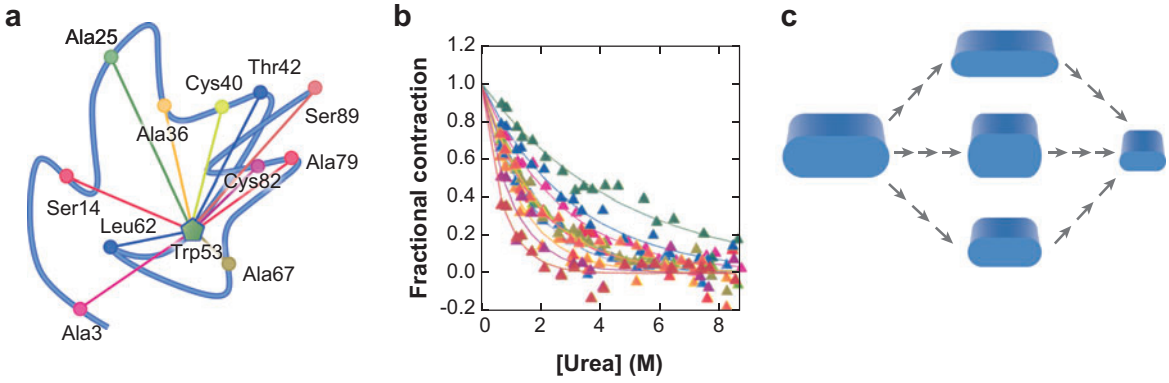

Figure 3

Heterogeneity in an early-collapse reaction. A specific intermediate, I_E, is populated at a few milliseconds of folding (104). (*a*) Eleven intramolecular distances in I_E, each shown by a colored line, were measured by FRET. (*b*) The 11 distances decrease gradually and asynchronously with a decrease in the urea concentration in which folding was carried out. (*c*) A multiple pathway scenario for the formation of I_E. Distance is shown as a measure of specific structure. In the top pathway, one distance contracts first, and then the other. In the middle pathway, the second distance contracts first. In the bottom pathway, both distances contract concurrently. All distances contract in a gradual manner.

Parallel Pathways in Unfolding

A clear example of a protein unfolding via two competing pathways is that of a titin domain (121). The pathway with a more compact TS is utilized at lower denaturant concentrations, and the pathway with a less-structured TS is utilized at higher denaturant concentrations. The switch in the utilization of unfolding pathways occurs because the less-structured TS with greater solvent-accessible surface area is more stabilized at higher denaturant concentration, as expected (22). The switch in pathways was detected because it led to an upward curvature in the denaturant-dependent unfolding kinetics: the unfolding rate at high denaturant concentrations was faster than expected.

Upward curvatures in the nonlinear free energy relationships for the unfolding of proteins are rare. It is possible that parallel unfolding pathways are absent for most proteins. But the presence of intermediates on the pathways, which causes a downward curvature (43, 123) in the free energy relationship for unfolding, is likely to compensate for any upward curvature. In unfolding studies of barstar, no upward curvatures were observed, but different structural probes yielded different unfolding rates, and the dependences of the observed unfolding rates on denaturant concentrations are dissimilar for the diverse probes in some but not all unfolding conditions. These results strongly suggest that unfolding occurs on two competing unfolding pathways (123). Additional kinetic unfolding experiments, including those using a pulsed sulfhydryl-labeling methodology (89), confirmed this interpretation. High pH-induced unfolding also appears to occur via at least two pathways (91). Measurement of the unfolding rates under refolding conditions, by coupling the unfolding reaction to a sulfhydryl-labeling reaction, indicated not only the existence of unfolding intermediates, but also that the unfolding pathways in urea and GdnHCl are different (111). Finally, multi-site FRET experiments indicated that different regions of the protein unfold independently of each other, again implying that unfolding occurs via parallel unfolding pathways (112).

