Protein Folding Pathways Studied by Pulsed- and Native-State Hydrogen Exchange

Yawen Bai†

Laboratory of Biochemistry, National Cancer Institute, Building 37, Room 6114E, National Institutes of Health, Bethesda, Maryland 20892

Received August 25, 2004

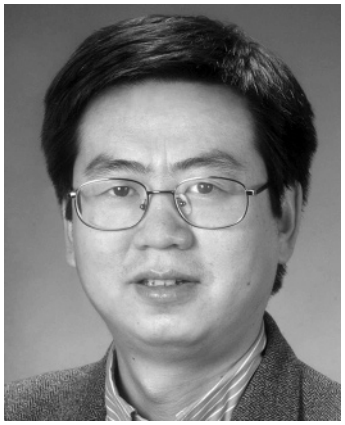
Contents

1. Introduction

Protein folding is the last step of the central dogma of molecular biology, which involves the transfer of genetic information from one-dimensional sequences of amino acids to three-dimensional structures. In the 1960s, Anfinsen and co-workers demonstrated that proteins could fold spontaneously from the unfolded state to the native state. It was hypothesized that the native state is the most stable conformation among the astronomical number of possible conformations available for a given polypeptide chain and that this native structure is determined by the amino acid sequence under the physiological environment.[1] Since then, studies of the process of protein folding and prediction of protein structures from protein sequences have become the two major topics in the field of protein folding. They have fascinated many researchers for more than three decades but remain unsolved.

Two recent events have further spurred protein folding studies. First, a number of human diseases, including Alzheimer's, type II diabetes, and CJD (a human version of mad cow disease) have been related to protein misfolding and aggregation. In some cases, it was shown that partially unfolded intermediates are the precursors for the formation of amyloid fibers.[2] Therefore, studies of the process of protein folding are important for understanding the mechanism of amyloid diseases. Second, the human genome project has accumulated a huge number of protein sequences. To understand the functional information encoded in these sequences, however, the three-dimensional structures of proteins need to be known. Thus, it is highly desirable to predict protein structures by a computational method. Because all computer programs for predicting high-resolution structures eventually will have to rely on physically realistic force-field parameters, detailed physical studies on the factors that affect the stability and folding of proteins are essential.

The central goal of current experimental studies of protein folding is to characterize the process of protein folding and to find the rules that control it. To characterize the pathway or the free-energy landscape of protein folding, which includes folding intermediates and transition states, two major methods have been developed to obtain their structural information at the level of residues: amide hydrogen exchange and protein engineering. Between the two methods, amide hydrogen exchange plays a dominant role in detecting and



Dr. Yawen Bai did his undergraduate training at the Jilin University in China and received his B.A. degree in biochemistry in 1986. He studied quantum chemistry as a graduate student at the Institute of Theoretical Chemistry at the Jilin University from 1986 to 1989. He earned his M.S. degree in theoretical chemistry at the University of Pennsylvania in 1991 and a Ph.D. degree in biophysics with S. Walter Englander at the Department of Biochemistry and Biophysics, Medical School, University of Pennsylvania, in 1994. From 1994 to 1997, he was a postdoc at the Scripps Research Institute, working with Peter E. Wright. In 1997, he joined the Laboratory of Biochemistry at the National Cancer Institute as a principal investigator.

† Telephone (301) 594-2375; fax (301) 402-3095; e-mail yawen@helix.nih.gov.

10.1021/cr040432i
Published on Web 03/14/2006

characterizing the partially unfolded intermediates, whereas protein engineering is more useful in obtaining structural information of the transition state. For the past 15 years, direct hydrogen exchange measurements have defined the structures of partially unfolded states under acidic conditions, which are not accessible to the usual crystallographic and NMR methods. The hydrogen exchange pulse-labeling method has specified the structure and stability of folding intermediates that exist for <1 s during folding processes. The native-state hydrogen exchange method has detected and characterized folding intermediates that exist as infinitesimally populated high-energy states under native conditions. A number of excellent reviews have been published on various aspects of these studies.[3−10] The focus of this review is on the recent results obtained from the pulse-labeling and native-state hydrogen exchange methods and their implications for understanding the mechanism of protein folding.

2. Amide Hydrogen Exchange of Proteins

2.1. Intrinsic Exchange Rates for Unfolded Polypeptides

Amide protons in polypeptides are chemically labile and can exchange with hydrogen isotopes in solvent water such as

$$>\mathrm{N{-}H} + \mathrm{D_2O} \rightarrow >\mathrm{N{-}H} + \mathrm{DOH} \qquad (1)$$

Because of the extreme pK_a values of main-chain amides, the exchange of their hydrogens with solvent is relatively slow and is catalyzed only by the strongest of aqueous acids and bases (hydronium and hydroxide ion). Thus, the exchange rate is pH-dependent. Figure 1 illustrates the exchange rate constants as a function of pD (pD = $\mathrm{pH_{read}}$ + 0.4)[11] for amide protons in unstructured poly D/L alanine. Here $\mathrm{pH_{read}}$ is the reading value from a pH-meter. At pD 7.0 and 20 °C, the exchange rate constant of an amide proton in an unfolded peptide, k_{int}, is affected mainly by the side chains of its two nearest amino acid residue neighbors. Both inductive[12] and steric blocking effects[13] are apparent. These effects have been characterized for all 20 amino acids using dipeptides as models.[14] k_{int} can now be predicted within a broad range of pH and temperature[14,15] (on-line at http://www.fccc.edu/research/labs/roder/sphere/). For unfolded polypeptides, the predicted k_{int} is likely to be within a factor of 2 of the measured k_{ex}.[14,16]

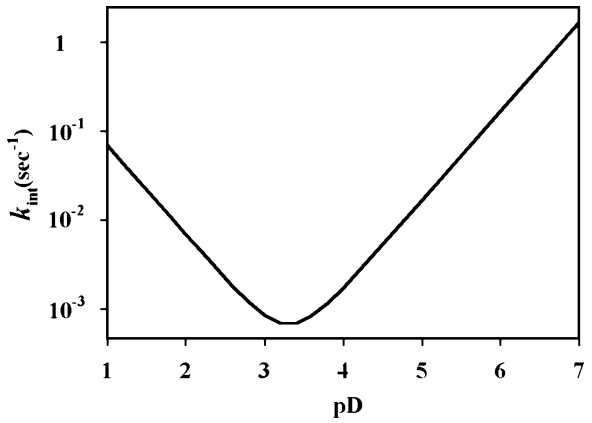


Figure 1. Intrinsic exchange rate constant as a function of pD (pD = $\mathrm{pH_{read}}$ + 0.4) at 20 °C for amide protons in PDLA.

2.2. Linderstrøm-Lang Model for Amide Hydrogen Exchange in Folded Proteins

In folded proteins, many amide protons are protected from exchange due to hydrogen bonding and burial in the native structure. The protection factor (PF = k_{int}/k_{ex}) provides information on the native structure and stability. Here k_{ex} is the experimentally measured exchange rate constant. Linderstrøm-Lang and his colleagues pictured a two-state situation and assumed that amide hydrogens can exchange with solvent hydrogens only when they are transiently exposed to solvent in some kind of closed-to-open reaction, as indicated in eq 2:[17]

$$\mathrm{NH(closed)} \underset{k_{cl}}{\overset{k_{op}}{\rightleftharpoons}} \mathrm{NH(open)} \xrightarrow{k_{int}} \mathrm{exchanged}; \quad K_{op} = k_{op}/k_{cl} \qquad (2)$$

Here, k_{op} is the kinetic opening-rate constant, and k_{cl} is the kinetic closing rate constant. Under steady-state conditions, the exchange rate, k_{ex}, determined by the above scheme is given by eq 3.

$$k_{ex} = k_{op}k_{int}/(k_{op} + k_{cl} + k_{int}) \qquad (3)$$

There are two extreme cases for this reaction scheme with stable structures ($k_{op} \ll k_{cl}$). (i) The closing reaction is much faster than the intrinsic exchange rate constants ($k_{cl} \gg k_{int}$), termed the EX2 condition. In this case, the exchange rate of any hydrogen, k_{ex}, is determined by its chemical exchange rate in the open form multiplied by the equilibrium opening constant, K_{op}.

$$k_{ex} = K_{op}k_{int} \qquad (4)$$

This leads to free energy for the dominant opening reaction, as represented by the following equation:

$$\Delta G_{HX} = -RT \ln K_{op} = -RT \ln(k_{ex}/k_{int}) = -RT \ln(1/PF) \qquad (5)$$

In this equation, R is the gas constant and T is the temperature. The free energy defined in eq 5, ΔG_{HX}, represents a combination of opening transitions from both structural unfolding and local fluctuations. (ii) The closing reaction is much slower than the intrinsic exchange rate constant ($k_{cl} \ll k_{int}$), termed the EX1 condition. In this case, the exchange rate, k_{ex}, is equal to the opening-rate constant k_{op}. For amide protons that can exchange only through global unfolding, the k_{op} will be the global unfolding-rate constant. A more general pre-steady-state solution for reaction scheme 2 without any assumptions about the relative magnitudes of k_{op}, k_{cl}, and k_{int} was also solved.[9,18]

2.3. Pulsed-Amide H/D Exchange Method

Amide hydrogen exchange can be used to characterize the structure and measure the stability of folding intermediates in a pulsed-amide H/D exchange experiment as illustrated in Figure 2.[19,20] In a typical experiment, the protein is initially unfolded in D_2O in concentrated denaturant or at low pH. Amide NH exchange to ND. Folding is initiated by a rapid dilution into a folding buffer in H_2O at the pH at which the folding experiment is normally performed, for example, at pH 5.0−7.0 and 5−20 °C. Formation of early-folding intermediates under these conditions is commonly faster than the average exchange rate constants. After some folding time

