Annu. Rev. Biophys. Biomol. Struct. 2000. 29:213–38
Copyright © 2000 by Annual Reviews. All rights reserved

PROTEIN FOLDING INTERMEDIATES AND PATHWAYS STUDIED BY HYDROGEN EXCHANGE

S. Walter Englander
Johnson Research Foundation; Department of Biochemistry and Biophysics, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania 19104; e-mail: walter@HX2.Med.upenn.Edu

Key Words protein folding, hydrogen exchange, intermediates, pathways, kinetic barriers

■ **Abstract** In order to solve the immensely difficult protein-folding problem, it will be necessary to characterize the barriers that slow folding and the intermediate structures that promote it. Although protein-folding intermediates are not accessible to the usual structural studies, hydrogen exchange (HX) methods have been able to detect and characterize intermediates in both kinetic and equilibrium modes—as transient kinetic folding intermediates on a subsecond time scale, as labile equilibrium molten globule intermediates under destabilizing conditions, and as infinitesimally populated intermediates in the high free-energy folding landscape under native conditions. Available results consistently indicate that protein-folding landscapes are dominated by a small number of discrete, metastable, native-like partially unfolded forms (PUFs). The PUFs appear to be produced, one from another, by the unfolding and refolding of the protein's intrinsically cooperative secondary structural elements, which can spontaneously create stepwise unfolding and refolding pathways. Kinetic experiments identify three kinds of barrier processes: (*a*) an initial intrinsic search-nucleation-collapse process that prepares the chain for intermediate formation by pinning it into a condensed coarsely native-like topology; (*b*) smaller search-dependent barriers that put the secondary structural units into place; and (*c*) optional error-dependent misfold-reorganization barriers that can cause slow folding, intermediate accumulation, and folding heterogeneity. These conclusions provide a coherent explanation for the grossly disparate folding behavior of different globular proteins in terms of distinct folding pathways.

CONTENTS

PERSPECTIVES AND OVERVIEW 214
HYDROGEN EXCHANGE 216
Some Principles 216
HX Labeling 217
INTERMEDIATES BY KINETIC HX LABELING 217
pH Competition 217

Pulse Labeling 219
Summary Results 221
MOLTEN GLOBULES: Intermediates by Equilibrium HX Labeling 221
INTERMEDIATES AT EQUILIBRIUM IN THE FOLDING LANDSCAPE 223
The Staircase Hypothesis 223
Native State HX 223
SUMMARY: Discrete Intermediates 225
ON OR OFF PATHWAY 225
THE KINETIC DETERMINANTS: Nucleation and Misfolding Barriers 227
Three-State Folding and Heterogeneity 227
Two-State Folding 229
CONCLUSIONS 230

PERSPECTIVES AND OVERVIEW

The fundamental questions in protein folding concern the structure of folding intermediates, the nature of the barrier processes that accompany intermediate formation, and the implications for folding pathways. Convincing answers have not emerged. Part of the problem is that different proteins exhibit wildly disparate folding behavior. This has led to sharply opposed hypotheses, often based on fundamentally different conceptions of the protein-folding landscape.

Experimental data are often interpreted in terms of the classical pathway paradigm which holds that unfolded polypeptides move toward their native structures through one or more distinct pathways defined by discrete intermediate forms (36, 42, 80, 81, 92). Conventional free-energy diagrams, as in Figure 1*A*, illustrate the system's important equilibrium and kinetic relationships. The wells portray the sequence and stability of intermediates and endstates, the barriers indicate their interconversion rates, and the reaction coordinate follows the important pathway events. One can choose to draw these diagrams in a multidimensional sense to indicate that the intermediates and the transition states retain considerable freedom, and one can include alternative folding routes, but that does not change the pathway paradigm.

A different paradigm is most often pictured in terms of a funnel-shaped reaction landscape (Figure 1*B*) to emphasize the multidimensionality of the folding reaction space of polypeptides (23, 40, 76, 97, 102, 135, 139). The vertical reaction-coordinate of the funnel is an increasingly negative scale of internal energy (enthalpy), often set proportional to the fraction of native contacts formed (Q). The XY dimension displays the conformational entropy of the polypeptide and its continuous decrease as folding proceeds. Since the determining free-energy parameter does not explicitly appear, the funnel representation itself is noncommittal with respect to intermediates and barriers. Detailed features have been suggested by experimental results with real proteins but especially by computer simulations of the folding of nonprotein models (22, 103, 126, 132) such as the lattice model shown. The central concept is that unfolded polymers diffuse energetically

downhill toward their native state through an unlimited number, essentially a continuum, of intermediates and paths. Rates are determined by the slope of the landscape and its roughness, essentially a continuum of small barriers. The multiple nonnative wells within and at the bottom of the funnel acknowledge the fact of specific intermediate accumulation in many proteins, but these are viewed as misfolded off-pathway species formed by the accidental occupation of nonobligatory states that act to slow rather than promote folding.

These fundamentally different viewpoints lead to very different interpretations of any given observation. For example, some proteins fold in a fast (msecs) two-state manner with no apparent intermediates. This might mean that distinct intermediates do not exist (Figure 1B) or that the initial kinetic barrier in a discrete folding pathway is rate-limiting so that the pathway intermediates are not seen (Figure 1A, top). Other proteins fold more slowly (secs), in a three-state manner and do accumulate intermediates. This might mean that an optional off-pathway intermediate, accidentally trapped in a deep off-pathway well, causes the slow folding (Figure 1B). Alternatively, both slow folding and the accumulation of one of the obligatory pathway intermediates might be caused by an optionally encountered error-dependent barrier (Figure 1A, bottom). Many proteins fold heterogeneously, with different molecules exhibiting different behavior. This might reflect multiple parallel pathways as in Figure 1B or a single pathway in which different molecules happen to encounter different error-dependent barriers and thus populate different intermediates and fold at different rates (Figure 1A, top plus bottom).

Are intermediates and pathways continuous or discrete? Clearly, insightful structural studies of the intermediates are required. This is difficult. Kinetic intermediates are very short lived; most proteins that have been studied fold to their native state in less than one second. Possible intermediates are invisible when folding is two-state, and they cannot be isolated for structural studies even when they do accumulate in three-state folding. Most experimental methods used to observe such fast processes are able to provide only nonspecific information about unresolved aromatic side chains (fluorescence, absorbance) or some averaged property of the polypeptide main chain (circular dichroism, scattering, infrared). Such signals provide rate information and can imply the presence of folding intermediates, but they do not define structures and cannot distinguish the fundamentally different viewpoints just described.

In principle these limitations do not apply to the ingenious theoretical approaches that have been constructed to simulate the entire kinetic folding history of single molecules. However, computational limitations have so far restricted these studies mainly to simplified models (e.g. Figure 1B) that lack properties such as main chain and side chain stereochemistry, secondary structure and hydrogen bonding, tertiary structural topology, solvent interaction, and realistic potential functions. One does not know a priori whether any or all of these factors are crucial or irrelevant for modeling protein-folding behavior.

To resolve these primary issues and then to delve more deeply into folding processes, it is necessary to obtain clear information on the structure, stability, and

interrelationships of the folding intermediates, whether continuous or discrete, that in fact carry real protein molecules from their unfolded state (U) to their native state (N). In this effort, methods based on hydrogen exchange (HX) have proven most useful. This article summarizes the pertinent HX developments, the results obtained, and their implications for the intermediates and the barriers in protein-folding pathways.

HYDROGEN EXCHANGE

How can hydrogen exchange data define the structure of nonnative molecules? HX results typically define the presence or absence of hydrogen bonding at a large number of identifiable residue amides in a protein. The pattern of H-bonding identifies secondary structure. One can then distinguish similarities to and differences from the known native structure. In addition, HX data can evaluate structural stability and flexibility at an amino acid–resolved level.

Some Principles

The basic facts of protein HX processes have been multiply recounted (30, 50, 55, 63, 143). Protons on polar side chains and on the main chain amides, distributed uniformly throughout all protein molecules, exchange naturally with solvent protons in a nonperturbing manner. Main chain NH protons are most informative. Their exchange rates depend on the protecting H-bonds that mark the regular elements of secondary structure and can evaluate protein structure, stability, and dynamics at very many probe points and at amino acid resolution.

Amide protons exchange with solvent by way of proton transfer reactions (45) that depend on a number of factors—pH, temperature, neighboring side chains, and the isotopes involved—all of which have been thoroughly calibrated (8, 33, 98). The exchange reaction requires direct contact between the catalyst ion and the amide group and so cannot occur when the group in question is involved in a protecting H-bond. For example, H-bonded groups in small molecules and at the aqueous surface of proteins exhibit slow exchange even though they are exposed to solvent. To permit exchange, the H-bond donor-acceptor pair must be separated, apparently by 5 Å or more (94), so that the proton can H-bond to and be carried away by solvent catalyst. Any given H-bond can be broken by the transient global unfolding reaction (stability), which severs all H-bonds (6, 9, 69, 77, 89, 121, 142), or by a local fluctuation that separates only the individual bond (flexibility) (94), or by subglobal unfolding reactions that break sets of neighboring H-bonds (7, 10, 53). We are especially interested in the latter because they must include the intermediates that come into play in protein-folding reactions.

In the usual kinetic limit, known as the EX2 condition (70), where structural reclosing is faster than the intrinsic unprotected exchange rate (k_{unp}), the HX rate

measured for any given hydrogen (k_{ex}) provides the transient opening equilibrium constant ($K_{op} = k_{ex}/k_{unp}$). From this the free energy of the determining opening reaction can be obtained (70, 87) ($\Delta G_{op} = -RT \ln K_{op} = -RT \ln k_{ex}/k_{unp}$). Therefore the measurement of HX rates and their sensitivity to ambient parameters can provide site-resolved information on the presence or absence of H-bonded structure, its stability, dynamics, and other properties.

But how does one make the measurements necessary to study ephemeral folding intermediates?