Parallel unfolding pathways operate for other proteins too. Pulsed HX-NMR studies of the unfolding of ribonuclease A indicate that the protein unfolds via two competing pathways, one of which presents a partially unfolded intermediate ensemble that also exists in slow equilibrium with the native protein (44). Elegant HX-MS studies of the unfolding of the ovomucoid third domain (4) indicate a manifold of unfolding and partial unfolding reactions. These results also bring into focus the question of whether similar partially unfolded forms of other proteins, identified by native-state HX-NMR (21, 45), also form via parallel unfolding reactions rather than a sequential reaction. This question is particularly important in the case of cytochrome *c* because these partially unfolded forms have been placed in a sequential series of events that occur after the rate-limiting step in folding (21).

Finally single-molecule experiments have also indicated that a protein can unfold in multiple ways starting at different points in the structure. Mechanical unfolding experiments with a fibronectin module (59), ubiquitin (11), and enhanced yellow fluorescent protein (84) indicate that unfolding occurs via multiple pathways. They can also identify the different structural events that distinguish the parallel pathways.

LOOKING BACK AND LOOKING AHEAD

Until recently, the identification of multiple folding routes and structural heterogeneity at the level of folding subpopulations was limited by the restricted number of structural probes available or utilized to investigate protein folding reactions. Now with the increased application of many different probes to

investigate folding and unfolding reactions, the heterogeneity inherent in protein folding reactions will become even more apparent. The nature of the heterogeneity that will unfold will test basic tenets of energy landscape theories of folding. In vitro studies of folding are now uncovering the possible ways in which any one protein sequence can fold, a few of which will be used in the cell depending on the conditions prevalent at different times.

The identification of multiple folding pathways leads to the question, What may cause a protein to switch from utilizing one pathway to utilizing another? Switching can happen upon a change in conditions prior to the event (56) or during the event. A change in folding conditions can differentially affect the stabilities of different structures in an intermediate ensemble (86, 91, 110) and can lead to a change in folding pathways. Ligand binding can do the same (51). The intrinsic stabilities of substructures in the protein play an important role, and folding pathways can be changed or switched by mutations that affect their relative stabilities (25, 62, 70). The utilization of a specific pathway may depend on many different factors in the cell; for example, interactions with chaperones may channel folding along one route when many routes are available (7). Evolution has ensured that folding is robust by providing alternative folding routes, and it will be important in future studies to understand what features of protein sequences allow any one protein to fold in different ways under different folding conditions.

SUMMARY POINTS

1. Multiple pathways may be available for protein folding and unfolding.
2. Different pathways may be utilized differentially under different folding conditions.
3. Folding may switch between alternative pathways upon a change in sequence or folding conditions.
4. Folding pathways are defined by progressively more structured folding intermediates and, perhaps in some cases, by a continuum of intermediates.
5. A folding intermediate is an ensemble of molecules with an ensemble-averaged structure.
6. The structural composition of an intermediate or TS ensemble may change under different folding conditions or upon mutation.
7. Folding intermediates may consist of subpopulations that have different structures but similar energies.
8. Subpopulations of molecules may fold independently of each other.

FUTURE ISSUES

1. How many different ways are there for a protein to fold? A few or many?
2. How does sequence determine the choice of a folding pathway?
3. At what stage during folding is structural heterogeneity in a folding intermediate lost?
4. Does a subpopulation of molecules possessing nonnative interactions or optional folding errors fold via the same pathway as a subpopulation that does not?

5. What are the new methods that need to be developed and applied for identifying and quantifying heterogeneity in protein folding reactions, both at the level of multiple routes and multiple subpopulations of molecules in intermediates?

DISCLOSURE STATEMENT

The author is not aware of any biases that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

I thank past and present members of my laboratory, whose work has shaped my thinking about protein folding. I also thank many other protein folding colleagues elsewhere, with whom I have had discussions about protein folding; especially because space limitations have made it impossible for me to acknowledge all their publications relevant to this review. I thank M.K. Mathew, and my students, for comments on the manuscript and assistance in preparing it.