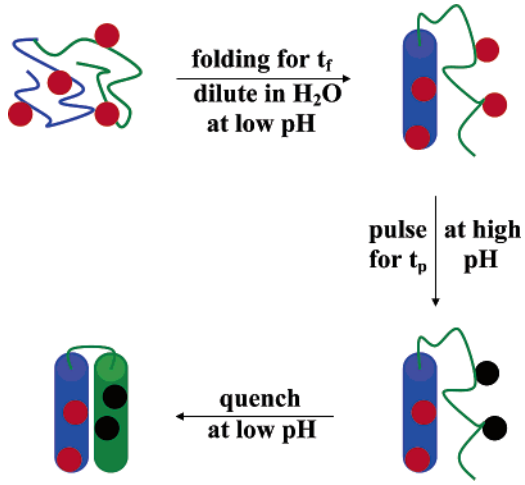


Figure 2. Illustration of the pulse-labeling procedure for detecting early-folding intermediates. Red and black balls represent D and H, respectively. t_f is the time for folding. t_p is the time for the high-pH pulse. The cylinders represent helical structures.

t_f, a brief H-labeling pulse, t_p, is applied by mixing with a high-pH buffer, for example, at pH 10 and 20 °C. The exchange time will be ~0.3 ms under such conditions. Amide deuterons that are not protected in the intermediates will exchange to NH, but those in already-formed structures are protected and remain as ND. In practice, multiple time points may be taken for either t_f or t_p. t_f may be in the range from milliseconds to several seconds. t_p may be in the range from 5 to 100 ms. A third mix into low pH terminates labeling. Within seconds the protein folds into its native state, which freezes the H−D labeling profile imposed before. The protein samples are concentrated. A 2D NMR spectrum or mass spectrometry can be taken.[19−21] In the NMR experiments, specific amide protons can be studied. Mass spectrometry, however, has the advantage of being able to identify the folding process involving parallel folding pathways.[21] A resolution at the level of peptide fragments can also be obtained by mass spectrometry when it is coupled with proteolysis.[22] Another advantage of mass spectrometry is that it can be used to study large proteins (>20 kDa).[23]

For a simple case in which the folding intermediate forms in sub-millisecond time scale and folds to the native state in the time scale of seconds, the fraction of the proton labeled in a pulse-labeling experiment is described in eq 6:

$$H_{label} = [1 - \exp(-PF \times k_{int} \times t_p)] \times [1 - \exp(k_f \times t_f)] \quad (6)$$

H_{label} can be measured at a series of different folding times from the start of folding (milliseconds) to its completion (seconds). The folding kinetics of the protein therefore can be monitored at multiple places of the structure. Because k_{int} is different for different amide protons at a given pH, the H_{label} is also different for different amide protons for a given pulse length t_p. The H_{label} alone is not always sufficient to determine which amide proton is protected in the intermediate. This problem is normally solved by changing the pH for the pulse while fixing t_f and t_p to measure the protection factor (PF) for each amide proton.[24,25] Figure 3 illustrates the plot from such an experiment.

A simplified version of the pulse-labeling method is the hydrogen exchange−competition method,[26] in which the high-pH pulse is omitted. Folding and exchange occur concurrently. Although it is limited to cases when the folding

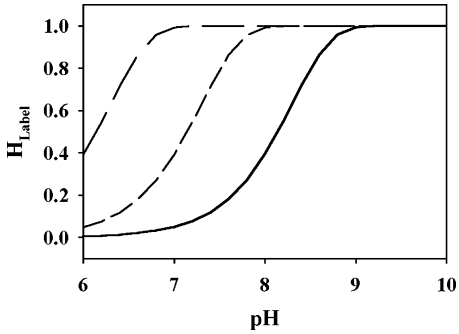


Figure 3. Illustration of the results for an early-folding intermediate populated in the sub-millisecond time scale from a pulse-labeling experiment. The parameters used are $t_f = 10$ ms, $t_p = 50$ ms, and $k_{int} = 100$ ms at pH 6.0. It is assumed that the conversion from the intermediate to the native state is much longer than t_p at the pH of pulse. The *X*-axis is the pH of pulse. The solid line represents an amide proton that is unfolded. The short- and long-dashed lines represent the amide protons that have protection factors of 10 and 100, respectively.

rate constant, k_f, is close to the intrinsic exchange rate constant, k_{int}, it is very useful to test whether a stable sub-millisecond intermediate exists because of its simplicity. The proton occupancy (H_{occ}) as a function of k_{int}, k_f, and t_f is described by eq 7.

$$H_{occ} = [PF \times k_{int}/(PF \times k_{int} + k_f)] \times [1 - \exp(-k_f \times t_f)] \quad (7)$$

2.4. Native-State Hydrogen Exchange Method

Hydrogen exchange pulse labeling is a kinetic method for characterizing folding intermediates. A native-state hydrogen exchange method was also developed to detect the equilibrium intermediates based on the exchange behavior of amide protons at low concentration of denaturants. Depending on the free-energy landscape of folding, amide protons can exchange through different processes. Figure 4 illustrates the

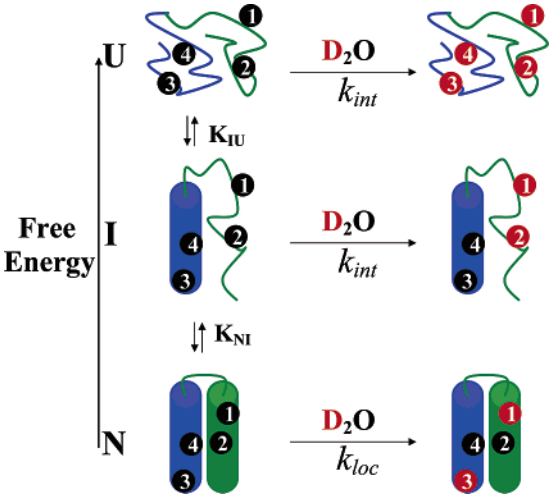


Figure 4. Illustration of three processes of hydrogen exchange for a protein with a partially unfolded intermediate: local fluctuation process (bottom); partial unfolding process (middle); and global unfolding (up).

three processes of exchange that occur for a three-state system. Amide protons that are not strongly protected or deeply buried in the native protein can exchange through local structural fluctuations by breaking one or two hydrogen

bonds without significantly exposing solvent-accessible surface area.[27−29] If the protein has a partially unfolded state that is more stable than the unfolded state, then the amide protons in the unfolded region of the intermediate can exchange in such partially unfolded states. All amide protons can also exchange from the fully unfolded states. The measured exchange rate constant is the sum of the exchange rate constants of all three processes, weighted by the unfolding equilibrium constants of the intermediate and the unfolded state:

$$k_{ex} = k_{loc} + K_{NI} \times k_{int} + K_{NU} \times k_{int} \quad (8)$$

That is, the exchange terms from the intermediate and the unfolded state are controlled by the unfolding equilibrium constants. Here, k_{loc} represents the exchange process from the native structure. K_{NI} and K_{NU} are the equilibrium unfolding constants. In a native-state hydrogen exchange experiment, the hydrogen exchange rates are measured at different concentrations of denaturant. The denaturant is used to perturb the equilibrium constants and help reveal the different exchange behaviors for amide protons in different regions of the protein. When ΔG_{HX} values of different amide protons are plotted against the concentration of denaturant concentrations, different structural segments that unfold cooperatively as structural units will converge together to different ΔG_{HX} values before joining the global ΔG_{NU}. Figure 5 illustrates the hydrogen exchange pattern for a protein with

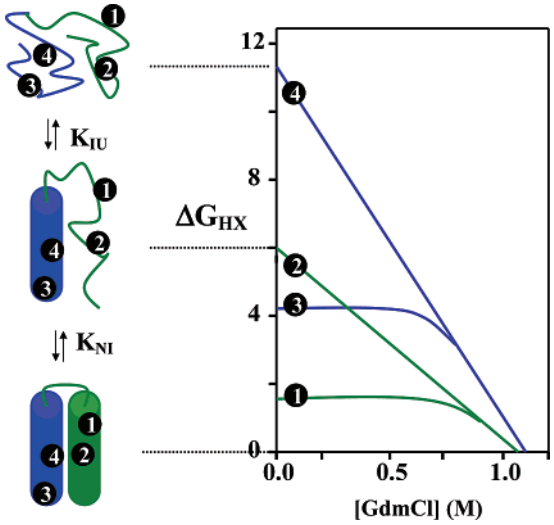


Figure 5. Illustration of native-state hydrogen exchange results: ΔG_{HX} values of amide proton versus [GdmCl] for a protein with one partially unfolded intermediate. Amide proton 2 dominantly exchanges in the partially unfolded intermediate. Amide proton 4 can exchange only through global unfolding. Because the above exchange processes involve the exposure of large solvent surface areas, the ΔG_{HX} values for these amide protons are sensitive to denaturant concentrations. The slopes of the ΔG_{HX} values versus denaturant concentration are proportional to the change of solvent-accessible surface area of the partial unfolding (proton 2) and global unfolding (proton 4). Amide protons 1 and 3 can exchange dominantly through local structure fluctuations at low concentration of denaturant without exposing significant solvent accessible surface area. Therefore, their ΔG_{HX} values are independent of denaturant concentration until the exchange process is taken over by the partial (for proton 1) and global (for proton 3) unfolding, as represented by the ΔG_{HX} values of protons 2 and 4. The intercepts of ΔG_{HX} values for protons 2 and 4 represent the unfolding free energies of the partially unfolded and the fully unfolded states at zero denaturant.[30]

a folding intermediate. On the basis of the exchange pattern, one can deduce the structure, stability, and exposed surface area of the intermediates from a native-state hydrogen exchange experiment.[30]