HX Labeling

HX is a chemical rather than a spectroscopic method and therefore can be used in a labeling mode. One can perform experiments that selectively emplace hydrogen isotope label at positions that are functionally interesting, under conditions most suitable for the experimental purpose. The hydrogen isotope profile can then be held essentially frozen in place by the native protein structure and by ambient conditions that ensure slow exchange while the labeled protein is placed into conditions suitable for analysis of the H-label by NMR or mass spectrometry (52). The results provide information about the presence and stability of H-bonded structure at many identifiable points through the protein.

HX labeling methods were first applied to the folding problem in the laboratory of RL Baldwin in experiments that used tritium labeling and liquid scintillation counting analysis (20, 21, 79, 123). The advent of two-dimensional NMR provided access to the inherent amino acid resolution of HX measurements (136). Recent progress in mass spectrometry extends these capabilities and can make them applicable to larger proteins (96, 146), especially when the analytical resolution is enhanced by fragment separation methods (47, 49, 120).

The combination of HX labeling with NMR analysis has made it possible to measure the structure and stability of populated intermediates that exist for less than 1 sec in kinetic folding (reviewed in 12, 51, 81, 141). The same methods have been used to define the H-bonded secondary structural elements that are present in molten globule forms, often thought to represent equilibrium analogs of kinetic intermediates (109). Most recently a method known as native state HX has made it possible to study partially unfolded intermediates that exist at equilibrium as infinitesimally populated, conformationally excited forms in the high free-energy conformational space of proteins under native conditions (7, 10).

INTERMEDIATES BY KINETIC HX LABELING

pH Competition

The pH competition method was introduced by Schmid & Baldwin using radioactive tritium labeling in studies of ribonuclease A (RNase A) (123). The method measures the rate of folding-dependent H-bond formation by competing the known

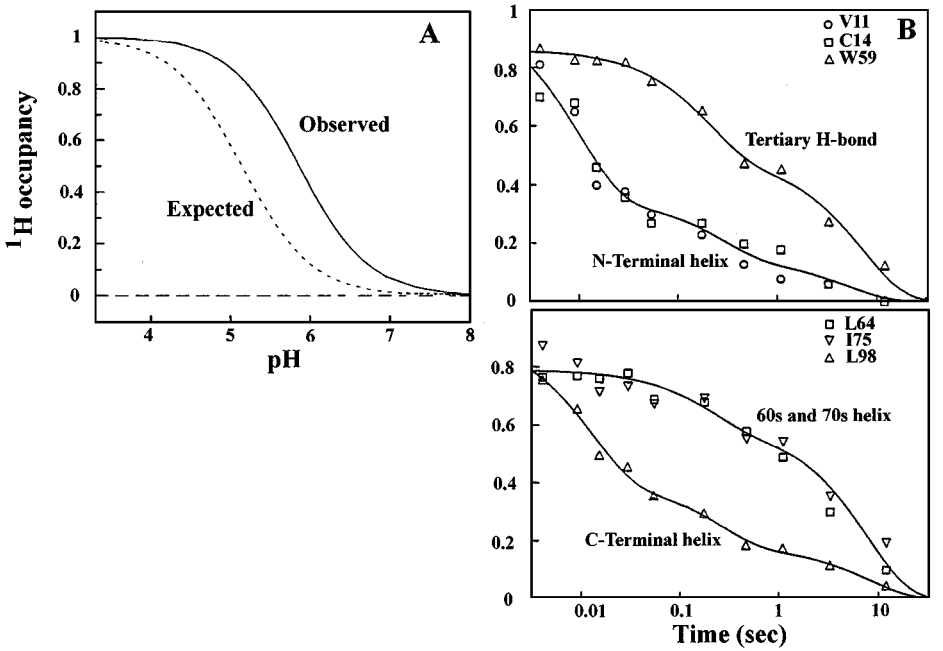

Figure 2 Modes of kinetic HX labeling. *A.* HX—pH competition, illustrating the expected HX labeling of a given amide NH site when no protection is present (*dashed line*) before native state formation and for a protection factor of 5 (*solid line*). *B.* HX pulse labeling data for cyt c (118) illustrating the selective early protection (H-bonding) of residues in the N- and C-terminal helices and folding heterogeneity.

rate of HX labeling against the to-be-determined rate of HX protection. Reviews are (78, 116).

In a typical pH competition experiment (Figure 2*A*), a denaturant-unfolded protein in H_2O is diluted into D_2O to initiate refolding in a stopped-flow apparatus that can mix solutions in ~1 msec. The pD of the diluent is varied. At low pD, HX is slow compared to folding and the protein folds before any of its amide NHs can exchange to ND. When the pD is increased, exchange is accelerated (catalyzed by OD^-) and competes more successfully with the folding rate. Amides then lose more of their H-label before the exchange period is terminated by hydrogen bond formation. The degree of H-label remaining at each amide is measured in the refolded native protein by NMR.

If the protein folds directly to the strongly protected native state in a two-state manner with folding rate constant k_f, each structurally protected amide will remain half labeled (H) when its intrinsic unprotected exchange rate (k_{unp}) equals the folding rate ($k_{ex} = k_{unp} = k_{ch}\,[OD^-] = k_f$). Here k_{ch} is the second-order exchange rate constant characteristic for each unprotected amide, and exchange is catalyzed

by OD^-. The expected curve of label remaining for each amide as a function of the labeling pD (Figure 2*A*) can be drawn from its known unprotected HX rate and the known folding rate, obtained from other signals. If some structure that fully protects an amide proton forms faster, then the labeling curve for that residue will be right-shifted to higher pD, from which the rate for H-bond formation can be obtained. If a partially structured intermediate is formed, the amides that are protected and unprotected in the intermediate will exhibit different behavior, from which the identity of the intermediate might be inferred. If some degree of HX protection (P) exists during the labeling period, then $k_{ex} = k_{unp}/P$. This too will right-shift the labeling curve to higher pD, from which P can be calculated.

The pH competition analysis is limited to proteins with folding rate essentially unaffected by pH over the pH range studied. Also, it is not easy to disentangle the complex folding behaviors that often occur, such as partial protection and intermediate formation, or to discern heterogeneous folding in which different fractions of the population fold at different rates.

Gladwin & Evans (62) modified the competition method to focus selectively on structure that forms rapidly, within the dead time of stopped-flow mixing experiments (<1 msec) (62). The unfolded protein is diluted into folding conditions in D_2O at various pD as before, but labeling is terminated after a few msecs by mixing into lower pD rather than by the folding event itself. As before, protection that already exists during the labeling period is seen as a rightward shift of the labeling curve to higher pD (Figure 2*A*). Gladwin & Evans stress the need to carefully account for the HX protection that is due to the early stages of slower folding events, occurring after rather than within or before the labeling period. Also HX rates are impressively sensitive to steric blocking (8). These factors may spuriously mimic significant generalized protection.

Pulse Labeling

The Baldwin laboratory also pioneered a related experiment known as HX pulse labeling, which they used together with tritium labeling to study a late intermediate caused to accumulate in the folding of RNase A by a nonnative proline isomer (79). The experiment uses a brief pulse of HX labeling to test for the presence of H-bonding (HX protection) at particular times during the folding process. When used with H-D exchange, stopped-flow, and NMR analysis, the method can track H-bond formation at many amides throughout the protein and thus determine the structure of folding intermediates that form, accumulate, and disappear on a multimillisecond time scale. This capability, initially demonstrated in RNase A (133) and cytochrome c (cyt c) (118), placed the study of folding intermediates on a firm structural basis and helped to transform the folding field into the vigorous effort that it has become. For reviews, see (12, 51, 141).

In the usual experimental design, a protein is initially unfolded at high denaturant in D_2O so that all of its amides are D-labeled. To start refolding, the solution is diluted into H_2O buffered at moderately low pH, e.g. pH 6 at 10°C, where the

exchange of freely exposed amides is slow ($k_{unp} \sim$ secs). After various refolding times, a second mix sharply raises the pH (pH 8.5 to 10) to initiate H-labeling ($k_{unp} \sim$ msecs), for up to 50 msecs, and then the pulse is terminated by a third mix that drops the pH back to low values. During the labeling pulse, amides not yet protected by hydrogen bonding are quickly labeled, whereas strongly protected amides are not. Refolding proceeds to completion. The H-D profile imposed during the labeling pulse is held by the native structure and ambient slow exchange conditions while the protein is analyzed, either by NMR to obtain site resolution or by mass spectrometry which can determine whether the protection of different elements of structure occurs within the same molecule (96). The results provide a snapshot of the H-bonded structure that exists within the brief period of the labeling pulse.

An example using NMR analysis is shown in Figure 2*B* for cyt c folding (118). The data illustrate some amino acids that participate in an intermediate with only the N- and C-terminal helices folded. About half of the molecules form the N/C intermediate in ~20 msec. Other fractions of the population follow a different time course, reaching N either more rapidly (~20%), without accumulating the N/C intermediate, or more slowly. More complete data like this can determine the identity of intermediates and their rates of production. Also the HX rates of the protected amides (stability; flexibility) can be measured. An unforeseen result is the observation of folding heterogeneity. Chemically identical molecules under identical conditions fold at different rates and populate different intermediates, or none at all.

Ideally the pulse method labels or fails to label amides in a yes-no fashion, depending on whether they are protected at the time of the pulse or not. In reality, different unprotected amides have very different intrinsic HX rates (8). Thus at any given pulse strength, some amides with slow intrinsic rates may escape full labeling even though they are unprotected; others with fast intrinsic rates may become partly labeled even though they are already weakly protected in reforming structure. The protection factors of the different amides within an intermediate can be probed by using a series of pH values in the labeling pulse (46, 134), as in the pH competition experiment. A thorough analysis would involve a pH series at each of a series of folding times, but this has not been done.

Some questions arise. Does the measured protection reflect the reversible unfolding of the entire intermediate, or parts thereof (stability), or much more local fluctuations (flexibility)? Comparison of protection factors for neighboring amides may distinguish these options (94). Is exchange in the EX2 limit? The usual tests (pH dependence; comparison of measured rates with intrinsic chemical rates) (9) may be able to tell. Is the pulse time sufficiently long to allow H-bonds to open even once? The length of the pH pulse can serve as a useful variable (134). More generally, the problems that are intrinsic to kinetic experimentation remain. Only populated intermediates can be observed. Are observed intermediates merely a small fraction of a vast repertoire of possible folding intermediates? Are they obligatory or optional? On-pathway or off?