LITERATURE CITED

1. Agashe VR, Shastry MC, Udgaonkar JB. 1995. Initial hydrophobic collapse in the folding of barstar. *Nature* 377:754–57
2. Akiyama K, Takahashi S, Ishimori K, Morishima I. 2000. Stepwise formation of α-helices during cytochrome *c* folding. *Nat. Struct. Biol.* 7:514–20
3. Ahmed Z, Beta IA, Mikhonin AV, Asher SA. 2005. UV-resonance Raman thermal unfolding study of Trp-cage shows that it is not a simple two-state miniprotein. *J. Am. Chem. Soc.* 127:10943–50
4. Arrington CB, Robertson AD. 2000. Correlated motions in native proteins from MS analysis of NH exchange: evidence for a manifold of unfolding reactions in ovomucoid third domain. *J. Mol. Biol.* 200:221–32
5. Baldwin RL. 1995. The nature of protein folding pathways: the classical versus the new view. *J. Biomol. NMR* 5:103–9
6. Baldwin RL, Rose GD. 1999. Is protein folding hierarchic? II. Folding intermediates and transition states. *Trends Biochem. Sci.* 24:77–83
7. Bhutani N, Udgaonkar JB. 2001. GroEL channels the folding of thioredoxin along one kinetic route. *J. Mol. Biol.* 314:1167–79
8. Bhuyan AK, Udgaonkar JB. 1999. Observation of multistate kinetics during the slow folding and unfolding of barstar. *Biochemistry* 38:9158–68
9. Bieri O, Wildegger G, Bachmann A, Wagner C, Kiefhaber T. 1999. A salt-induced kinetic intermediate is on a new parallel pathway of lysozyme folding. *Biochemistry* 38:12460–70
10. Brujic J, Hermans RIZ, Garcia-Manyes S, Walther KA, Fernandez JM. 2007. Dwell time distribution analysis of polyprotein unfolding using force clamp spectroscopy. *Biophys. J.* 92:2896–903
11. Butler JS, Loh SN. 2005. Kinetic partitioning during folding of the p53 DNA binding domain. *J. Mol. Biol.* 350:906–18

12. Chattopadhyay K, Elson EL, Frieden C. 2005. The kinetics of conformational fluctuations in an unfolded protein measured by fluorescence methods. *Proc. Natl. Acad. Sci. USA* 102:2385–89
13. Chavez LL, Gosavi S, Jennings PA, Onuchic JN. 2006. Multiple routes lead to the native state in the energy landscape of the beta-trefoil family. *Proc. Natl. Acad. Sci. USA* 103:10254–58
14. Chen H, Rhoades E, Butler JS, Loh SN, Webb WW. 2007. Dynamics of equilibrium structural fluctuations of apomyoglobin measured by fluorescence correlation spectroscopy. *Proc. Natl. Acad. Sci. USA* 104:10459–64
15. Crespo MD, Simpson ER, Searle MS. 2006. Population of on-pathway intermediates in the folding of ubiquitin. *J. Mol. Biol.* 360:1053–66
16. Dill KA, Chan HS. 1997. From Levinthal to pathways to funnels. *Nat. Struct. Biol.* 4:10–19