2.5. Experimental Results

Both the pulsed- and native-state hydrogen exchange experiments have been applied to a number of proteins, and some of the earlier experiments have been reexamined. They will be summarized here. For the proteins that both methods have been applied to, the relationship between the kinetic and equilibrium results will be discussed. It is very important to realize that the pulse-labeling and native-state hydrogen exchange methods detect partially unfolded intermediates in very different ways. In pulse-labeling experiments, only the intermediates that exist before the rate-limiting step and are more stable than the unfolded state can be detected. Intermediates that exist after the rate-limiting step are not observable (see Figure 6). In addition, the pulse-labeling

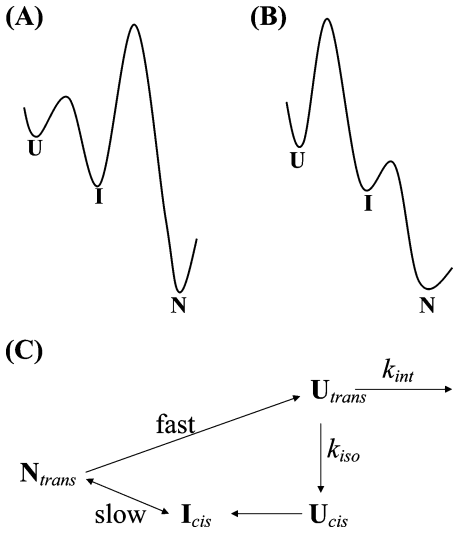


Figure 6. Illustration of kinetically detectable and undetectable folding intermediates. U, I, and N represent the unfolded state, the partially unfolded intermediate, and the native state, respectively. (A) The intermediate exists before the rate-limiting step. It accumulates during folding and is detectable in kinetic folding experiments. (B) The intermediate exists after the rate-limiting step and does not accumulate during kinetic folding. Therefore, it is not detectable in the pulse-labeling experiment. Both intermediates are detectable by the native-state hydrogen exchange under EX2 conditions (see detailed discussion in Bai and Englander[39]). (C) An intermediate involving proline isomerization may not be detectable by native-state hydrogen exchange when k_{int} is faster than k_{iso}. This is because before U_{trans} becomes U_{cis}, the amide protons in U_{trans} have already occurred. However, I_{cis} can be easily detectable in the conventional kinetic folding experiment where U_{cis} already exists in the unfolded state.

experiment requires that early-folding intermediates be more stable than the unfolded state at high pH during the pulse period. In other words, it is possible that early-folding intermediates may exist at low pH but not at high pH. In such cases, folding intermediates at low pH may not be detected in pulse-labeling experiments. For the native-state hydrogen exchange method, both early and late partially unfolded intermediates in principle can be detected as long as they are more stable than the unfolded state. However, the experiment needs to be done under EX2 conditions. In addition, the method does not detect the folding intermediates

accumulated before the rate-limiting step, due to slow cis−trans proline isomerization,[29] whereas such intermediates may be dominantly detected in the pulse-labeling experiments. Therefore, it is critical to know whether the kinetic folding intermediates identified in the pulse-labeling experiment are caused by proline isomerization. Another unappreciated problem is that transient intermolecular aggregations could cause amide proton protections in the pulse-labeling experiment. They can be mistakenly interpreted as a population of early-folding intermediates. Available examples are classified into five groups.

Group I. Partially unfolded intermediates with similar structures were observed in both pulsed- and native-state hydrogen exchange experiments. These examples include cytochrome *c* (cyt *c*) and ribonuclease H (RNase H).

Cyt c (104 Amino Acids). Cyt *c* structure consists of three helices, several ω-loops, and a heme that is covalently linked to the N-terminal helix. It is one of the two proteins that were first studied using a pulsed-amide H/D exchange and two-dimensional NMR.[19] The initial pulse-labeling experiment was done at pH 6 and 10 °C. A quick drop of 20% in H_{label} occurred for all amide protons that can be studied, suggesting that a fraction of the molecules fold to the native state through a fast-folding pathway. In a slow-folding pathway, amide protons in the N- and C-terminal helices are protected with a time constant of ∼20 ms, involving ∼40% of the molecules, suggesting population of an intermediate with the formation of the N- and C-terminal helices. Then there is a slower phase with a time constant of ∼3 s. The stability of the intermediate has been measured by changing the pH of the pulse.[25] Recent studies found that the ∼3 s phase is due to intermolecular aggregations.[31] The native-state hydrogen exchange was performed at pD 7 and 30 °C. Four equilibrium intermediates were identified.[30,32] One of the equilibrium intermediates has the N- and C-terminal helices folded. Population of the kinetic intermediate has been attributed to a misligation between heme and the side chain of a non-native His ligand.[33,34]

RNase H (155 Amino Acids). RNase H is one of the best-studied large proteins (>120 amino aids). The structure of RNase H consists of several α-helices (A−D) and four β-stands (1−4). Its folding pathway has been characterized by all available methods. The pulse-labeling experiments suggest that a partially unfolded intermediate forms in the sub-millisecond time scale before the rate-limiting transition state.[35] The intermediate involves the formation of helices A and D and strand 4. There are no probes for helix B. Other amide protons are protected in the time scale of ∼1 s. The native-state hydrogen exchange suggests that there are two partially unfolded intermediates.[36] The first intermediate also involves the formation of helices A and D. The second intermediate has the helices A, B, and D and strand 4 folded. Thus, both the first and second intermediates that were observed in the native-state hydrogen exchange experiment have formed in the kinetic intermediate detected in the pulse-labeling experiment. The structure of the kinetic intermediate also resembles those intermediates characterized under acidic conditions[37] and by a protein engineering study.[38]

Group II. Partially unfolded intermediates were identified in the native-state hydrogen exchange experiment, but they were not detected in the pulse-labeling experiments. This group includes cyt *c*, Rd-apocytochrome b_{562} (Rd-apocyt b_{562}), and barnase. It was suggested that the intermediates exist after the rate-limiting transition state.

Cyt c (104 Amino Acids). When the pulse-labeling experiment was performed for cyt *c* at pH 4.9, which avoids the misligation between a His residue and heme, no intermediates were detectable.[33,34] Because native-state hydrogen exchange has identified several partially unfolded states,[30,32] it was postulated that all of the intermediates exist after the rate-limiting transition step at pH 4.9.[39,40] The unfolding kinetics of these intermediates have been characterized by a kinetic native-state hydrogen exchange experiment and suggest that cyt *c* folds and unfolds sequentially through these intermediates.[41] A hydrogen exchange−competition experiment on cyt *c* in the presence of 0.4 M Na_2SO_4 showed some protection of amide protons (protection factor of 5−8) at the sub-millisecond time scale for essentially all amide protons that could be studied.[42] However, no protection was found in the absence of 0.4 M Na_2SO_4.

Rd-apocyt b_{562} (106 Amino Acids). Cytochrome b_{562} is a four-helix bundle protein with a heme ligand. A native-state hydrogen exchange experiment identified two partially unfolded intermediates in apocytochrome b_{562},[43] in which the non-covalently bound heme has been removed. The apo-protein was redesigned to include a Trp residue for kinetic folding studies.[44] The redesigned apo-protein, Rd-apocyt b_{562}, was also studied using the native-state hydrogen exchange experiment. It also identified two partially unfolded intermediates.[45] In the kinetic folding experiment, however, no early-folding intermediates were found,[44] again suggesting that the equilibrium intermediates exist after the rate-limiting step. One of the partially unfolded intermediates was populated using a protein engineering method.[46] Its structure was solved using multidimensional NMR (see later discussion).

Barnase (110 Amino Acids). The structure of barnase consists of four β-stands and two short α-helices. The initial pulse-labeling experiment of barnase showed complex folding kinetics.[47,48] Some amide protons were protected much earlier than others. The kinetic traces of H_{label} as a function of folding time for several amide protons were multiexponential. It was concluded that there is at least one early-folding intermediate that is ∼3 kcal/mol more stable than the fully unfolded state. Recent studies indicate that the complex folding behavior is due to intermolecular aggregation.[49,50] No stable early intermediates could be identified in the kinetic folding for barnase.[51,52]

The initial native-state hydrogen exchange experiment failed to identify any partially unfolded intermediate because it was done under near-EX1 conditions.[53] A recent native-state hydrogen exchange experiment has indicated that an intermediate with an ω-loop unfolded could exist. This was further confirmed using a protein engineering method.[54] Because the intermediate was not detected in the kinetic folding experiment, it is also postulated that the intermediate exists after the rate-limiting step. An early-folding intermediate with formation of the two short α-helices has been suggested on the basis of the earlier studies of the acid denatured state and computer simulation studies,[52] but there is no protection for the amide protons in them,[50] suggesting the proposed intermediate, if it exists at all, is less stable than the fully unfolded state.

Group III. No intermediates with native proline conformation were identified in either the pulsed- or native-state hydrogen exchange experiments. These proteins include ribonuclease A (RNase A), protein A B-domain, protein G B1-domain, CD2.D1, protein L B-domain, ubiquitin, and

chymotrypsin inhibitor 2 (CI2).