Summary Results

Kinetic HX labeling has been applied to a number of proteins with similar and noncontroversial results (25, 72, 75, 90, 91, 100, 113, 119). In the case of two-state folding, one sees all of the amides gain protection at the folding rate. In three-state folding, one generally sees behavior like that in Figure 2, showing the population of a discrete native-like intermediate. Some intermediates are clearly early with only a small fraction of the protein formed, as in Figure 2*B*. Others are closer to the entire native structure.

Significant folding heterogeneity is often observed. Folding heterogeneity has been interpreted variably in terms of multiple parallel pathways with different intermediates (113), or as a single folding pathway interrupted by alternative error-dependent barriers (Figure 1*A*, top plus bottom) (128). A decision in favor of two parallel pathways has often been further interpreted as support for the funnel model (Figure 1*B*), erroneously so since that model pictures an unlimited continuum of paths and intermediates and no distinct on-pathway forms.

Of major interest is the general observation from much HX labeling work that intermediates seen for various proteins always represent pieces of the native protein. An early concern was that these experiments would overinterpret apparently native-like structure because only amides that are protected in the native protein can be measured (34). The many native-like *patterns* of H-bonding that have now been seen for intermediates with various conformations and degrees of advancement lessen this concern although it is true that detailed nonnative interactions are unlikely to be detected.

MOLTEN GLOBULES: Intermediates by Equilibrium HX Labeling

A number of proteins when placed in mildly destabilizing conditions, especially low pH, assume an expanded but still somewhat structured form called a (or *the*) molten globule (59, 83, 108–110). Ptitsyn and his coworkers proposed that these forms represent a new thermodynamic state, structurally compact (globular), with some secondary structural content but lacking rigid tertiary packing (molten) (43), and that they represent equilibrium analogs of kinetic folding intermediates (93, 110) (see especially the work of Kuwajima and coworkers 82–84). This view appeared to be inconsistent with the conventional concept of proteins as monolithically cooperative two-state entities (although see Figure 4) and was poorly received by protein chemists for many years. The thermodynamic status of molten globules hinges on whether they are connected to neighboring forms by a second-order continuum or separated by a first-order cooperative transition (107), analogous to the difference between continuous and distinct intermediates indicated in Figure 1.

We are more concerned here with the structural status of molten globules which, due to their dynamically disordered tertiary structure, cannot be specified by the usual X-ray and NMR methods except in the most well ordered examples (57, 115). Fortunately, the secondary structural elements that are present can be specified by HX labeling. Typically, the molten globule is placed into D_2O and allowed to exchange for increasing times. Amides in unstructured regions become rapidly deuterated, perhaps at the unprotected amide rate, while H-bonded amides exchange their proton label more slowly. Timed samples are returned to native conditions where the imposed H-D exchange profile is locked in and can be analyzed by NMR or mass spectrometry.

A turning point in molten globule studies came with the demonstration by HX labeling of native-like elements of secondary structure in the classical molten globule of α-lactalbumin (18), the namesake molten globule of cyt c (74), and the apomyoglobin pH 4 molten globule (68). The demonstration that molten globular structure lies between U and N modifies the long-entrenched concept that proteins are monolithically cooperative two-state entitites (see Figure 4). Impressively, the very same partial structure found in the apomyoglobin molten globule was demonstrated in a kinetic folding intermediate by HX pulse labeling (75). A similar correspondence has been found for RNase H (38, 114). These results support the view that molten globules prepared at equilibrium represent analogs of kinetic folding intermediates (109), at least in some cases.

The powerful implications of the molten globule hypothesis led to an outpouring of new studies describing many partially folded proteins at equilibrium and also in kinetic experiments (3, 17, 28, 29, 38, 88, 96, 99, 112, 124, 125). These intermediates were invariably asserted to represent molten globules. Accordingly the concept of molten globule structure and thermodynamic character (65, 107) has evolved and become more synonymous with intermediates in general (37, 41, 109).

HX protection factors in molten globules are typically low, well under 1000. This presumably reflects their low structural rigidity which can be expected to facilitate local H-bond breaking fluctuations (94). Accordingly a pattern of protection factors may be seen that is different from the more rigid native protein (74, 125). Nevertheless, the easily discerned H-bonding pattern specifies the secondary structure that is present and implies supporting tertiary interactions. In all of these studies, one finds native-like secondary structural elements.

Whether all of these forms represent *true* molten globules, or whether *the* molten globule is something other than any generic partially folded intermediate is not at issue here. The large molten globule literature documents the same important conclusion that one can draw from the kinetic HX labeling literature discussed above. Distinct native-like structures are selected, form abundantly, and can be stably maintained. Further, in some cases these equilibrium intermediates have been shown to match the structure of kinetic intermediates observed independently by HX pulse labeling (75, 114).

INTERMEDIATES AT EQUILIBRIUM IN THE FOLDING LANDSCAPE

In principle, protein molecules must continually cycle through all of the high free-energy states implied in Figures 1*A* and 1*B* and must populate each one as dictated by the Boltzmann distribution. Thus proteins unfold and refold even under fully native conditions. In favorable cases a native state HX method can characterize some of the intermediate forms.

The Staircase Hypothesis

The high free-energy states that populate the folding landscape include all of the partially folded intermediates utilized by protein molecules in folding from U to N and all of the partially unfolded intermediates, reaching from N to U, that determine HX reactions. In light of this commonality, Englander & Kallenbach suggested that progressively slower HX rates measured at equilibrium under native conditions might reflect an energetically uphill staircase of increasingly unfolded forms down which proteins might step in their folding sequence (50, 54). The staircase hypothesis was later discussed by Woodward and her coworkers (77, 140), especially the possibility that the slowest protons to exchange might show the first part of a protein to refold, based on newly available HX labeling results for pancreatic trypsin inhibitor (BPTI) (119), cyt c (118), and lysozyme (113).

In fact the results available for these and other proteins provide rather ambiguous support for these hypotheses (31, 32, 48). In cyt c, it is true that the slowest exchanging protons reflect the final global unfolding (9) and participate in the initial N/C helix intermediate found by HX pulse labeling (118). However, most of the residues in the C-terminal helix and all in the N-terminal helix exchange much faster. Similar ambiguities are seen in lysozyme (112, 113). In BPTI insufficient refolding data is available to judge (119).

Thus straightforward HX data although suggestive do not strongly support the staircase model of Englander & Kallenbach or the more limited last-out first-in model of Woodward and coworkers. The general problem is that HX in many proteins is dominated by local fluctuations that break only one H-bond (94). These forms cannot show us important folding intermediates, and they obscure the measurement of larger unfoldings that might.

Native State HX

A method known as native state HX (7, 10, 53) can ferret out the large unfoldings that might reveal the structure of folding intermediates. In this approach, one selectively promotes the large unfoldings so that they come to dominate the exchange of the many hydrogens that they expose. HX results can then identify the otherwise invisible partially unfolded forms (PUFs) and measure their stability.

The number and distribution of these forms can help to map the high free-energy reaction landscape between N and U. Do the intermediates describe an undifferentiated continuum, as in Figure 1*B*, or do a small number of native-like metastable intermediates occupy distinct free-energy wells that might shape a stepwise folding pathway, as in Figure 1*A*? A successful native state HX analysis should be able to document the true situation in either case.

An example in Figure 3 shows HX data for all of the amides in the 60s helix of cyt c and some in the right loop. HX was measured in increasing concentrations of denaturant to selectively promote the larger unfolding reactions. Data for the various amide hydrogens are plotted in terms of the free energy for the responsible opening reactions. These hydrogens exchange initially by way of different local fluctuations, shown by their insensitivity to denaturant and the very different ΔG_{op} values for neighboring NHs. The one exception is the Leu68 amide proton. It can exchange only by way of a large, denaturant sensitive, high free-energy unfolding. When this large unfolding is sufficiently promoted by increasing denaturant, all of the amide protons in the 60s helix merge into the same HX isotherm, i.e. they all become dominated by the same large unfolding. The measurable amides in the right loop do the same. This result reveals a state with the entire 60s helix and perhaps the green loop cooperatively unfolded but with other structure still intact (identified by the still protected residues, namely the N- and C-helices, shown as the blue unit in Figure 3*B*). Similar behavior is seen for the other secondary structural units in cyt c (10, 95), which merge into HX isotherms (indicated by the dashed lines in Figure 3*A*, color coded to match the protein segments that they represent in Figure 3*B*). These results show that the cyt c protein is made up of four cooperative units. In lattice model terminology, cyt c might be considered to correspond to four connected beads rather than 124 independently folding amino acids.

The native state HX experiment has been applied with analogous results to RNAse H (26) and apocyt b562 (60, 61), and in a more limited way to barstar (19) and a hyperthermophilic rubredoxin (66). One does not see intermediates distributed in a continuous manner on a downsloping (free) energy scale as in funnel models. Protein intermediates appear to be formed by the unfolding and refolding of entire secondary structural elements or sets thereof. The cooperative tertiary packing of secondary structural units greatly increases their stability but does not destroy their separate intrinsically cooperative nature. Approaches are being developed that may explain the determinants of the protein unfolding-refolding reactions (4, 67).

Clarke et al note the failure of the native state HX experiment to find intermediates—either continuous or discrete—in barnase and CI2 (31). The native state HX method does set certain requirements for any given protein in order to serve as a good model for these studies, as has been described (7, 48, 95). For example barnase may fail because it diverges from the EX2 limit at low denaturant (39). Clarke et al have also noted the inability of an equilibrium result all by itself to specify a kinetic sequence. Englander and coworkers list kinetic and equilibrium results

that support the intermediates and pathways implied by the native state HX results (48, 56). Some are noted below.