17. Dinner AR, Sali A, Smith LJ, Dobson CM, Karplus M. 2000. Understanding protein folding via free-energy surfaces from theory and experiment. *Trends Biochem. Sci.* 25:331–39
18. Eaton WA, Munoz V, Hagen SJ, Jas GS, Lapidus LJ, et al. 2000. Fast kinetics and mechanisms in protein folding. *Annu. Rev. Biophys. Biomol. Struct.* 29:327–59
19. **Ellison PA, Cavagnero S. 2006. Role of unfolded state heterogeneity and en-route ruggedness in protein folding kinetics. *Protein Sci.* 15:564–82**
20. Elove GA, Bhuyan AK, Roder H. 1994. Kinetic mechanism of cytochrome *c* folding: involvement of the heme and its ligands. *Biochemistry* 33:6925–35
21. **Englander SW. 2000. Protein folding intermediates and pathways studied by hydrogen exchange. *Annu. Rev. Biophys. Biomol. Struct.* 29:213–38**
22. Fersht AR, Itzhaki LS, ElMarry NF, Matthews JM, Oltzen DE. 1994. Single versus parallel pathways of protein folding and fractional formation of structure in the transition state. *Proc. Natl. Acad. Sci. USA* 91:10426–29
23. Fersht AR, Daggett V. 2002. Protein folding and unfolding at atomic resolution. *Cell* 108:573–82
24. Fersht AR, Sato S. 2004. φ-value analysis and the nature of protein folding transition states. *Proc. Natl. Acad. Sci. USA* 101:7976–81
25. Garcia C, Nishimura C, Cavagnero S, Dyson HJ, Wright PE. 2000. Changes in the apomyoglobin pathway caused by mutation of the distal histidine residue. *Biochemistry* 39:2894–901
26. Garel JL, Baldwin RL. 1973. Both the fast and slow refolding reactions of ribonuclease A yield native enzyme. *Proc. Natl. Acad. Sci. USA* 70:3347–51
27. Georgescu RE, Lee JH, Goldberg ME, Tasayko ML, Chaffotte AF. 1998. Proline isomerization-independent accumulation of early intermediate and heterogeneity of the folding pathways of a mixed alpha/beta protein, *Escherichia coli* thioredoxin. *Biochemistry* 37:10286–97
28. Gianni S, Travaglini-Allocatelli C, Cutruzolla F, Brunori M, Shastry MC, Roder H. 2003. Parallel pathways in cytochrome *c*551 folding. *J. Mol. Biol.* 330:1145–52
29. Goldbeck RA, Tomas YG, Chen E, Esquerra RM, Kliger DS. 1999. Multiple pathways on a protein-folding energy landscape: kinetic evidence. *Proc. Natl. Acad. Sci. USA* 96:2782–87
30. Goldenberg DP. 1999. Finding the right fold. *Nat. Struct. Biol.* 6:987–90
31. Gorski SA, Capaldi AP, Kleanthous C, Radford SE. 2001. Acidic conditions stabilize intermediates populated during the folding of Im7 and Im9. *J. Mol. Biol.* 312:849–63