RNase A (124 Amino Acids). RNase A consists of several β-strands and an α-helix. It also includes four disulfide bonds. RNase A is the other protein that was first studied using the H/D exchange pulse-labeling method and 2D-NMR.[20] The folding kinetics of the initial pulse-labeling experiment was complex. Two intermediates were found in the pulse-labeling experiments.[20,24] These intermediates, however, are associated with proline isomerization. No kinetic folding intermediates were detectable in the absence of proline isomerization when a double-jump experiment coupled with pulse-labeling was performed[55] and analyzed properly.[56] The native-state hydrogen exchange of RNase A also showed the absence of detectable folding intermediates.[29,57,58] Because native-state hydrogen exchange can detect the folding pathway only in the absence of proline isomerization,[29] the pulse-labeling result is consistent with that of the native state hydrogen exchange; that is, no intermediate is detectable on the fast folding pathway in the absence of proline isomerization.

Ubiquitin (76 Amino Acids). Ubiquitin's structure consists of four β-strands packed with an α-helix. In the initial pulse-labeling experiment, the majority of the amide protons were protected against labeling in a major kinetic phase on the 10 ms time scale.[59] A region of irregular structure, including Tyr-59, Ile-61, and Leu-69, is protected later. There is also a minor slow protection phase for all amide protons, which is attributed to the cis−trans isomerization of prolines. An earlier stopped-flow study of kinetic folding suggested that there might be a stable sub-millisecond folding intermediate in the presence of 0.4 M Na_2SO_4.[60] More recent experimental study and a reanalysis of the earlier experimental results, however, have suggested that no early folding intermediates exist.[61−63]

CI2 (64 Amino Acids). The structure of CI2 has several β-strands packed with an α-helix. Kinetic folding studies indicate that CI2 folds in an apparently two-state manner.[64,65] Results from native-state hydrogen exchange studies on CI2[66] also led to the conclusion that no partially unfolded intermediates were detectable, but the pattern of ΔG_{HX} versus [GdmCl] appears to be somewhat complex. The reason could be, in part, that incorrect intrinsic exchange rates might have been used in calculating the ΔG_{HX} values.

Protein A B-Domain (58 Amino Acids). Protein A B-domain is a three-helix bundle protein. Kinetic folding experiments indicate that this protein folds in an apparent two-state manner.[67−69] No partially unfolded intermediates were identifiable by the native-state hydrogen exchange and protein engineering studies.[70]

Protein G B1 Domain (56 Amino Acids). The structure of protein G B1 domain has four β-strands packed with an α-helix. Earlier H/D exchange-folding competition[71] and stopped-flow[72] experiments suggested the existence of an early-folding intermediate. Later studies and a reanalysis of the earlier data have led to the conclusion that early-folding intermediates do not exist.[63]

Protein L (62 Amino Acids). Protein L has a structure similar to the protein G B1 domain. No intermediates were identified in both kinetic folding and native-state hydrogen exchange experiments.[73,74]

CD2.D1 (98 Residues). CD2.D1 is a β-sheet protein. The pulse-labeling experiment in the presence of 0.4 M Na_2SO_4 showed protection factors of <6 for 37 of the 41 NHs measured.[75] Native-state hydrogen exchange measurements found no stable intermediate.[76] In the absence of 0.4 M Na_2SO_4, the logarithm of the folding rate is a linear function of denaturant concentration, suggesting the absence of stable intermediates before the rate-limiting transition state.

Group IV. Only a pulse-labeling or native-state hydrogen exchange experiment has been performed. These proteins are hen egg white lysozyme, apomyoglobin, LysN, bovine β-lactoglobulin, interlukin-1β, OspA, triosephosphate isomerase, ribonuclease T1, thioredoxin, barstar, and an SH3 domain variant.

Hen Egg White Lysozyme (HEWL; 129 Amino Acids). HEWL has two domains. One domain contains α-helical structures (α-domain), and the other has the structure of β-strands (β-domain). Pulse-labeling experiments revealed parallel folding pathways as detected by both NMR and mass spectrometry.[21,77] A small fraction ($\sim$20%) folds quickly to the native state.[78] The other fraction ($\sim$80%) involves the formation of the α-domain as an early folding intermediate.

Apomyoglobin (AMb; 158 Amino Acids). AMb is an α-helical protein. It has six α-helices. The pulse-labeling experiment revealed a stepwise folding process. A, G, and H helices fold in a sub-millisecond time scale. Subsequent folding includes the formation of helix B and the rest of the protein and occurs on the time scale of seconds.[79] The structures of the two kinetic intermediates are similar to those populated at pH 4.1.[80,81] The equilibrium intermediate at pH 4.1 has been extensively characterized using multidimensional NMR.[82]

LysN (146 Amino Acids). The structure of LysN has an OB-fold motif composed of a structurally conserved five-stranded β-barrel capped by a poorly conserved α-helix between strands β3 and β4. Two additional α-helices flank the N terminus of the OB-fold. No kinetic folding studies have been performed on this protein. The results of native-state hydrogen exchange indicate that the N-terminal α-helices in LysN are able to unfold independently[83] and lead to an intermediate that is $\sim$3 kcal/mol more stable than the unfolded state. In contrast to the findings of cyt *c* and RNase H, however, the unfolding of the two α-helices appears to be less cooperative, as indicated by the broadness of ΔG_{HX} distribution at high denaturant concentrations.

Interleukin-1β (153 Amino Acids) (1L2H). This is a β-sheet protein with eight prolines. The pulse-labeling experiment on this protein has been performed and monitored by both NMR at pH 5 and 4 °C[84] and mass spectrometry at pH 5 and 25 °C.[85] In the experiment monitored by NMR, formation of two intermediates was suggested to occur on the time scale from 0.7 to 1.5 s and from 15 to 25 s, respectively. In the recent pulse-labeling experiment by mass spectrometry, only one intermediate was observed. It formed within $\sim$0.1 s. The conversion from the intermediate to the native state takes $\sim$40 s. No native-state hydrogen exchange experiment has been performed on this protein.

OspA (274 Amino Acids). OspA is a predominantly β-sheet protein containing a unique single-layer β-sheet. Native-state hydrogen exchange has identified five unfolding units,[86] leading to four partially unfolded intermediates. A kinetic folding experiment using stopped-flow has not been performed.

Triosephosphate Isomerase. This is a very large $(\beta/\alpha)_8$ barrel protein. The equilibrium intermediates have been studied using the misincorporation proton−alkyl exchange (MPAX).[87] The principle of this method is the same as the native-state hydrogen exchange except it monitors the

reaction behavior of the side chains of Cys residues. A misincorporated cysteine residue buried in the protein structure is protected from solvent and is unreactive in the native state. When exposed to solvent by local fluctuations, partial unfolding, or global unfolding, the cysteine sulfhydryl can react with an alkylating reagent to generate an irreversibly modified species. Thus, as in native-state hydrogen exchange, the partially unfolded intermediates could be identified by monitoring the exchange rates of sulfhydryl reaction as a function of denaturant concentration. Two partially unfolded forms have been identified.

Ribonuclease T_1 (RNase T_1; 104 Amino Acids). RNase T_1 is a homologue of barnase. It has two disulfide bonds. The folding kinetics in the pulse-labeling experiment was as complex as that of the initial pulse-labeling experiment on barnase.[88] The time course for deuterium protection is biphasic, with 60−80% of the protein molecules showing rapid hydrogen bond formation. The other 20−40% of the molecules are protected in a slow phase with a rate constant that has a lower limit of 0.01 s^{-1}. It is also known that RNase T1 populates an intermediate due to proline isomerization.[89]

Thioredoxin (TRX; 108 Amino Acids). TRX is a β-sheet protein. Pulse-labeling experiments have not been performed. No partially unfolded intermediates were identified in the native-state hydrogen exchange experiments.[90]

Barstar (89 Amino Acids). Barstar is an α/β protein. One proline (Pro48) has a cis conformation in the native state. The kinetic folding of barstar is heterogeneous, involving intermediates due to proline isomerization.[91] Native-state hydrogen exchange experiments suggest the existence of two partially unfolded intermediates.[92] In addition, two partially unfolded intermediates have been suggested to populate during kinetic unfolding of barstar by optical probes.[93]

SH3 Variant (65 Amino Acids). The folding behavior of the α-spectrin Src homology mutant with the substitution of N47−D48 by a sequence KITVNGKTYE that tends to form a stable β-hairpin was studied.[94] A kinetic folding experiment showed that the logarithm of the folding rate constant is curved. The native-state hydrogen exchange suggests that a partially unfolding intermediate with the substituted β-hairpin folded. No pulse-labeling experiment was performed. It is not clear whether this intermediate exists before or after the rate-limiting step.

Group V. Kinetic folding intermediates observed in the pulse-labeling experiment were not observed or were apparently different from those detected in the native-state hydrogen exchange experiment.