SUMMARY: Discrete Intermediates

The information just summarized, obtained especially from HX experiments, in both equilibrium and kinetic modes, on many proteins, shows unquestionably that selected partially folded intermediates exist. In all cases the intermediates turn out to represent partial replicas of the native protein. Further, the intermediate forms are robust. The same intermediates are maintained over a broad range of destabilizing conditions including denaturants, temperature, and pressure. The intermediates utilize the built-in intrinsically cooperative secondary structural units of the protein. It appears that proteins naturally stabilize not only their unique native state but in addition a small number of discrete partially folded forms.

Why then do proteins most often melt in a highly cooperative two-state manner? Secondary structural elements packed in a globular protein cooperatively stabilize each other. When global stability is sufficiently reduced in melting studies, the loss of any unit causes the entire protein to unfold in what appears to be a highly cooperative two-state process. However, when stability is high, the individual elements can express their individually cooperative nature and unfold separately without destabilizing the entire molecule. The free-energy relationships that determine measurable unfolding patterns in conditions of low and high stability are illustrated in Figure 4.

At the first level, the various results just described help to map the equilibrium contours of the high free-energy reaction surface. In any given intermediate, distinct secondary structures are formed and others are unfolded. That is, the intermediates occupy discrete wells, albeit with some width, in the folding landscape. The immense collection of all other possible forms must also exist, as implied in Figure 1*B*, but apparently only at higher free-energy levels than the few selected metastable intermediates. In an equilibrium sense, these results are as expected from the classical view of the high free-energy folding landscape as suggested by the free-energy wells in Figure 1*A*, but not from the multiform continuum view in Figure 1*B*.

ON OR OFF PATHWAY

What is the role of the discrete intermediates in kinetic folding? Do these intermediates represent productive, even obligatory, on-pathway forms or are they off-pathway forms in deep local minima that inhibit rather than promote folding? Considerable evidence points to a productive pathway role.

In cyt c, the same N/C helix intermediate seen in the native state HX experiment (10) is found also as a kinetic folding intermediate by HX pulse labeling

(118). These separate measurements even find the same stability, ~3 kcal/mol (5). The N/C helix intermediate accumulates when the green segment (Figure 3*B*) is trapped, which is the next unit in line in the suggested folding sequence. Other results (144) indicate that the several PUFs are just the forms needed to construct a sequential unfolding pathway. These are, in order of increasing free energy, the entire red loop unfolded, the red plus yellow loops unfolded, those two plus the 60s helix and green loop unfolded, and finally all of these together with the blue helices to produce the globally unfolded state. If so, then the same sequence in the reverse direction must dominate refolding, as in the staircase model, since these experiments are done at equilibrium. These states are suggested in Figure 1A by I_1, I_2, I_3, and U.

In apomyoglobin, the same A/G/H helix intermediate seen in the equilibrium molten globule (68) is seen also as a kinetic folding intermediate (75). Jamin & Baldwin provide further evidence that defined apomyoglobin intermediates, with the native-like AGH and B helices formed, occur sequentially in a productive pathway (73). In RNase H, the same A/D helix intermediate is seen as a molten globule at low pH (38), as a high free energy intermediate under native conditions (26), and as a kinetic folding intermediate (114).

Bai has demonstrated the on-pathway nature of the kinetically blocked N/C helix intermediate of cyt c by showing that it reaches the native state faster than it could cycle back through the unfolded state (5). Laurents et al (85) have demonstrated the on-pathway nature of a kinetic RNase A intermediate by extending the length of the HX labeling pulse to show that the intermediate moves forward to the native state without going back through the unfolded state where the protected H-label would have been lost. In these advances, on-pathway is interpreted to mean that the trapped intermediate does not have to fully unfold in order to resume forward folding, although some degree of unfolding may well occur.

Intermediate accumulation in the slow three-state folding of a number of proteins does often depend on the trapping of some misfolded structure (see below) but apparently not in deep off-pathway wells. In cyt c the misfold-dependent slowing can be produced even at pH 4.9 where the histidine to heme misligation itself is unstable (pKa is 5.6) and contributes no stability to the intermediate populated (128). The same is true in BPTI, where two cysteine residues are prematurely buried in the reforming native structure so that a solvent-catalyzed disulfide bridge formation is blocked (137). A RNase T1 kinetic intermediate appears to show an analogous structural situation in which a mis-isomerized proline has steered some segments out of place where they become trapped, whereas other parts of the intermediate are impressively native-like, as shown by NMR (11). The nonnative proline isomer provides no obvious stabilization energy to the populated RNase T1 intermediate. The same is true of proline-dependent slowing and intermediate accumulation in general. In all of these cases, some distinct native-like intermediate is seen and it achieves major population. This is as expected for a discrete folding pathway but seems unlikely in a situation where many different off-pathway wells may be accidentally encountered.

These results argue that the discrete native-like metastable intermediates seen to populate the reaction surface act as kinetic intermediates in discrete stepwise folding pathways. It should be stressed that distinct folding pathways need not be rigorously sequential. For example, in the four helix bundle ($H_1H_2H_3H_4$) apo cyt b562 protein, a branched folding sequence suggested by native state HX can be written as U to H_2H_3 to $H_1H_2H_3$ or $H_2H_3H_4$ to $H_1H_2H_3H_4$ = N (60, 61). A similar conclusion is indicated for barstar (145).

These considerations are in line with the framework and hierarchic condensation models for protein folding which focus on a central role for native-like intermediates based on secondary structure formation (14, 15), although the present considerations do not bear on whether secondary or tertiary structure forms first.

THE KINETIC DETERMINANTS: Nucleation and Misfolding Barriers

The different folding paradigms considered in Figure 1 provide very different explanations for the disparate folding behavior shown by different proteins. In the continuum view (Figure 1*B*) fast two-state folding proceeds down a smooth landscape without specific intermediates, while slow three-state folding and heterogeneity are caused by the accidental trapping of non-obligatory intermediates in deep off-pathway wells. A classical pathway with discrete obligatory intermediates requires a different explanation. One alternative is that two-state folding occurs because the rate-limiting process is the first step in the pathway, whereas three-state folding and heterogeneity are caused by later, optional, error-dependent barriers.

Three-State Folding and Heterogeneity

Are proteins that fold in a two-state or a three-state manner different in some fundamental way? Apparently not. For example, cyt c can be induced to fold in either a fast two-state manner (msecs) or a much slower three-state manner (∼sec), even though folding occurs under the same conditions in both cases (128). This special capability has made it possible to study the determining kinetic barriers and their structural bases.

When cyt c is unfolded, the weak Met80-S ligation to the heme iron dissociates and can be replaced by one of the two histidines in the green loop (Figure 3B) in a pH-dependent way with a pKa of 5.6. On dilution into folding conditions at pH 5, either two-state or three-state folding follows depending on the pH in the initial unfolding condition. At low unfolding pH, the histidines are protonated and do not bind the heme. On dilution into pH 5, cyt c folding is then fast and two-state. When the unfolding pH is above 5.6, a neutral histidine misligates to the heme iron, forcing the green loop segment in Figure 3 to the wrong side of the heme. The segment is trapped out of place by the early chain collapse when folding begins.

The N/C helix intermediate (blue in Figure 3) forms but then further folding is blocked. The newly inserted slow step ("misfold-reorganization") involves significant back unfolding, demonstrated by a reverse denaturant effect and by the fact that it requires about 300 msec while the histidine deligation by itself occurs in 15 msec. Evidently the slow step represents the time-consuming reorganization process necessary to free the trapped green segment so it can take its turn in the folding sequence (128).

The work on cyt c misfolding stimulated the influential 'New View' article of Baldwin (13) and a commentary by Creighton (35), which discussed the cyt c results in light of the view, based on theoretical studies, that discrete intermediates are actually misfolded off-pathway forms trapped in deep energy wells (Figure 1*B*). The cyt c study has therefore been widely cited as questioning the on-pathway nature of folding intermediates. In fact, Sosnick et al (128) considered that the misfolded form represents a normally occurring native-like on-pathway intermediate with, *in addition*, a mislocated and trapped segment.

An analogous situation has been seen for disulfide-reduced BPTI. A final step in BPTI refolding can be blocked by the native-like burial of two cysteines so that disulfide bridge formation cannot be catalyzed by added glutathione. As in cyt c, a kinetic barrier that involves some back unfolding necessary to allow disulfide formation slows folding and causes the intermediate to accumulate (137). As in cyt c, the intermediate is native-like but includes in addition an error, the premature burial of the unbridged cysteines.

A similar situation appears in RNase T1 where a nonnative proline isomer leads to slow folding (11). An intermediate accumulates in which part of the protein is strikingly native-like and could be characterized by NMR, while other segments are caught in a nonnative conformation. Apparently the well-known tendency of misisomerized prolines to produce slow three-state folding can also operate by inserting misfold-reorganization barriers.

In all of these cases, forward folding is blocked by some distinct structural misfolding error. Folding pauses. A native-like but corrupted error-containing intermediate accumulates. A time-consuming reorganization process is necessary to free the trapped groups so forward folding can resume. It does not seem to be the *population* of the misfolded intermediate state that *causes* the slow folding, for example in some deep well. Rather, slow folding and intermediate accumulation are both caused by the time-requiring error correction process, which can be viewed as an optionally inserted kinetic reorganization barrier (128), as in Figure 1*A*. The probabilistic nature of misfolding errors makes folding heterogeneous.

Protein folding can be described mathematically as a multidimensional process, as suggested by the funnel landscape diagram. In a multidimensional world, it is difficult to picture how on pathway barriers could block folding since proteins might simply bypass them in some higher dimension. Features that might slow folding would be limited to entropic search barriers and deep, necessarily optional wells. In reality, proteins operate in a three-dimensional world and kinetic folding barriers represent structural processes that should be considered in a 3-D structural

sense. The cases just noted provide examples of structurally based barriers that cannot be simply circumnavigated. Similar considerations hold for the issue of obligatory intermediates. The previous discussion suggests that in this case the determining structural issues involve built-in intrinsically cooperative structural elements and the interactions that allow early intermediates to stabilize later ones.