19. A lucid presentation of the effects of structural heterogeneity on the kinetics of folding.

21. A thought-provoking review of the application of HX methods to the study of protein folding.

32. Gruebele M. 2005. Downhill protein folding: evolution meets physics. *C. R. Biol.* 328:701–12
33. Gulotta M, Rogatsky E, Callender RH, Dyer RB. 2003. Primary folding dynamics of sperm whale apomyoglobin: core formation. *Biophys. J.* 84:1090–118
34. Hagen SJ. 2007. Probe-dependent and nonexponential relaxation kinetics: unreliable signatures of downhill protein folding. *Proteins* 68:205–17
35. Hagen SJ, Hofrichter A, Szabo A, Eaton WA. 1997. Diffusion-limited contact formation in unfolded cytochrome *c*: estimating the maximum rate of protein folding. *Proc. Natl. Acad. Sci. USA* 93:11615–17
36. **Harrison SC, Durbin R. 1985. Is there a single pathway for the folding of a polypeptide chain? *Proc. Natl. Acad. Sci. USA* 82:4028–30**
37. Heidary DK, Gross LA, Roy M, Jennings PA. 1997. Evidence for an obligatory intermediate in the folding of interleukin-1β. *Nat. Struct. Biol.* 4:725–31
38. Holtzer ME, Lovett EG, D'Avignon DA, Holtzer A. 1997. Thermal unfolding in a GCN4-like leucine zipper: $^{13}C^{\alpha}$ NMR chemical shifts and local unfolding curves. *Biophys. J.* 73:1031–41
39. Houry WA, Scheraga HA. 1996. Structure of a hydrophobically collapsed intermediate on the conformational folding pathway of ribonuclease A probed by hydrogen-deuterium exchange. *Biochemistry* 35:11734–46
40. Jackson SE. 1998. How do small single-domain proteins fold? *Fold. Des.* 3:81–91
41. Jencks WP. 1989. How does the calcium pump pump calcium? *J. Biol. Chem.* 264:18855–58
42. Jennings PA, Finn BE, Jones BE, Matthews CR. 1993. A reexamination of the folding mechanism of dihydrofolate reductase from *Escherichia coli*: verification and refinement of a four-channel model. *Biochemistry* 32:3783–89
43. Jonsson T, Waldburger CD, Sauer RT. 1996. Non-linear free energy relationships in Arc repressor unfolding imply the existence of unstable, native-like folding intermediates. *Biochemistry* 35:4795–802
44. Juneja J, Udgaonkar JB. 2002. Characterization of the unfolding of ribonuclease A by a pulsed hydrogen exchange study: evidence for competing pathways for unfolding. *Biochemistry* 41:2641–54
45. Juneja J, Udgaonkar JB. 2003. NMR studies of protein folding. *Curr. Sci.* 84:157–72
46. Kamagata K, Sawano Y, Tanokura M, Kuwajima K. 2003. Multiple parallel-pathway folding of proline-free staphylococcal nuclease. *J. Mol. Biol.* 332:1143–53
47. Karplus M, Weaver DL. 1976. Protein-folding dynamics. *Nature* 260:404–6
48. Kiefhaber T, Quaas R, Hahn U, Schmid FX. 1990. Folding of ribonuclease T1.2. Kinetic models for the folding and unfolding reactions. *Biochemistry* 29:3061–70
49. Konermann L, Simmons DA. 2003. Protein folding kinetics and mechanisms studied by pulse labeling and mass spectrometry. *Mass Spectrom. Rev.* 22:1–26
50. Korzhnev DM, Salvatella X, Vendruscolo M, Di Nardo AA, Davidson AR, et al. 2004. Low populated folding intermediates of Fyn SH3 characterized by relaxation dispersion NMR. *Nature* 430:586–90
51. Krantz BA, Sosnick TR. 2001. Engineered metal binding sites map the heterogeneous folding landscape of a coiled coil. *Nat. Struct. Biol.* 8:1043–47
52. Krishna MM, Englander SW. 2007. A unified mechanism for protein folding: predetermined pathways with optional errors. *Protein Sci.* 16:449–64
53. Kuwajima K, Hiraoka Y, Ikeguchi M, Sugai S. 1985. Comparison of the transient folding intermediates in lysozyme and α-lactalbumin. *Biochemistry* 24:874–81