T4 Lysozyme (164 Amino Acids). T4 lysozyme has two visibly separated domains in the structure of the native state. The N-terminal domain has both α-helices and β-strands. The C-terminal domain is completely α-helical. This protein has three prolines. In the pulse-labeling experiment, intermediates were found to populate in the burst phase in ~10 ms.[95] The intermediate was converted to the native state with a time constant of ~170 ms at pH 6.0 and 23 °C in the presence of 1.5 M urea. The protection factors were measured after 32 ms of folding by varying the pH of a 41 ms pulse and were found to be in the range of 20−200. The very early folding in the burst phase involves formation of the E helix in the C-terminal domain and the β-sheet residues in the N domain. Other regions are formed in later steps. In the native-state hydrogen exchange experiment, a partially unfolded intermediate that involves only the unfolding of the N-terminal domain was identified.[96] These results have led to the suggestion that the partially unfolded intermediate seen in the native-state hydrogen exchange experiment is not on the kinetic folding pathway; rather, it is a kinetically isolated intermediate as in $U \leftrightarrow I_{kinetic} \leftrightarrow N \leftrightarrow I_{NHX}$. Here, $I_{kinetic}$ represents the kinetic folding intermediate observed in the pulse-labeling experiment and I_{NHX} is the intermediate detected by the native-state hydrogen exchange. However, protein engineering studies indicate that only the mutations in the C-terminal domain affect the folding rate,[97] suggesting that the formation of the C-terminal domain is the rate-limiting step. Further experiments are needed to clarify the contradictory results.

Staphylococcal Nuclease (SNase A; 149 Amino Acids). SNase A is an α/β protein. It has six proline residues. A pulse-labeling experiment was performed with the P117G mutant.[98] The amide proton occupancy of 39 amides was determined from 5 ms to 10 s, and three kinetic phases were identified. The protection of amide protons observed within the 5-ms dead time is consistent with the formation of a folding intermediate containing the β-sheet but not the α-helices. However, more intermediates need to be invoked to explain the slow-folding phases. Native-state hydrogen exchange was also performed for this protein. No partially unfolded intermediates were identified,[99] suggesting that the kinetic intermediates seen in the pulse-labeling experiments might all be related to proline isomerization like those seen in the folding RNase A.

Bovine β-Lactoglobulin (162 Amino Acids). Bovine β-Lactoglobulin is a predominantly β-sheet protein. The protein has eight prolines and two disulfide bonds. One of the disulfide bonds (106−119) holds the two central β-strands. At neutral pH, it forms a dimer. The native-state hydrogen exchange experiment at pH 2.1 and 37 °C showed that the protein essentially behaves as a single unit with some noncooperative fraying at the edge of the β-strands, which was attributed to the dissociation of dimers.[100] A pulse-labeling experiment was performed manually at pD 2.9 and 6 °C.[101] Folding intermediates were clearly observable 1 s after the initiation of folding. It has not been determined whether these intermediates are associated with proline isomerization. Again, the kinetic folding intermediates could be due to proline isomerizaton.

3. Hydrogen Exchange Results and the Mechanism of Protein Folding

3.1. Solving Large-Scale Conformational Search Problem by Discrete Folding Intermediates

One of the key questions in protein folding has been how protein molecules solve the large-scale conformational search problem. This question was raised at the end of the 1960s by Levinthal.[102−104] It was argued that protein molecules could not do a random search to find the native state in a biologically meaningful time scale because there are an astronomical number of conformations for a given polypeptide chain with even 100 amino acids. Thus, for proteins to fold efficiently, some defined folding pathways should exist. Accordingly, it was speculated that partially unfolded intermediates may form during folding and help to solve the large-scale conformational search problem.[102] For example, if each amino acid contributes three conformations, a polypeptide chain with 100 amino acids will have 3^{100} or ~10^{48} conformations. If the conversion from one conformation to another takes 10^{-12} s, a random search to find the

native structure will take 10^{36} s. However, if the molecule folds in four steps in a sequential manner, the total number of conformations to be searched is $4 \times 3^{25} \sim 10^{12}$. Thus, the native structure can be found by a stepwise folding in 1 s. In other words, folding by sequential intermediates is a very efficient way to reduce the search time.

This hypothesis led to the search for folding intermediates using kinetic methods with optical probes, such as fluorescence and circular dichroism. By the early 1980s, folding intermediates in several proteins had been found.[105] However, their structures could not be defined specifically, which left the possibility that there might be a large number of nonspecific partially unfolded intermediates on different pathways, which led to a jigsaw puzzle model.[106] In 1988, the pulse-labeling results on cyt *c*[19] and RNase A[20] provided the first examples for the existence of specific intermediates. Subsequent pulse-labeling studies on many other proteins such as lysozyme,[21,77,107] apomyoglobin,[79] and ribonuclease H[35] further supported the existence of specific intermediates. However, these are either large (>120 amino acids), which could have multiple domains, or involve misligations (cyt *c* intermediate) and proline isomerization (RNase A) during the process of folding. Recent studies on smaller single-domain proteins (<120 residues) have shown that they generally do not have detectable kinetic intermediates in the absence of stabilizing reagent (such as 0.4 M Na_2SO_4), disulfide bonds, proline isomerization, misligation, and aggregations. For example, hydrogen exchange experiments have shown that there are no significant amide proton protections for the proposed early-folding intermediates in the folding of ubiquitin, barnase, CD1, protein G B1-domain, and RNase A (see group II and III proteins).

The critical evidence for the existence of multiple folding intermediates in small single-domain proteins came from the native-state hydrogen exchange experiments. However, these intermediates exist after the rate-limiting transition state (see group II proteins) and therefore evade the detection by the conventional kinetic studies. It should be noted that the kinetically detectable intermediates and the hidden intermediates play the same role in the large-scale conformational search, although they may have different roles in determining the rate-limiting steps (see discussions in ref 54).

3.2. On- and Off-Pathway Nature of Folding Intermediate

Detection of partially unfolded intermediates by hydrogen exchange confirms that defined intermediates exist. To prove that intermediates solve the large-scale conformational search problem, it needs to be demonstrated that the intermediates are on the folding pathway, as in U ⇔ I ⇔ N. In other words, protein molecules have to go through them to reach the native state in a three-state system. It should not be off-pathway, as in I ⇔ U ⇔ N, or kinetically isolated, as in U ⇔ N ⇔ I.

A simple observation of an intermediate in a kinetic folding experiment alone does not necessarily mean that the intermediate is on the folding pathway. This is because when the unfolded state is less stable than the intermediate state, the unfolding process will not be detected in the folding from I ⇔ U ⇔ N. Therefore, specific demonstration is needed. The on- and off-pathway issue on folding intermediates has been studied for several proteins. These include the misligated intermediate of cyt *c*,[111] the misfolded intermediate of IM7,[109] and the intermediates with non-native proline conformation of RNase A.[110] These intermediates have been found to be on the folding pathway.[108−111] In the case of cyt *c* and RNase H, the structures of the kinetic intermediates identified in the pulse-labeling experiments are similar to those of intermediates characterized in the native-state hydrogen exchange experiments (group I). However, the structure of the kinetic folding intermediate of T4 lysozyme is different from that of the equilibrium intermediates identified by the native-state hydrogen exchange experiment. The kinetic intermediate has parts of the structures in both the N- and C-terminal domains folded, whereas only the C-terminal domain is folded in the equilibrium intermediate state. It is important to point out, however, that very high concentrations of protein (~7 mg/mL in the refolding process) were used in the pulse-labeling experiment of T4 lysozyme.[95] Because the kinetic folding intermediates of barnase observed in the earlier pulse-labeling experiment have been shown to be caused by transient intermolecular aggregations, the possibility that the observed kinetic intermediate in T4 lysozyme is also due to intermolecular aggregation cannot be simply ruled out.

3.3. High-Resolution Structure of a Hidden Folding Intermediate

Although folding intermediates have been identified and characterized for many proteins by pulsed- and native-state hydrogen exchange, the resolution of the structures of these intermediates is only at the level of residues. An important aspect of protein folding studies is to determine the high-resolution structures of folding intermediates at atomic resolution. A high-resolution structure provides the ultimate evidence for the existence of specific intermediates. This has been achieved recently. Using a native-state hydrogen exchange-directed protein engineering approach,[46] one of the hidden folding intermediates of Rd-apocyt b_{562} (see Figure 7) has been populated, and its structure has been

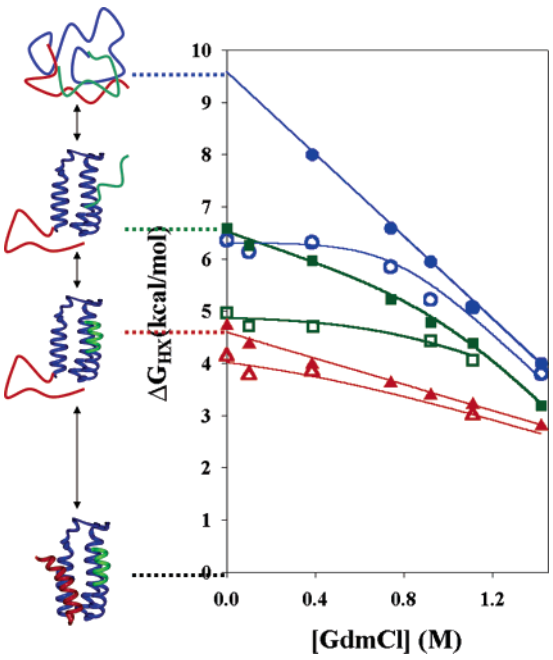


Figure 7. Native-state hydrogen exchange results of Rd-apocyt b_{562} along with the partially unfolded structures. Unfolding of the N-terminal helix (red) leads to the PUF2. Further unfolding of a part of the C-terminal helix leads to PUF1. Two amide protons are shown with corresponding colors for each region in the native structure.