Two-State Folding

Sosnick et al (127, 129) further exploited the two-state to three-state switch in cyt c to study the position along the reaction coordinate of the rate-limiting barrier in two-state folding. The rate-limiting two-state barrier was found to be identical to the initial barrier in three-state folding (compare Figures 1*A* top and bottom). They have the same thermodynamic parameters (ΔG^*, ΔH^*, ΔS^*) and the same structural surface burial parameters (complex GmCl dependence expressed as m^* in the chevron folding limb and an earlier rollover). In three-state folding, the initial barrier obviously precedes (or accompanies) the formation of the early N/C helix intermediate. Therefore the identical barrier in two-state folding must also occur before (or during) formation of the early N/C helix intermediate, i.e. at the initial step of the pathway.

Sosnick et al extensively characterized the initial barrier and concluded that it represents an energetically uphill search for interactions that produce a native-like chain topology and can nucleate chain collapse (56, 127, 129). There is general agreement for a nucleation barrier in protein folding (1, 58, 71, 101, 127, 129, 131) but different workers endow the barrier with different characteristics and place it at different positions in the folding sequence. The rate-limiting nucleation step in two-state folding is often assigned to be late in the folding process. This is because (*a*) a large fraction of the surface that is buried in the native protein becomes buried in the transition state and (*b*) nucleation is assumed to lead directly to the final product. Rather, the cyt c results place the intrinsic rate-limiting barrier at the very first step, before intermediate formation. What is nucleated is the initial collapse step, and this is accompanied by a large surface burial. It appears that multiple native-like interactions are necessary to overcome the loop closure entropy that resists chain collapse, and that chain collapse in turn is necessary to form a docking surface that can stabilize subsequent secondary structure formation (127). Thus a massive nucleated collapse occurs as the first committed step in folding. The need for an initial long-range search to find a correct set of native interactions is supported by the correlation of two-state folding rates in many proteins with the contact order, the averaged sequence distance of interacting residues (106).

The large-scale search necessary to implement the initial native-like chain collapse most often requires a msec time scale and can be obscured by an even earlier sub-millisecond burst phase chain contraction (117). Considerable evidence now suggests that the kinetic burst phase does not reflect the formation of an intermediate in the usual sense. Rather it represents a more random solvent-dependent

polymer-like chain contraction on dilution from a good solvent (high denaturant) to a poor solvent (low denaturant) (2, 27, 56, 111, 127, 128, 130). In studied cases, the fast contraction produces signals (CD, fluorescence) identical to the U state characteristic for the same low denaturant (56, 111, 127, 130). The chain must then search for the native-like topology within the condensed phase prior to intermediate formation, leading to a kinetic slowing (rollover).

The nonspecific burst phase chain contraction is prominent at very low denaturant where hydrophobic interactions are strong and random interactions are probable. The slower nucleated collapse that produces the topologically native-like chain condensation can be seen in isolation at higher denaturant (on the chevron) where interactions are weaker and undirected burst phase interactions are suppressed. Lattice models simulate these two very different collapse modes when interactions are set to be either strong or weak, respectively (64).

If the very first on-pathway barrier is intrinsically large, the occurrence of two-state folding simply requires that the subsequent intrinsic barriers are smaller. In fact very fast on-pathway events have now been seen in a number of proteins and peptides (16, 24, 44, 105, 122, 138). It appears that ultrafast behavior can be explained in terms of events that occur after the initial nucleation barrier has been surpassed, since these experiments generally start with a certain amount of structure already in place (56).

CONCLUSIONS

The HX methods reviewed here have been able to provide fairly detailed structural information under the difficult conditions that are most pertinent to the folding problem. Many intermediate structures have now been characterized in kinetic folding experiments on a subsecond time scale, in labile molten globule forms under nonnative conditions, and in the high free-energy landscape under native conditions.

Results available show that distinct intermediates exist. They all turn out to represent partial replicas of the native protein. Clearly proteins are able to stabilize not only their lowest free-energy native state but also a small number of discrete partially folded native-like forms. These forms utilize as building blocks the intrinsically cooperative secondary structural elements of the protein and accordingly are robustly maintained over a wide range of conditions. The immense collection of all other possible forms must also exist, as implied in Figure 1*B*, but apparently only at higher free-energy levels than the few selected metastable intermediates (wells in Figure 1*A*). These conclusions relate to the equilibrium shape of the reaction landscape.

Additional evidence favors the view that these intermediates form distinct pathways in kinetic folding. It appears that globular proteins under folding conditions naturally condense into a native-like topology that acts to promote the formation of and to stabilize native-like secondary structural elements. The reversible folding

and unfolding of these intermediates will naturally produce stepwise pathways that lead from U to N.

Kinetic folding experiments are consistent with discrete pathways that include three kinds of kinetic barriers, all of which can be seen to represent conformational searches for the specific intermediates. An initial conformational search culminates by finding a set of interactions that pins the chain into a coarsely native-like topology, nucleating chain collapse and preparing the way for subsequent propagation steps. Propagation depends on short range searches that add cooperative secondary structural units to pre-existing structure. When all goes well, the initial search then becomes the rate-limiting process, intermediates do not accumulate, and folding appears to be a two-state process. However, proteins tend to make errors and misfold. Error correction involves some reorganizational back unfolding and can be slow, inserting an effective barrier that can slow folding and cause the prior normally occurring but flawed intermediate to accumulate. The probabilistic nature of error formation leads naturally to folding heterogeneity. This view, illustrated in Figure 1*A* (top plus bottom), provides a straightforward explanation for the apparently contradictory folding behavior of different proteins.

Theoretical simulations of non-protein models project a different view, with an unlimited number of continuously distributed intermediates and pathways, and slow folding when specific intermediates are accidentally populated (Figure 1*B*). However, experimental results for real globular proteins discussed here demonstrate the existence of a small number of distinct native-like folding intermediates and their likely participation in distinct pathways. The probable reason for this discrepancy is that the models usually simulated are analogous to molecules with a single cooperative unit. They do not contain the separable independently cooperative secondary structural units that determine intermediate formation in the folding of real globular proteins. To investigate theoretically the kinds of intermediates and pathways that are utilized by typical globular proteins, it will be necessary to study more protein-like models with separable cooperative units (86, 104, 135).

If proteins utilize as folding units their few cooperative structural elements rather than their many separate amino acid residues, then the folding problem is immensely simplified. The astronomical number of possible folding intermediates considered by Levinthal becomes irrelevant. The previously enigmatic amino acid code for the folding pathway becomes the very same code that determines the final native structure. More broadly, this conclusion would have important implications for protein design, both in the laboratory and in biological evolution, and perhaps for the goal of predicting protein-folding pathways and their resulting native structures.

ACKNOWLEDGMENTS

Helpful discussions with Leland Mayne, Tobin Sosnick, Yawen Bai, Buzz Baldwin, and Ken Dill are gratefully acknowledged. This work was supported by the National Institutes of Health and the Mathers Charitable Trust.

Visit the Annual Reviews home page at www.AnnualReviews.org

LITERATURE CITED

1. Abkevich VI, Gutin AM, Shakhnovich EI. 1994. Specific nucleus as the transition state for protein folding: evidence from the lattice model. *Biochemistry* 33:10026–36
2. Agashe VR, Shastry MC, Udgaonkar JB. 1995. Initial hydrophobic collapse in the folding of barstar. *Nature* 377:754–57
3. Alexandrescu AT, Evans PA, Pitkeathly M, Baum J, Dobson CM. 1993. Structure and dynamics of the acid-denatured molten globule state of α-lactalbumin: A two-dimension NMR study. *Biochemistry* 32:1707–18
4. Bahar J, Wallqvist A, Covell DG, Jernigan RL. 1998. Correlation between native state hydrogen exchange and cooperative residue fluctuations from a simple model. *Biochemistry* 37:1067–75
5. Bai Y. 1999. Kinetic evidence for an on-pathway intermediate in the folding of cytochrome c. *Proc. Natl. Acad. Sci. USA* 96:477–80
6. Bai Y, Englander JJ, Mayne L, Milne JS, Englander SW. 1995. Thermodynamic parameters from hydrogen exchange measurements. *Methods Enzymol.* 259:344–56
7. Bai Y, Englander SW. 1996. Future directions in folding: the multi-state nature of protein structure. *Proteins: Struct. Funct. Genet.* 24:145–51
8. Bai Y, Milne JS, Mayne L, Englander SW. 1993. Primary structure effects on peptide group hydrogen exchange. *Proteins: Struct. Funct. Genet.* 17:75–86
9. Bai Y, Milne JS, Mayne L, Englander SW. 1994. Protein stability parameters measured by hydrogen exchange. *Proteins: Struct. Funct. Genet.* 20:4–14
10. Bai Y, Sosnick TR, Mayne L, Englander SW. 1995. Protein folding intermediates studied by native state hydrogen exchange. *Science* 269:192–97
11. Balbach J, Steegborn C, Schindler T, Schmid FX. 1999. A protein folding intermediate of ribonuclease T1 characterized at high resolution by 1D and 2D NMR spectroscopy. *J. Mol. Biol.* 285:829–42
12. Baldwin RL. 1993. Pulsed H/D-exchange studies of folding intermediates. *Curr. Opin. Struct. Biol.* 3:84–91
13. Baldwin RL. 1995. The nature of protein folding pathways: the classical versus the new view. *J. Biomol. NMR* 5:103–9
14. Baldwin RL, Rose GD. 1999. Is protein folding hierarchic? I. Local structure and peptide folding. *Trends Biochem. Sci.* 24:26–33
15. Baldwin RL, Rose GD. 1999. Is protein folding hierarchic? II. Folding intermediates and transition states. *Trends Biochem. Sci.* 24:77–83
16. Ballew RM, Sabelko J, Gruebele M. 1996. Direct observation of fast protein folding: the initial collapse of apomyoglobin. *Proc. Natl. Acad. Sci. USA* 93:5759–64
17. Barrick D, Baldwin RL. 1993. The molten globule intermediate of apomyoglobin and the process of protein folding. *Protein Sci.* 2:869–876
18. Baum J, Dobson CM, Evans PA, Hanley C. 1989. Characterization of a partly folded protein by NMR methods: studies on the molten globule state of guinea pig α-lactalbumin. *Biochemistry* 28:7–13
19. Bhuyan AK, Udgaonkar JB. 1998. Two structural subdomains of barstar detected by rapid mixing NMR measurement of amide hydrogen exchange. *Proteins: Struct. Funct. Genet.* 30:296–308
20. Brems DN, Baldwin RL. 1984. Amide proton exchange used to monitor the formation of a stable a-helix by residues 3-13 during folding of ribonuclease S. *J. Mol. Biol.* 180:1141–56
21. Brems DN, Baldwin RL. 1985. Protection