36. One of the first reports to bring out the possibility of multiple pathways to the attention of experimental scientists.

54. Kuzmenika EV, Heyes CD, Nienhaus GU. 2005. Single molecule FRET study of denaturant induced unfolding of RNAse H. *J. Mol. Biol.* 357:313–24
55. Lakshmikanth GS, Sridevi K, Krishnamoorthy G, Udgaonkar JB. 2001. Structure is lost incrementally during the unfolding of barstar. *Nat. Struct. Biol.* 8:799–804
56. Leeson DT, Gai F, Rodriguez HM, Gregoret LM, Dyer RB. 2000. Protein folding and unfolding on a complex energy landscape. *Proc. Natl. Acad. Sci. USA* 97:2527–32
57. Li H, Frieden C. 2007. Comparison of C40/82A and P27A C40/82A barstar mutants using ^{19}F NMR. *Biochemistry* 46:4337–47
58. Li H, Frieden C. 2007. Observation of sequential steps in the folding of intestinal fatty acid binding protein using a slow folding mutant and ^{19}F NMR. *Proc. Natl. Acad. Sci. USA* 104:11993–98
59. Li L, Huang HH, Badilla CL, Fernandez JM. 2005. Mechanical unfolding intermediates observed by single-molecule force spectroscopy in a fibronectin type III module. *J. Mol. Biol.* 345:817–26
60. Lindberg MO, Tangrot J, Oliveberg M. 2002. Complete change of the protein folding transition state upon circular permutation. *Nat. Struct. Biol.* 9:818–22
61. Linse S, Linse B. 2007. Protein folding through kinetic discrimination. *J. Am. Chem. Soc.* 129:8481–86
62. Lowe AR, Itzhaki LS. 2007. Rational redesign of the folding pathway of a modular protein. *Proc. Natl. Acad. Sci. USA* 104:2679–84
63. Martinez JC, Serrano L. 1999. The folding transition state between SH3 domains is conformationally restricted and evolutionarily conserved. *Nat. Struct. Biol.* 6:1016–24
64. Matagne A, Chung EW, Ball LJ, Radford SE, Robinson CV, Dobson CM. 1998. The origin of the alpha-domain intermediate in the folding of hen lysozyme. *J. Mol. Biol.* 277:997–1005
65. Meekhof AE, Freund SMV. 1999. Probing residual structure and backbone dynamics on the milli- to picosecond time scale in a urea-denatured fibronectin type III domain. *J. Mol. Biol.* 286:579–92
66. Miranker A, Robinson CV, Radford SE, Aplin RT, Dobson CM. 1993. Detection of transient protein folding populations by mass spectrometry. *Science* 262:896–900
67. Mok KH, Kuhn LT, Goez M, Day IJ, Lin JC, et al. 2007. A preexisting hydrophobic collapse in the unfolded state of an ultrafast folding protein. *Nature* 447:106–9
68. Myers J, Oas TG. 2002. Mechanisms of fast protein folding. *Annu. Rev. Biochem.* 71:783–815
69. Nath U, Udgaonkar JB. 1995. Perturbation of a tertiary hydrogen bond in barstar by mutagenesis of the sole His residue to Gln leads to accumulation of at least one equilibrium folding intermediate. *Biochemistry* 34:1702–13
70. Nauli S, Kuhlman B, Baker D. 2001. Computer-based redesign of a protein folding pathway. *Nat. Struct. Biol.* 8:602–5
71. Nettels D, Gopich IV, Hoffmann A, Schuler B. 2007. Ultrafast dynamics of protein collapse from single-molecule photon statistics. *Proc. Natl. Acad. Sci. USA* 104:2655–60
72. Nishimura C, Dyson HJ, Wright PE. 2002. The apomyoglobin folding pathway revisited: structural heterogeneity in the kinetic burst phase intermediate. *J. Mol. Biol.* 322:483–89
73. Nishimura C, Dyson HJ, Wright PE. 2006. Identification of native and non-native structure in kinetic folding intermediates of apomyoglobin. *J. Mol. Biol.* 355:139–56
74. Nishimura C, Prytulla S, Dysin HJ, Wright PE. 2000. Conservation of folding pathways in evolutionary distant globin sequences. *Nat. Struct. Biol.* 7:679–86