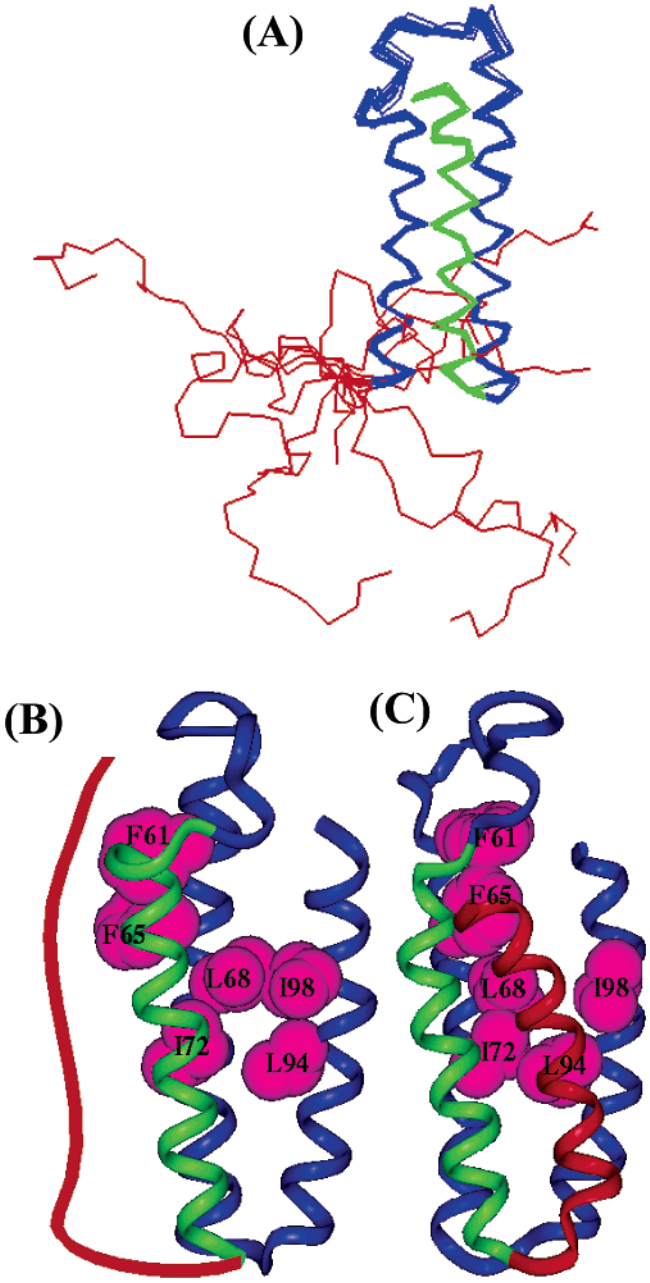


Figure 8. Structure of PUF2 determined by NMR: (A) C_α traces for 10 NMR structures. They are superimposed on the folded region of the structure; (B) illustration of non-native-like hydrophobic packing in the intermediate; (C) corresponding hydrophobic residues in the native state.

solved by multidimensional NMR methods (see Figure 8).[112]

The high-resolution structure of the intermediate has immediate implications for interpreting experimental results from amide hydrogen exchange and protein engineering studies. It is commonly assumed that protection of amide hydrogens in protein folding intermediates is due to native-like hydrogen bonding. Similarly, in protein engineering studies,[113,114] mutational effects on the stability of partially unfolded states (intermediates or transition states) are generally assumed to arise from native-like side-chain interactions. The native-like backbone conformation of the intermediates of Rd-apocyt b_{562} provides experimental support for the native-like hydrogen-bonding assumption. However, the broad non-native-like side-chain interactions observed in these intermediates indicate that the interpretation of mutation results in terms of native-like side-chain interactions may not be well founded.

This result also has important implications for understanding the mechanism of folding. The fact that Rd-apocyt b_{562} folds very rapidly with a rate constant of $\sim 10^4$ s^{-1} at 25 °C[45] indicates that side-chain packing or repacking is not intrinsically slow and cannot be the reason for population of early-folding intermediates, as previously suggested in the molten globule model.[115,116] This conclusion is further supported by recent mutational and theoretical studies in which non-native hydrophobic packing interactions have been found to actually stabilize the transition state and accelerate folding[117] and the number of folded residues and the topology of the transition states have been shown to be the dominant rate-limiting factors.[118]

3.4. Solving Large-Scale Conformational Search Problem by a Funnel-like Energy Landscape

The numerous folding pathway view proposed in the jigsaw puzzle model[106] has been further addressed on the basis of the statistical mechanics of polymers and computer simulation with simple lattice models.[119−121] These theoretical studies assumed that folding occurs on a funnel-like energy landscape through numerous different folding pathways. Accordingly, it is argued that mass action law may not be applicable.[121] Instead, Kramer's theory on diffusive reactions may be more appropriate for analyzing the kinetic folding.[122] In this funnel-like landscape theory, when a specific intermediate is populated in the folding process, it is considered as a kinetic trap that slows folding.[123] Moreover, the funnel-like energy landscape view also implies a general correlation between folding stability and folding rates. The stability of the transition state ensembles will be acquired largely by making use of interactions present in the native-state structure.[124]

The identification of specific early-folding intermediates in the pulse-labeling experiments in larger proteins and hidden intermediates by native-state hydrogen exchange in small proteins appears to be inconsistent with the general view of the funnel-like energy landscape. In particular, the high-resolution structure of the intermediate of Rd-apocyt b_{562} indicates that the intermediate can be very specific, providing strong evidence for the use of mass action law in analyzing folding kinetics. In addition, despite the broad non-native-like interactions in the intermediate state of Rd-apocyt b_{562}, the intermediate is not kinetically trapped because it exists after the rate-limiting step and Rd-apocyt b_{562} folds very rapidly with a k_f of $\sim 10^4$ s^{-1}. Moreover, destabilizing mutations in the N-terminal helix of Rd-apocyt b_{562} have no effect on the folding rates.[45] This is contrary to predictions of the funnel-like energy landscape theory but can be interpreted easily by the classical view because the N-terminal helix is not formed in the rate-limiting transition state and, therefore, makes no contribution to the stability of the rate-limiting transition state.[45]

3.5. Cooperativity Hypothesis of Protein Folding

A major cause for the inconsistency between the experimental results of naturally occurring proteins and the funnel-like energy landscape theory has recently been revealed by Chan and co-workers.[125] In a series of papers,[126−129] Chan and co-workers examined the past lattice models and found that all of them failed to reproduce the typical folding

behavior of apparently two-state proteins: a linear chevron plot (the linear dependence of the logarithm of the folding rate on denaturant concentrations). For example, in the most recently published lattice model,[130] the chevron plot is curved and the folding rate changed by a factor of only 2. In real proteins, however, it is normally linear and covers more than an order of magnitude of folding rate. More importantly, Chan and co-workers further showed that both many-body (cooperative) interactions and a near-Levinthal energy landscape (near-flat energy landscape) for the unfolded state, which have been lacking in the earlier lattice models, are necessary to reproduce a meaningful range of linear dependence of folding rate on denaturant concentrations. Specific partially unfolded intermediates similar to the hidden intermediates observed in the native-state hydrogen exchange experiments in real proteins have also emerged when the above two factors were incorporated empirically in the new lattice model.[131]

Because a funnel-like energy landscape can efficiently solve the large-scale conformational search problem, the key question becomes why small proteins do not appear to use it to solve the conformational search problem. One possible reason is that naturally occurring proteins may have evolved to have highly specific structures, which requires a rather flat energy landscape or even an uphill energy landscape for the unfolded conformations with cooperative stabilization of folding intermediates and native state.[132] A simple funnel-like energy landscape is not chosen by evolution because it lacks the folding cooperativity necessary for the well-defined structures.

Available data in the literature appear to be consistent with this cooperativity hypothesis. Many proteins perform very specific functions. These proteins need to have well-folded specific structures to ensure they can have the functional specificity. Therefore, a stringent evolutionary pressure is needed to achieve such results. This is consistent with the fact that it has been very difficult to de novo design proteins with the typical structural properties of naturally occurring proteins. The most successfully de novo designed or redesigned proteins are still dynamically more flexible than typical naturally occurring proteins. For example, the de novo designed three-helix bundle protein (α3D)[133] and the computer-redesigned protein G B1 domain[134] have more dynamic motions than naturally occurring proteins as judged by dynamics studies using NMR.[135,136] Another computer-redesigned protein G B1 domain[137] and a de novo designed protein with a new fold (Top7) by computational method[138] have much larger temperature factors (B-factor) than those in the naturally occurring proteins. These dynamic features might explain why the folding energy landscapes of computer-designed proteins tend to have a low folding free energy barrier.[139] Moreover, these studies further showed that naturally occurring proteins can be redesigned to be thermodynamically more stable and fold more quickly,[134,140] indicating that major evolutionary pressure is neither to optimize the thermodynamic stability nor to increase the folding rates of proteins.