of amide protons in folding intermediates of ribonuclease A measured by pH-pulse exchange curves. *Biochemistry* 24:1689–93
22. Brooks CL 3rd. 1998. Simulations of protein folding and unfolding. *Curr. Opin. Struct. Biol.* 8:222–26
23. Bryngelson JD, Onuchic JN, Socci ND, Wolynes PG. 1995. Funnels, pathways, and the energy landscape of protein folding:a synthesis. *Proteins: Struct. Funct. Genet.* 21:167–95
24. Burton RE, Huang GS, Dougherty MA, Fullbright PW, Oas TG. 1996. Microsecond protein folding through a compact transition state. *J. Mol. Biol.* 263:311–22
25. Bycroft M, Matouschek A, Kellis JT Jr, Serrano L, Fersht AR. 1990. Detection and characterization of a folding intermediate in barnase by NMR. *Nature* 346:488–90
26. Chamberlain AK, Handel TM, Marqusee S. 1996. Detection of rare partially folded molecules in equilibrium with the native conformation of RNase H. *Nature Struct. Biol.* 3:782–87
27. Chan CK, Hu Y, Takahashi S, Rousseau DL, Eaton WA, et al. 1997. Submillisecond protein folding kinetics studied by ultrarapid mixing. *Proc. Natl. Acad. Sci. USA* 94:1779–84
28. Chung EW, Nettleton EJ, Morgan CJ, Gross M, Miranker A, et al. 1997. Hydrogen exchange properties of proteins in native and denatured states monitored by mass spectrometry and NMR. *Protein Sci.* 6:1316–24
29. Chyan CL, Wormald C, Dobson CM, Evans PA, Baum J. 1993. Structure and stability of the molten globule state of guinea pig α-lactalbumin: a hydrogen exchange study. *Biochemistry* 32:5681–91
30. Clarke J, Itzhaki LS. 1998. Hydrogen exchange and protein folding. *Curr. Opin. Struct. Biol.* 8:112–18
31. Clarke J, Itzhaki LS, Fersht AR. 1997. Hydrogen exchange at equilibrium: a short cut for analysing protein–folding pathways? *Trends Biochem. Sci.* 22:284–87
32. Clarke J, Itzhaki LS, Fersht AR. 1998. A reply to Englander and Woodward. *Trends Biochem Sci.* 23:379–81
33. Connelly GP, Bai Y, Jeng M-F, Mayne L, Englander SW. 1993. Isotope effects in peptide group hydrogen exchange. *Proteins: Struct. Funct. Genet.* 17:87–92
34. Creighton TE. 1991. Characterizing intermediates in protein folding. *Current Biol.* 1:8–10
35. Creighton TE. 1994. The energetic ups and downs of protein folding. *Nature Struct. Biol.* 1:135–38
36. Creighton TE. 1995. Protein folding. An unfolding story. *Current Biol.* 5:353–56
37. Creighton TE. 1997. How important is the molten globule for correct protein folding? *Trends Biochem. Sci.* 22:6–10
38. Dabora JM, Pelton JG, Marqusee S. 1996. Structure of the acid state of *Escherichia coli* ribonuclease HI. *Biochemistry* 35:11951–58
39. Dalby PA, Clarke J, Johnson CM, Fersht AR. 1998. Folding intermediates of wild type and mutants of barnase. II. Correlation of changes in equilibrium amide exchange etc. *J. Mol. Biol.* 276:647–56
40. Dill KA, Chan HS. 1997. From Levinthal to pathways to funnels. *Nature Struct. Biol.* 4:10–19
41. Dill KA, Shortle D. 1991. Denatured states of proteins. *Annu. Rev. Biochem.* 60:795–825
42. Dobson CM. 1991. Characterization of folding intermediates. *Curr. Opin. Struct. Biol.* 1:22–27
43. Dolgikh DA, Gilmanshin RI, Brazhnikov EV, Bychkova VE, Semisotnov GV, et al. 1981. Alpha–lactalbumin:compact state with fluctuating tertiary structure? *FEBS Lett.* 136:311–315
44. Eaton WA, Munoz V, Thompson PA, Henry ER, Hofrichter J. 1998. Kinetics and dynamics of α–helices, β–hairpins, and fast-folding proteins. *Accts. Chem. Res.* 31:745–54

45. Eigen M. 1964. Proton transfer, acid-base catalysis, and enzymatic hydrolysis. *Angew. Chem. Intl. Ed. English* 3:1–19
46. Elöve GA, Roder H. 1990. Structure and stability of cytochrome c folding intermediates. In *Protein Refolding*, ed. G Georgiou, E De Bernadez-Clark, pp. 50–63. Washington, DC: Am. Chem. Soc.
47. Englander JJ, Rumbley JN, Englander SW. 1998. Intramolecular signal transduction: cross subunit effects in the hemoglobin T-state. *J. Mol. Biol.* 284:1707–16
48. Englander SW. 1998. Native state HX. *Trends Biochem. Sci.* 23:378
49. Englander SW, Calhoun DB, Englander JJ, Kallenbach NR, Liem RKH, et al. 1980. Individual breathing reactions measured in hemoglobin by hydrogen exchange methods. *Biophys. J.* 32:577–90
50. Englander SW, Kallenbach NR. 1984. Hydrogen exchange and structural dynamics of proteins and nucleic-acids. *Q. Rev. Biophys.* 16:521–655
51. Englander SW, Mayne L. 1992. Protein folding studied using hydrogen-exchange labeling and two-dimensional NMR. *Annu. Rev. Biophys. Biomol. Struct.* 21:243–65
52. Englander SW, Mayne L, McKinnie RE, Paterson Y, Englander JJ. 1991. Protein interaction, folding and energetics studied by hydrogen exchange labeling. In *Molecular Conformation and Biological Interactions*, ed. P Balaram, S Ramaseshan, pp. 245–68. Bangalore: Indian Acad. Sci.
53. Englander SW, Mayne LC, Bai Y, Sosnick TR. 1997. Hydrogen exchange:the modern legacy of Linderstrom-Lang. *Protein Sci.* 6:1101–9
54. Englander SW, Milne J. 1990. The protein folding problem studied by hydrogen exchange and 2D NMR. In *Protein Structure and Function*, ed. ZZH Karachi/New York:Twel
55. Englander SW, Sosnick TR, Englander JJ, Mayne L. 1996. Mechanisms and uses of hydrogen exchange. *Curr. Opin. Struc. Biol.* 6:18–23
56. Englander SW, Sosnick TR, Mayne LC, Shtilerman M, Qi PX, et al. 1998. Fast and slow folding in cytochrome c. *Accts. Chem. Res.* 31:737–44
57. Feng Y, Sligar SG, Wand AJ. 1994. Solution structure of apocytochrome b562. *Nature Struct. Biol.* 1:30–35
58. Fersht AR. 1997. Nucleation mechanisms in protein folding. *Curr. Opin. Struct. Biol.* 7:3–9
59. Fink AL. 1995. Compact intermediate states in protein folding. *Annu. Rev. Biophys. Biomol. Struct.* 24:495–522
60. Fuentes EJ, Wand AJ. 1998. Local dynamics and stability of apocytochrome b562 examined by hydrogen exchange. *Biochemistry* 37:3687–98
61. Fuentes EJ, Wand AJ. 1998. The local stability and dynamics of apocytochrome b562 examined by the dependence of hydrogen exchange on hydrostatic pressure. *Biochemistry* 37:3687–98
62. Gladwin ST, Evans PA. 1996. Structure of very early folding intermediates: new insights through a variant of hydrogen exchange labeling. *Folding Design* 1:407–17
63. Gregory RB, Rosenberg A. 1986. Protein conformational dynamics measured by hydrogen isotope exchange techniques. *Methods Enzymol.* 131:448–508
64. Gutin AM, Abkevich VI, Shakhnovich EI. 1995. Is burst hydrophobic collapse necessary for protein folding? *Biochemistry* 34:3066–76
65. Haynie DT, Freire E. 1993. Structural energetics of the molten globule state. *Proteins: Struct. Funct. Genet.* 16:115–40
66. Hiller R, Zhou ZH, Adams MWW, Englander SW. 1997. Stability and dynamics in a hyperthermophilic protein with melting temperature close to 200°C. *Proc. Natl. Acad. Sci. USA* 94:11329–32
67. Hilser VJ, Freire E. 1996. Structure based calculation of the equilibrium folding pathway of proteins. Correlation with hydrogen exchange protection factors. *J. Mol. Biol.* 262:756–72