70. A report on how stabilization of substructures can be used to switch a folding pathway.

75. Odefey C, Mayr L, Schmid FX. 1995. Non-prolyl cis-trans peptide bond isomerization as a rate-determining step in protein unfolding and refolding. *J. Mol. Biol.* 245:69–78
76. Olofsson M, Hansson S, Heideberg L, Logan DT, Oliveberg M. 2007. Formation of S6 structures with divergent amino acid composition and pathway flexibility within partly overlapping foldons. *J. Mol. Biol.* 365:237–48
77. Onuchic JN, Luthey-Schulten Z, Wolynes PG. 1997. Theory of protein folding: the energy landscape perspective. *Annu. Rev. Phys. Chem.* 48:545–600
78. Otzen DE, Oliveberg M. 1999. Salt-induced detour through compact regions of the protein folding landscape. *Proc. Natl. Acad. Sci. USA* 96:11746–51
79. Ozkan SB, Bahar I, Dill KA. 2001. Transition states and the meaning of N-values in protein folding kinetics. *Nat. Struct. Biol.* 8:765–69
80. Ozkan SB, Dill KA, Bahar I. 2002. Fast folding kinetics, hidden intermediates and the sequential stabilization model. *Protein Sci.* 11:1958–70
81. Pande VS, Grosberg AY, Tanaka T, Rokhsar DS. 1998. Pathways in protein folding: Is a new view needed? *Curr. Opin. Struct. Biol.* 8:68–79
82. Pappenberger G, Aygyn H, Engels JW, Reimer U, Fischer G, Kiefhaber T. 2001. Non-prolyl cis peptide bonds in unfolded proteins cause complex folding kinetics. *Nat. Struct. Biol.* 8:452–58
83. Patra AK, Udgaonkar JB. 2007. Characterization of the folding and unfolding of single-chain monellin: evidence for multiple intermediates and multiple pathways. *Biochemistry* 46:11727–43
84. Perez-Jimenez R, Garcia-Manyes S, Ainavarapu SR, Fernandez JM. 2006. Mechanical unfolding pathways of the enhanced yellow fluorescent protein revealed by single molecule force microscopy. *J. Biol. Chem.* 281:40010–14
85. Pletneva EV, Gray HB, Winkler JR. 2005. Snapshots of cytochrome *c* folding. *Proc. Natl. Acad. Sci. USA* 102:18397–402
86. Pradeep L, Udgaonkar JB. 2002. Differential salt-induced stabilization of structure in the initial folding intermediate ensemble of barstar. *J. Mol. Biol.* 324:331–47
87. Pradeep L, Udgaonkar JB. 2004. Osmolytes induce structure in an early intermediate on the folding pathway of barstar. *J. Biol. Chem.* 279:40303–13
88. Radford SE, Dobson CM, Evans PA. 1992. The folding of hen lysozyme involves partially structured intermediates and multiple pathways. *Nature* 358:302–7
89. Ramachandran S, Rami BR, Udgaonkar JB. 2000. Measurements of cysteine activity during protein unfolding suggest the presence of competing pathways. *J. Mol. Biol.* 297:733–45
90. Rami BR, Krishnamoorthy G, Udgaonkar JB. 2003. Dynamics of the core tryptophan during the formation of a productive molten globule intermediate of barstar. *Biochemistry* 42:7986–8000
91. Rami BR, Udgaonkar JB. 2001. pH-jump induced folding and unfolding studies of barstar: evidence for multiple folding and unfolding pathways. *Biochemistry* 40:15267–79
92. Rami BR, Udgaonkar JB. 2002. Mechanism of formation of a productive molten globule form of barstar. *Biochemistry* 41:1710–16
93. Sabelko J, Ervin J, Gruebele M. 1999. Observation of strange kinetics in protein folding. *Proc. Natl. Acad. Sci. USA* 96:6031–36
94. Sadqi M, Fushman D, Munoz V. 2006. Atom-by-atom analysis of global downhill protein folding. *Nature* 442:317–21
95. Saigo S, Shibayama N. 2003. Highly nonexponential kinetics in the early phase refolding of proteins at low temperatures. *Biochemistry* 42:9669–76
96. Sanchez IE, Kiefhaber T. 2003. Evidence for sequential barriers and obligatory intermediates in apparent two-state folding. *J. Mol. Biol.* 325:367–76