3.6. Early-Folding Intermediates in Larger Proteins

Small proteins with fewer than 120 amino acids usually do not populate early-folding intermediates in the absence of the stabilizing reagent such as 0.4 M Na_2SO_4. It has been suggested that the rate-limiting process of the folding of small proteins is a search for a transition state with native-like topology.[40,141] The number of effectively folded residues in the transition state, which is controlled by the topology of the structure, is also found to be the dominant factor that determines the folding rate.[142] In contrast, the folding behavior of larger proteins has been less studied. The major questions are why larger proteins commonly populate early-folding intermediates during folding and what determines their folding rates. Currently, there are several hypotheses. The first is the molten globule model, which postulates that protein folding occurs in two steps: (1) formation of secondary structures on the time scale of milliseconds and (2) slow packing of tertiary interactions on the time scale of seconds.[116,143] The folding behavior of small proteins suggests that the molten-globule model is unlikely to be a general model for protein folding because these proteins fold in an apparent two-state manner without the population of early-folding intermediates. However, it has been argued recently that side-chain packing may still depend on the chain length of proteins. Tertiary packing may be easier to solve in small proteins than in large proteins.[144] The second hypothesis is that early-folding intermediates are populated due to the existence of misfolding barriers.[40] However, except in cases involving proline isomerization, Fe^{3+}−His misligation, and disulfide isomerization, the exact feature of misfolding is still not identified. Recent studies suggest that intermolecular aggregation may be considered as another factor for misfolding. This may be an important issue for the early pulse-labeling experiments because they were done at relatively high protein concentrations. Reexamination of the earlier results is needed to clarify this issue. In addition, an effort should be made to solve a high-resolution structure of an early-folding intermediate to see whether non-native topology exists because topology has been found to be important for determining the folding rate of small proteins. The third hypothesis is that larger proteins are made of multiple domains. Early-folding intermediates exist because different domains fold with different rates.[112,145] The domain that folds quickly will fold early and forms the folded region in the folding intermediate.

3.7. EX1 Hydrogen Exchange and Protein Unfolding

One important aspect of more recent protein folding studies by hydrogen exchange is the exploration of protein unfolding at high pH under native and EX1 conditions (see early introduction on EX1 mechanism) at which unfolding experiments cannot be performed directly. The rate constants of unfolding obtained from these studies have been very useful for providing information on protein unfolding. Different unfolding rates measured from several amide protons in the turkey ovomucoid third domain suggest the existence of a number of unfolding intermediates.[146,147] In the case of barnase, the unfolding rate constant measured by hydrogen exchange at low concentration of denaturant shows that the logarithm of the unfolding rate constant is not a linear function of denaturant concentration and helps to illustrate the absence of a previously identified stable sub-millisecond early-folding intermediate.[50,52] The nonlinear behavior also suggests that a hidden on-pathway folding intermediate might exist after the rate-limiting transition state. A native-state hydrogen exchange experiment under EX1 conditions on cyt *c* has provided important kinetic information for a sequential unfolding/folding pathway.[148] It is shown that partially unfolded intermediates can be identified on the basis of the

exchanging rates of amide protons measured at high pH. Four distinct unfolding steps were found to be in the order as postulated in the earlier studies. Exchange under EX1 conditions has also been used to study the dynamic conversion from one folding intermediate to another in a large protein with a single-layer β-sheet.[86] More recent work on CD1 revealed the switch between EX1 and EX2 at high pH, suggesting that a high-energy intermediate may have been observed.[149] A more complete coverage of the exploration of the EX1 mechanism has been presented in a recent review.[150]

4. Acknowledgment

I thank S. Walter Englander and Hue Sun Chan for sending their manuscripts before publication and for stimulating discussions and the reviewers for their helpful comments.

5. References

(1) Anfinsen, C. B. *Science* **1973**, *181*, 223.
(2) Canet, D.; Last, A. M.; Archer, D. B.; Redfield, C.; Robinson, C. V.; Dobson, C. M. *Nat. Struct. Biol.* **2002**, *9*, 308.
(3) Baldwin, R. L. *Curr. Opin. Struct. Biol.* **1993**, *3*, 84.
(4) Englander, S. W.; Mayne, L. *Annu. Rev. Biophys. Biomol. Struct.* **1992**, *21*, 243.
(5) Li, R.; Woodward, C. *Protein Sci.* **1999**, *8*, 1571.
(6) Dempsey, C. E. *Prog. Nucl. Magn. Reson. Spectrosc.* **2001**, *39*, 135.
(7) Rumbley, J.; Hoang, L.; Mayne, L.; Englander, S. W. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 105.
(8) Englander, S. W. *Annu. Rev. Biophys. Biomol. Struct.* **2000**, *29*, 213.
(9) Krishna, M. M. G.; Hoang, L.; Lin, Y.; Englander, S. W. *Methods* **2004**, *34*, 51.
(10) Wildes, D.; Marqusee, S. *Methods Enzymol.* **2004**, *380*, 328.
(11) Glasoe, P. F.; Long, F. A. *J. Phys. Chem.* **1960**, *64*, 188.
(12) Molday, R. S.; Englander, S. W.; Kallen, R. G. *Biochemistry* **1972**, *11*, 150.
(13) Bai, Y.; Milne, J. S.; Mayne, L.; Englander, S. W. *Proteins* **1993**, *17*, 75.
(14) Bai, Y.; Milne, J. S.; Mayne, L.; Englander, S. W. *Proteins* **1993**, *17*, 75.
(15) Connelly, G. P.; Bai, Y.; Jeng, M.-F.; Englander, S. W. *Proteins* **1993**, *17*, 87.
(16) Huyghues-Despointes, B. M.; Scholtz, J. M.; Pace, C. N. *Nat. Struct. Biol.* **1999**, *6*, 910.
(17) Hvidt, A.; Nielsen, S. O. *Adv. Protein Chem.* **1966**, *21*, 287.
(18) Qian, H.; Chan, S. I. *J. Mol. Biol.* **1999**, *286*, 607.
(19) Roder, H.; Elove, G. A.; Englander, S. W. *Nature* **1988**, *335*, 700.
(20) Udgaonkar, J. B.; Baldwin, R. L. *Nature* **1988**, *335*, 694.
(21) Miranker, A.; Robinson, C. V.; Radford, S. E.; Aplin, R. T.; Dobson, C. M. *Science* **1993**, *262*, 896.
(22) Zhang, Z.; Smith, D. L. *Protein Sci.* **1993**, *2*, 522.
(23) Pan, H.; Raza, A. S.; Smith, D. L. *J. Mol. Biol.* **2004**, *336*, 1251.
(24) Udgaonkar, J. B.; Baldwin, R. L. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 8197.
(25) Elove, G. A.; Roder, H. *ACS Symp. Ser.* **1991**, *No. 470*, 50.
(26) Schmid, F. X.; Baldwin, R. L. *J. Mol. Biol.* **1979**, *135*, 199.
(27) Mayo, S. L.; Baldwin, R. L. *Science* **1993**, *262*, 873.
(28) Qian, H.; Mayo, S. L.; Morton, A. *Biochemistry* **1994**, *33*, 8167.
(29) Bai, Y.; Milne, J. S.; Mayne, L.; Englander, S. W. *Proteins* **1994**, *20*, 4.
(30) Bai, Y.; Sosnick, T. R.; Mayne, L.; Englander, S. W. *Science* **1995**, *269*, 192.
(31) Nawrocki, J. P.; Chu, R. A.; Pannell, L. K.; Bai, Y. *J. Mol. Biol.* **1999**, *293*, 991.
(32) Krishna, M. M. G.; Lin, Y.; Rumbley, J. N.; Englander, S. W. *J. Mol. Biol.* **2003**, *331*, 29.
(33) Sosnick, T. R.; Mayne, L.; Hiller, R.; Englander, S. W. *Nat. Struct. Biol.* **1994**, *1*, 149.
(34) Elove, G. A.; Bhuyan, A. K.; Roder, H. *Biochemistry* **1994**, *33*, 6925.
(35) Raschke, T. M.; Marqusee, S. *Nat. Struct. Biol.* **1997**, *4*, 298.
(36) Chamberlain, A. K.; Handel, T. M.; Marqusee, S. *Nat. Struct. Biol.* **1996**, *3*, 782.
(37) Debora, J.; Marqusee, S. *Biochemistry* **1996**, *35*, 11951.
(38) Raschke, T. M.; Kho, J.; Marqusee, S. *Nat. Struct. Biol.* **1999**, *6*, 825.
(39) Bai, Y.; Englander, S. W. *Proteins* **1996**, *24*, 145.
(40) Sosnick, T. R.; Mayne, L.; Englander, S. W. *Proteins* **1996**, *24*, 413.
(41) Hoang, L.; Bedard, S.; Krishna, M. M. G.; Lin, Y.; Englander, S. W. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 12173.
(42) Sauder, J. M.; Roder, H. *Folding Des.* **1998**, *3*, 293.
(43) Fuentes, E. J.; Wand, A. J. *Biochemistry* **1998**, *37*, 9877.
(44) Chu, R. A.; Takei, J.; Knowles, J. R.; Andrykovitch, M.; Pei, W.; Kajava, A. V.; Steinbach, P.; Ji, X.; Bai, Y. *J. Mol. Biol.* **2002**, *323*, 253.
(45) Chu, R. A.; Pei, W. H.; Takei, J.; Bai, Y. *Biochemistry* **2002**, *41*, 7998.
(46) Takei, J.; Pei, W.; Vu, D.; Bai, Y. *Biochemistry* **2002**, *41*, 12308.
(47) Bycroft, M.; Matouschek, A.; Kellis, J. T., Jr.; Serrano, L.; Fersht, A. R. *Nature* **1990**, *346*, 488.
(48) Matouschek, A.; Serrano, L.; Meiering, E. M.; Bycroft, M.; Fersht, A. R. *J. Mol. Biol.* **1992**, *224*, 837.
(49) Chu, R.-A.; Takei, J.; Barchi, J. J., Jr.; Bai, Y. *Biochemistry* **1999**, *38*, 14119.
(50) Chu, R. A.; Bai, Y. *J. Mol. Biol.* **2002**, *315*, 761.
(51) Takei, J.; Chu, R. A.; Bai, Y. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 10796.
(52) Khan, F.; Chuang, J. I.; Gianni, S.; Fersht, A. R. *J. Mol. Biol.* **2003**, *333*, 169.
(53) Clarke, J.; Fersht, A. R. *Folding Des.* **1996**, *1*, 243.
(54) Vu, N. D.; Feng, H.; Bai, Y. *Biochemistry* **2004**, *43*, 3346.
(55) Houry, W. A.; Scheraga, H. A. *Biochemistry* **1996**, *35*, 11734.
(56) Qi, P. X.; Sosnick, T. R.; Englander, S. W. *Nat. Struct. Biol.* **1998**, *5*, 882.
(57) Mayo, S. L.; Baldwin, R. L. *Science* **1993**, *262*, 873.
(58) Qian, H.; Mayo, S. L.; Morton, A. *Biochemistry* **1994**, *33*, 8167.
(59) Briggs, M. S.; Roder, H. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 2017.
(60) Khorasanizadeh, S.; Peters, I. D.; Roder, H. *Nat. Struct. Biol.* **1996**, *3*, 193.
(61) Gladwin, S. T.; Evans, P. A. *Folding Des.* **1996**, *1*, 407.
(62) Krantz, B. A.; Sosnick, T. R. *Biochemistry* **2000**, *39*, 11696.
(63) Krantz, B. A.; Mayne, L.; Rumbley, J.; Englander, S. W.; Sosnick, T. R. *J. Mol. Biol.* **2002**, *324*, 1.
(64) Jackson, S. E.; Fersht, A. R. *Biochemistry* **1991**, *30*, 10428.
(65) Jackson, S. E.; Fersht, A. R. *Biochemistry* **1991**, *30*, 10436.
(66) Itzhaki, L. S.; Neira, J. L.; Fersht, A. R. *J. Mol. Biol.* **1997**, *270*, 89.
(67) Myers, J. K.; Oas, T. G. *Nat. Struct. Biol.* **2001**, *8*, 552.
(68) Dimitriadis, G.; Drysdale, A.; Myers, J. K.; Arora, P.; Radford, S. E.; Oas, T. G.; Smith, D. A. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 3809.
(69) Sato, S.; Fersht, A. R. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 6952.
(70) Bai, Y.; Karimi, A.; Dyson, H. J.; Wright, P. E. *Protein Sci.* **1997**, *6*, 1449.
(71) Kuszewski, J.; Clore, G. M.; Gronenborn, A. M. *Protein Sci.* **1994**, *3*, 1945.
(72) Park, S. H.; Shastry, M. C.; Roder, H. *Nat. Struct. Biol.* **1999**, *6*, 943.
(73) Yi, Q.; Baker, D. *Protein Sci.* **1996**, *5*, 1060.
(74) Yi, Q.; Scalley, M. L.; Simons, K. T.; Gladwin, S. T.; Baker, D. *Folding Des.* **1997**, *2*, 271.
(75) Parker, M. J.; Dempsey, C. E.; Lorch, M.; Clarke, A. R. *Biochemistry* **1997**, *36*, 13396.
(76) Parker, M. J.; Marqusee, S. *J. Mol. Biol.* **2001**, *305*, 593.
(77) Radford, S. E.; Dobson, C. M.; Evans, P. A. *Nature* **1992**, *358*, 302.
(78) Kiefhaber, T. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 9029.
(79) Jennings, P. A.; Wright, P. E. *Science* **1993**, *262*, 892.
(80) Hughson, F. M.; Wright, P. E.; Baldwin, R. L. *Science* **1990**, *249*, 1544.
(81) Loh, S. N.; Kay, M. S.; Baldwin, R. L. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 5446.
(82) Eliezer, D.; Yao, J.; Dyson, H. J.; Wright, P. E. *Nat. Struct. Biol.* **1998**, *5*, 148.
(83) Alexandrescu, A. T.; Jaravine, V. A.; Dames, S. A.; Lamour, F. P. *J. Mol. Biol.* **1999**, *289*, 1041.
(84) Varley, P.; Gronenborn, A. M.; Christensen, H.; Wingfield, P. T.; Pain, R. H.; Clore, G. M. *Science* **1993**, *260*, 1110.
(85) Heidary, D. K.; Gross, L. A.; Roy, M.; Jennings, P. A. *Nat. Struct. Biol.* **1997**, *4*, 725.
(86) Yan, S.; Kennedy, S. D.; Koide, S. *J. Mol. Biol.* **2002**, *323*, 363.
(87) Silverman, J. A.; Harbury, P. B. *J. Mol. Biol.* **2002**, *324*, 1031.
(88) Mullins, L. S.; Pace, C. N.; Raushel, F. M. *Biochemistry* **1993**, *32*, 6152.
(89) Balbach, J.; Steegborn, C.; Schindler, T.; Schmid, F. X. *J. Mol. Biol.* **1999**, *285*, 829.
(90) Bhutani, N.; Udgaonkar, J. B. *Protein Sci.* **2003**, *12*, 1719.
(91) Schreiber, G.; Fersht, A. R. *Biochemistry* **1993**, *32*, 11195.
(92) Bhuyan, A. K.; Udgaonkar, J. B. *Proteins* **1998**, *30*, 295.
(93) Zaidi, F. N.; Nath, U.; Udgaonkar, J. B. *Nat. Struct. Biol.* **1997**, *4*, 1016.