68. Hughson FM, Wright PE, Baldwin RL. 1990. Structural characterization of a partly folded apomyoglobin intermediate. *Science* 249:1544–48
69. Huyghues-Despointes BMP, Scholtz JM, Pace CN. 1999. Protein conformational stabilities can be determined from hydrogen exchange rates. *Nature Struct. Biol.* 6:910–12
70. Hvidt A, Nielsen SO. 1966. Hydrogen exchange in proteins. *Adv. Protein Chem.* 21:287–386
71. Itzhaki LS, Otzen DE, Fersht AR. 1995. The structure of the transition state for folding of chymotrypsin inhibitor 2 analysed by protein engineering methods:evidence for a nucleation–condensation mechanism for protein folding. *J. Mol. Biol.* 254:260–88
72. Jacobs MD, Fox RO. 1994. Staphylococcal nuclease folding intermediate characterized by hydrogen exchange and NMR spectroscopy. *Proc. Natl. Acad. Sci. USA* 91:449–53
73. Jamin M, Baldwin RL. 1999. Two forms of the pH 4 folding intermediate of apomyoglobin. *J. Mol. Biol.* 276:491–504
74. Jeng M-F, Englander SW, Elöve GA, Wand AJ, Roder H. 1990. Structural description of acid-denatured cytochrome c by hydrogen exchange and 2D NMR. *Biochemistry* 29:10433–37
75. Jennings PA, Wright PE. 1993. Formation of a molten globule intermediate early in the kinetic folding pathway of apomyoglobin. *Science* 262:892–96
76. Karplus M, Sali A. 1995. Theoretical studies of protein folding and unfolding. *Curr. Opin. Struct. Biol.* 5:58–73
77. Kim KS, Fuchs JA, Woodward CK. 1993. Hydrogen exchange identifies native-state motional domains important in protein folding. *Biochemistry* 32:9600–8
78. Kim PS. 1986. Amide proton exchange as a probe of protein folding pathways. *Methods Enzymol.* 131:136–57
79. Kim PS, Baldwin RL. 1980. Structural intermediates trapped during the folding of ribonuclease A by amide proton exchange. *Biochemistry* 19:6124–29
80. Kim PS, Baldwin RL. 1982. Specific intermediates in the folding reactions of small proteins and the mechanism of protein folding. *Annu. Rev. Biochem.* 51:459–89
81. Kim PS, Baldwin RL. 1990. Intermediates in the folding reactions of small proteins. *Annu. Rev. Biochem.* 59:631–60
82. Kita N, Kuwajima K, Nitta K, Sugai S. 1976. Equilibrium and kinetics of the unfolding of alpha–lactalbumin by guanidine hydrochloride (II). *Biochim. Biophys. Acta* 427:350–58
83. Kuwajima K. 1989. The molten globule state as a clue for understanding the folding and cooperativity of globular protein structure. *Proteins: Struct. Funct. Genet.* 6:87–103
84. Kuwajima K, Nitta K, Yoneyama M, Sugai S. 1976. Three-state denaturation of alpha-lactalbumin by guanidine hydrochloride. *J. Mol. Biol.* 106:359–73
85. Laurents DV, Bruix M, Jamin M, Baldwin RL. 1998. A pulse-chase-competition experiment to determine if a folding intermediate is on or off-pathway:application to ribonuclease A. *J. Mol. Biol.* 283:669–78
86. Lazaridis T, Karplus M. 1997. "New View" of protein folding reconciled with the old through multiple unfolding simulations. *Science* 278:1928–31
87. Linderstrøm-Lang KU. 1958. Deuterium exchange and protein structure. In *Symposium on Protein Structure*, ed. A. Neuberger. London: Methuen
88. Loh SN, Kay MS, Baldwin RL. 1995. Structure and stability of a second molten globule intermediate in the apomyoglobin folding pathway. *Proc. Natl. Acad. Sci. USA* 92:5446–50
89. Loh SN, Prehoda KE, Wang J, Markley JL. 1993. Hydrogen exchange in unligated and ligated staphylococcal nuclease. *Biochemistry* 32:11022–28

90. Lu J, Dahlquist FW. 1992. Detection and characterization of an early folding intermediate of T4 lysozyme using pulsed hydrogen exchange and two-dimensional NMR. *Biochemistry* 31:4749–56
91. Matouschek A, Serrano L, Meiering EM, Bycroft M, Fersht AR. 1992. The folding of an enzyme V. 1H/2H exchange-nuclear magnetic resonance studies on the folding pathway of barnase: complementarity to and agreement with protein engineering studies. *J. Mol. Biol.* 224:837–45
92. Matthews CR. 1991. The mechanism of protein folding. *Curr. Opin. Struct. Biol.* 1:28–35
93. Matthews CR. 1993. Pathways of protein folding. *Annu. Rev. Biochem.* 62:653–83
94. Milne JS, Mayne L, Roder H, Wand AJ, Englander SW. 1998. Determinants of protein hydrogen exchange studied in equine cytochrome c. *Protein Sci.* 7:739–45
95. Milne JS, Xu Y, Mayne LC, Englander SW. 1999. Experimental study of the protein folding landscape:unfolding reactions in cytochrome c. *J. Mol. Biol.* 290:811–22
96. Miranker A, Robinson CV, Radford AE, Dobson CM. 1996. Investigation of protein folding by mass spectrometry. *FASEB J.* 10:93–101
97. Mirny LA, Abkevich V, Shakhnovich EI. 1996. Universality and diversity of folding scenarios:a comprehensive analysis with the aid of a lattice model. *Folding Design* 1103–116
98. Molday RS, Englander SW, Kallen RG. 1972. Primary structure effects on peptide group hydrogen exchange. *Biochemistry* 11:150–58
99. Morozova LA, Haynie DT, Arico-Muendel C, Van Dael H, Dobson CM. 1995. Structural basis of the stability of a lysozyme molten globule. *Nature Struct. Biol.* 2:871–75
100. Mullins LS, Pace CN, Raushel FM. 1993. Investigation of ribonuclease T1 folding intermediates by hydrogen-deuterium amide exchange-two-dimensional NMR spectroscopy. *Biochemistry* 32:6152–56
101. Oliveberg M. 1998. Alternative explanations for multi-state kinetics in protein folding: transient aggregation and changing transition state ensembles. *Accts. Chem. Res.* 31:765–72
102. Onuchic JN, Wolynes PG, Luthey-Schulten Z, Socci ND. 1995. Toward an outline of the topography of a realistic protein-folding funnel. *Proc. Natl. Acad. Sci. USA* 92:3626–30
103. Pande VS, Grosberg A, Tanaka T, Rokhsar DS. 1998. Pathways for protein folding: is a new view needed? *Curr. Opin. Struct. Biol.* 8:68–79
104. Pande VS, Rokhsar DS. 1999. Folding pathway of a lattice model for proteins. *Proc. Natl. Acad. Sci.* 96:1273–78
105. Phillips CM, Mizutani Y, Hochstrasser RM. 1995. Ultrafast thermally induced unfolding of RNase A. *Proc. Natl. Acad. Sci. USA* 92:7292–96
106. Plaxco KW, Simons KT, Baker D. 1998. Contact order, transition state placement and the refolding rates of single domain proteins. *J. Mol. Biol.* 277:985–94
107. Privalov PL. 1996. Intermediate states in protein folding. *J. Mol. Biol.* 258:707–25
108. Ptitsyn OB. 1987. Protein folding: hypotheses and experiments. *J. Protein Chem.* 6:273–93
109. Ptitsyn OB. 1995. Molten globule and protein folding. *Adv. Protein Chem.* 47: 83–229
110. Ptitsyn OB. 1995. Structures of folding intermediates. *Curr. Opin. Struc. Biol.* 5:74–78
111. Qi PX, Sosnick TR, Englander SW. 1998. The burst phase in ribonuclease A folding: solvent dependence of the unfolded state. *Nat. Struct. Biol.* 5:882–84
112. Radford SE, Buck M, Topping KD, Dobson CM, Evans PA. 1992. Hydrogen exchange in native and denatured states

of hen egg-white lysozyme. *Proteins: Struct. Funct. Genet.* 14:237–48
113. Radford SE, Dobson CM, Evans PA. 1992. The folding of hen lysozyme involves partially structured intermediates and multiple pathways. *Nature* 358:302–7
114. Raschke TM, Marqusee S. 1997. The kinetic folding intermediate of ribonuclease H resembles the acid molten globule and partially unfolded molecules detected under native conditions. *Nature Struct. Biol.* 4:298–304
115. Redfield C, Smith RA, Dobson CM. 1994. Structural characterization of a highly-ordered 'molten globule' at low pH. *Nature Struct. Biol.* 1:23–29
116. Roder H. 1989. Structural characterization of protein folding intermediates by proton magnetic resonance and hydrogen exchange. *Methods Enzymol.* 176:447–73
117. Roder H, Colon W. 1997. Kinetic role of early intermediates in protein folding. *Curr. Opin. Struct. Biol.* 7:15–28
118. Roder H, Elöve GA, Englander SW. 1988. Structural characterization of folding intermediates in cytochrome c by H-exchange labeling and proton NMR. *Nature* 335:700–4
119. Roder H, Wuthrich K. 1986. Protein folding kinetics by combined use of rapid mixing techniques and NMR observation of individual amide protons. *Proteins: Struc. Funct. Genet.* 1:34–42
120. Rosa JJ, Richards FM. 1979. An experimental procedure for increasing the structural resolution of chemical hydrogen exchange measurements on proteins: application to ribonuclease S peptide. *J. Mol. Biol.* 133:399–416
121. Rosenberg A, Chakravarti K. 1968. Studies of hydrogen exchange in proteins. I. The exchange kinetics of bovine carbonic anhydrase. *J. Biol. Chem.* 243:5193–5201
122. Sabelko J, Ervin J, Gruebele M. 1999. Observation of strange kinetics in protein folding [see comments]. *Proc. Natl. Acad. Sci. USA* 96:6031–36
123. Schmid FX, Baldwin RL. 1979. Detection of an early intermediate in the folding of ribonuclease A by protection of amide protons against exchange. *J. Mol. Biol.* 135:199–215
124. Schulman BA, Kim PS, Dobson CM, Redfield C. 1997. A residue-specific NMR view of the non-cooperative unfolding of a molten globule. *Nature Struct. Biol.* 4:630–34
125. Schulman BA, Redfield C, Peng ZY, Dobson CM, Kim PS. 1995. Different subdomains are most protected from hydrogen exchange in the molten globule and native states of human alpha-lactalbumin. *J. Mol. Biol.* 253:651–57
126. Shakhnovich EI. 1997. Theoretical studies of protein-folding thermodynamics and kinetics. *Curr. Opin. Struct. Biol.* 7:29–40
127. Sosnick TR, Mayne L, Englander SW. 1996. Molecular collapse: the rate-limiting step in two-state cytochrome c folding. *Proteins: Struct. Funct. Genet.* 24:413–26
128. Sosnick TR, Mayne L, Hiller R, Englander SW. 1994. The barriers in protein folding. *Nature Struct. Biol.* 1:149–56
129. Sosnick TR, Mayne LC, Hiller R, Englander SW. 1995. The barriers in protein folding. In *Peptide and Protein Folding Workshop*, ed. W. F. Degrado. Philadelphia: Int. Bus. Commun.
130. Sosnick TR, Shtilerman MD, Mayne L, Englander SW. 1997. Ultrafast signals in protein folding and the polypeptide contracted state. *Proc. Natl. Acad. Sci. USA* 94:8545–50
131. Thirumalai D, Guo Z. 1995. Nucleation mechanism for protein folding and theoretical predictions for hydrogen-exchange experiments. *Biopolymers* 35:137–40
132. Thirumalai D, Klimov DK. 1999. Deciphering the timescales and mechanisms of protein folding using minimal off-lattice