88. One of the first experimental reports on two parallel pathways for folding.

97. Saxena AK, Udgaonkar JB, Krishnamoorthy G. 2006. Characterization of intramolecular distances and site-specific dynamics in chemically unfolded barstar: evidence for denaturant-dependent nonrandom structure. *J. Mol. Biol.* 359:174–89
98. Schwarzinger S, Wright PE, Dyson HJ. 2002. Molecular hinges in protein folding: the urea denatured state of apomyoglobin. *Biochemistry* 41:12681–86
99. Shastry MC, Roder H. 1998. Evidence for barrier-limited protein folding kinetics on the microsecond time scale. *Nat. Struct. Biol.* 5:385–92
100. Shastry MC, Udgaonkar JB. 1995. The folding mechanism of barstar: evidence for multiple pathways and multiple intermediates. *J. Mol. Biol.* 247:1013–27
101. Shimada J, Shakhnovich EI. 2002. The ensemble folding kinetics of protein G from an all-atom Monte Carlo simulation. *Proc. Natl. Acad. Sci. USA* 99:11175–80
102. Shortle D. 2002. The expanded denatured state: an ensemble of conformations trapped in a locally encoded topological space. *Adv. Protein Chem.* 62:1–23
103. Sinha KK, Udgaonkar JB. 2005. 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.* 353:704–18
104. Sinha KK, Udgaonkar JB. 2007. Dissecting the nonspecific and specific components of the initial folding reaction of barstar by multi-site FRET experiments. *J. Mol. Biol.* 370:385–405
105. Snow CD, Sorin EJ, Rhee YM, Pande VS. 2005. How well can simulation predict protein folding kinetics and thermodynamics? *Annu. Rev. Biophys. Biomol. Struct.* 34:43–69
106. Song J, Jamin N, Gilquin B, Vita C, Menez A. 1999. A gradual disruption of tight side-chain packing: 2D ^{1}H-NMR characterization of acid-induced unfolding of CHABII. *Nat. Struct. Biol.* 6:129–34
107. Sosnick TR, Mayne L, Hiller R, Englander SW. 1994. The barriers in protein folding. *Nat. Struct Biol.* 1:149–57
108. Spudich GM, Miller EJ, Marqusee S. 2004. Destabilization of the *Escherichia coli* RNase H kinetic intermediate: switching between a two-state and three-state mechanism. *J. Mol. Biol.* 335:609–18
109. Sridevi K, Juneja J, Bhuyan AK, Udgaonkar JB. 2000. 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.* 302:479–95
110. Sridevi K, Lakshmikanth GS, Krishnamoorthy G, Udgaonkar JB. 2004. Increasing stability reduces conformational heterogeneity in a protein folding intermediate ensemble. *J. Mol. Biol.* 337:699–711
111. Sridevi K, Udgaonkar JB. 2002. Unfolding rates of barstar determined in native and low denaturant conditions indicate the presence of intermediates. *Biochemistry* 41:1568–78
112. Sridevi K, Udgaonkar JB. 2003. Surface expansion is independent of and occurs faster than core salvation during the unfolding of barstar. *Biochemistry* 42:1551–63
113. Udgaonkar JB, Baldwin RL. 1990. Early folding intermediate of ribonuclease A. *Proc. Natl. Acad. Sci. USA* 87:8197–201
114. Viguera AR, Serrano L, Wilmanns M. 1996. Different folding transition states may result in the same native structure. *Nat. Struct. Biol.* 3:874–80
115. Wallace LA, Matthews CR. 2002. Sequential vs parallel protein-folding mechanisms: experimental tests for complex folding reactions. *Biophys. Chem.* 101:113–31
116. Wani AH, Udgaonkar JB. 2006. HX-ESI-MS and optical studies of the unfolding of thioredoxin indicate stabilization of a partially unfolded, aggregation-competent intermediate at low pH. *Biochemistry* 45:11226–38

110. A multisite, TRFRET study of heterogeneity in a folding intermediate, which showed that heterogeneity is less prevalent under conditions that confer more stability.

117. Wedemeyer WJ, Welker E, Scheraga HA. 2002. Proline cis-trans isomerization and protein folding. *Biochemistry* 41:14637–44
118. Weissman JS. 1995. All roads lead to Rome? The multiple pathways of protein folding. *Chem. Biol.* 2:255–60
119. Wildegger G, Kiefhaber T. 1997. Three-state model for lysozyme folding: triangular folding mechanism with an energetically trapped intermediate. *J. Mol. Biol.* 270:294–304
120. Wintrode P, Rojsajjakul T, Vadrevu R, Matthews CR, Smith DL. 2005. An obligatory intermediate controls the folding of the α-subunit of tryptophan synthase, a TIM barrel protein. *J. Mol. Biol.* 347:911–19
121. Wright CF, Lindorff-Lagen K, Randles LG, Clarke J. 2003. Parallel protein-unfolding pathways revealed and mapped. *Nat. Struct. Biol.* 10:658–62
122. Yang H, Smith DL. 1997. Kinetics of cytochrome *c* folding examined by hydrogen exchange and mass spectrometry. *Biochemistry* 36:14992–99
123. Zaidi FN, Nath U, Udgaonkar JB. 1997. Multiple intermediates and transition states during protein unfolding. *Nat. Struct. Biol.* 4:1016–24
124. Zarrine-Afsar A, Larson SM, Davidson AR. 2005. The family feud: Do proteins with similar structures fold via the same pathway? *Curr. Opin. Struct. Biol.* 15:42–49
125. Zwanzig R. 1997. Two-state models of protein folding kinetics. *Proc. Natl. Acad. Sci. USA* 94:148–50
126. Zwanzig R, Szabo A, Bagchi B. 1992. Levinthal's paradox. *Proc. Natl. Acad. Sci. USA* 89:20–22

119. An educative account of how kinetic methods can demonstrate the operation of two competitive pathways of folding.

121. Some of the best evidence for multiple pathways being available for unfolding.

126. Explains how cumulative selection of native substructures during folding provides a solution to the Levinthal Paradox.