models. *Curr. Opin. Struct. Biol.* 9:197–207
133. Udgaonkar JB, Baldwin RL. 1988. NMR evidence for an early framework intermediate on the folding pathway of ribonuclease A. *Nature* 335:694–99
134. Udgaonkar JB, Baldwin RL. 1990. Early folding intermediate of ribonuclease A. *Proc. Natl. Acad. Sci. USA* 87:8197–8201
135. Veitshans T, Klimov D, Thirumalai D. 1997. Protein folding kinetics:timescales, pathways and energy landscapes in terms of sequence-dependent properties. *Folding Design* 2:1–22
136. Wagner G, Wüthrich K. 1982. Amide proton exchange and surface conformation of BPTI in solution:studies with 2D NMR. *J. Mol. Biol.* 160:343–61
137. Weissman JS, Kim PS. 1995. A kinetic explanation for the rearrangement pathway of BPTI folding. *Nature Struct. Biol.* 2:1123–30
138. Williams S, Causgrove TP, Gilmanshin R, Fang KS, Callender RH, et al. 1996. Fast events in protein folding:helix melting and formation in a small peptide. *Biochemistry* 35:691–97
139. Wolynes PG, Onuchic JN, Thirumalai D. 1995. Navigating the folding routes. *Science* 267:1619–20
140. Woodward C. 1993. Is the slow exchange core the protein folding core? *Trends Biochem. Sci.* 18:359–60
141. Woodward CK. 1994. Hydrogen exchange rates and protein folding. *Curr. Opin. Struc. Biol.* 4:112–16
142. Woodward CK, Hilton BD. 1979. Hydrogen exchange kinetics and internal motions in proteins and nucleic acids. *Annu. Rev. Biophys. Bioeng.* 8:99–127
143. Woodward CK, Simon I, Tuchsen E. 1982. Hydrogen exchange and the dynamic structure of proteins. *Mol. Cell. Biochem.* 48:135–60
144. Xu Y, Mayne L, Englander SW. 1998. Evidence for an unfolding and refolding pathway in cytochrome c. *Nature Struct. Biol.* 5:774–78
145. Zaidi FN, Nath U, Udgaonkar JB. 1997. Multiple intermediates and transition states during protein unfolding. *Nature Struct. Biol.* 4:1016–24
146. Zhang Z, Post CB, Smith DL. 1996. Amide hydrogen exchange determined by mass spectrometry:application to rabbit muscle aldolase. *Biochemistry* 35:779–91

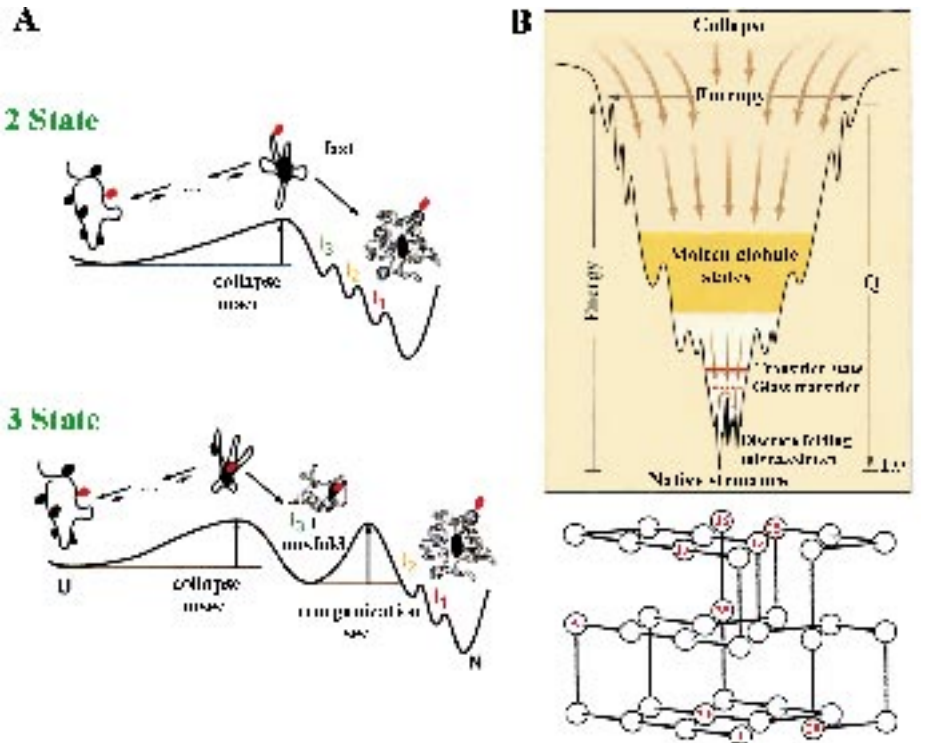

Figure 1 Alternative views of the folding reaction landscape. A. Conventional free-energy diagrams that represent two-state folding with an initial limiting barrier (*top*), three-state folding with the same initial barrier and a subsequently inserted misfold-reorganization barrier (bottom), and kinetic intermediates hidden or revealed by these barriers. Following Sosnick et al (127–129), the initial barrier is a search process that brings large hydrophobes (*black residues*) together to pin the chain into a native-like topology in the nucleated collapse. The segment carrying the *red residue* remains free in the nucleated collapse that leads to fast two-state folding. When it is internally trapped, it produces a kinetic reorganization barrier that causes slow folding, intermediate accumulation, and folding heterogeneity (127–129). B. A funnel-shaped reaction landscape, adapted from Wolynes et al (139), and a typical on-lattice model used to explore it, adapted from Abkevich et al (1). This landscape posits an unlimited number of intermediates and pathways. The accidental trapping and accumulation of defined intermediates causes slow folding and heterogeneity.

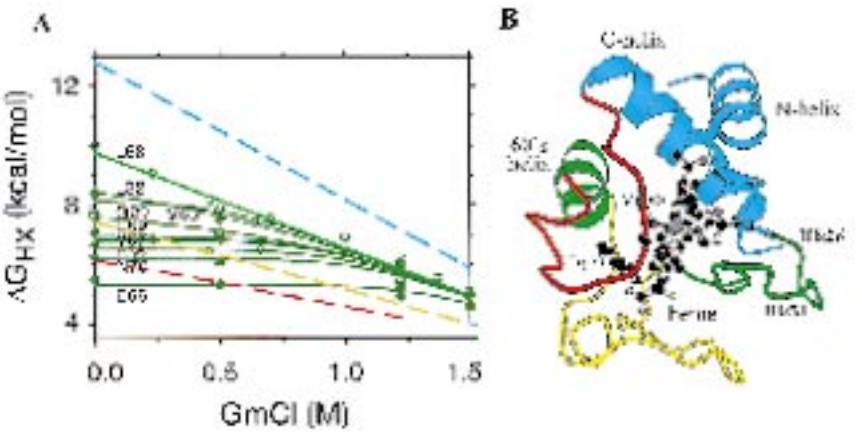

Figure 3 Native-state HX results for cyt c (10, 95). Data are shown for all of the slow amide hydrogens in the *green* cooperative element (60s helix plus the histidine-containing loop), which exchange through local fluctuations initially but become dominated by a recognizable larger unfolding when it is promoted by increasing GmCl. The *dashed lines*, color coded to match the cyt c diagram in panel B, indicate HX isotherms produced by sets of hydrogens that identify other cooperatively unfolding structural units. Conditions are pD 7, 30°C.

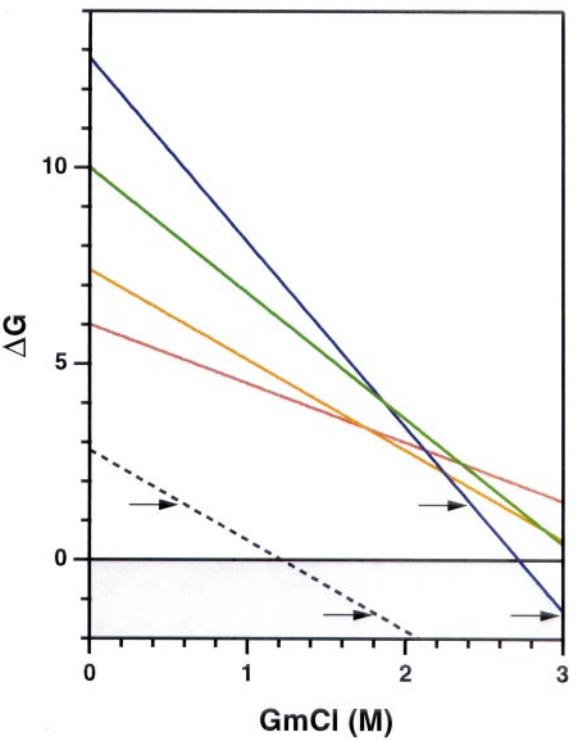

Figure 4 Free-energy cross over curve for the cooperative units in cyt c obtained from the native state HX results in Figure 3 (7). When stability is high, subglobal unfoldings can be seen separately by native state HX. When denaturant is increased, the larger global unfolding is most sharply promoted and, where it can be measured by the usual means (*arrows*), appears to be a two-state process. The *dashed curve* suggests a subglobal unfolding that crosses N earlier, which would produce a molten globule and would make the global unfolding transition appear to be less cooperative.