(94) Viguera, A. R.; Serrano, L. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 5730.
(95) Lu, J.; Dahlquist, F. W. *Biochemistry* **1992**, *31*, 4749.
(96) Llinas, M.; Gillespie, B.; Dahlquist, F. W.; Marqusee, S. *Nat. Struct. Biol.* **1999**, *6*, 1072.
(97) Gassner, N. C.; Baase, W. A.; Lindstrom, J. D.; Lu, J.; Dahlquist, F. W.; Matthews, B. W. *Biochemistry* **1999**, *38*, 14451.
(98) Jacobs, M. D.; Fox, R. O. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 449.
(99) Wooll, J. O.; Wrabl, J. Q.; Hilser, V. J. *J. Mol. Biol.* **2000**, *301*, 247.
(100) Ragona, L.; Fogolari, F.; Romagnoli, S.; Zetta, L.; Maubois, J. L.; Molinari, H. *J. Mol. Biol.* **1999**, *293*, 953.
(101) Forge, V.; Hoshino, M.; Kuwata, K.; Arai, M.; Kuwajima, K.; Batt, C. A.; Goto, Y. *J. Mol. Biol.* **2000**, *296*, 1039.
(102) Baldwin, R. L. *Protein Sci.* **1997**, *6*, 2031.
(103) Levinthal, C. In *Mossbauer Spectroscopy of Biological System*; Debrunner, P., Tsibris, J. C., Munck, E., Eds.; University of Illinois Press: Urbana, IL, 1969; pp 22−24.
(104) Levinthal, C. *J. Chim. Phys.* **1968**, *65*, 44.
(105) Kim, P. S.; Baldwin, R. L. *Annu. Rev. Biochem.* **1982**, *51*, 459.
(106) Harrison, S. C.; Durbin, R. *Proc. Natl. Acad. Sci. U.S.A.* **1985**, *82*, 4028.
(107) Miranker, A.; Radford, S. E.; Karplus, M.; Dobson, C. M. *Nature* **1991**, *349*, 633.
(108) Heidary, D. K.; O'Neill, J. C., Jr.; Roy, M.; Jennings, P. A. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 5866.
(109) Capaldi, A. P.; Kleanthous, C.; Radford, S. E. *Nat. Struct. Biol.* **2002**, *9*, 209.
(110) Laurents, D. V.; Bruix, M.; Jamin, M.; Baldwin, R. L. *J. Mol. Biol.* **1998**, *283*, 669.
(111) Bai, Y. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 477.
(112) Feng, H.; Takei, J.; Lipsitz, R.; Tjandra, N.; Bai, Y. *Biochemistry* **2003**, *42*, 12461.
(113) Matthews, C. R.; Hurle, M. R. *Bioessays* **1987**, *6*, 254.
(114) Fersht, A. R.; Matouschek, A.; Serrano, L. *J. Mol. Biol.* **1992**, *224*, 771.
(115) Shakhnovich, E. I.; Finkelstein, A. V. *Biopolymers* **1989**, *28*, 1667.
(116) Ptitsyn, O. B. *Trends Biochem. Sci.* **1995**, *20*, 376.
(117) Di Nardo, A. A.; Korzhnev, D. M.; Stogios, P. J.; Zarrine-Afsav, A.; Kay, L. E.; Davidson, A. R. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 7954.
(118) Clementi, C.; Plotkin, S. S. *Protein Sci.* **2004**, *13*, 1750.
(119) Brygelson, J. D.; Onuchic, J. N.; Socci, N. D.; Wolynes, P. G. *Proteins* **1995**, *21*, 167.
(120) Sali, A.; Shakhnovich, E.; Karplus, M. *Nature* **1994**, *369*, 248.
(121) Dill, K. A.; Chan, H. S. *Nat. Struct. Biol.* **1997**, *4*, 10.
(122) Onuchic, J. N.; Socci, N. D.; Luthey-Schulten, Z.; Wolynes, P. G. *Folding Des.* **1996**, *1*, 441.
(123) Wolynes, P. G.; Onuchic, J. N.; Thirumalai, D. *Science* **1995**, *267*, 1619.
(124) Onuchic, J.; Wolynes, P. G. *Curr. Opin. Struct. Biol.* **2004**, *14*, 70.
(125) Chan, H. S.; Shimizu, S.; Kaya, H. *Methods Enzymol.* **2004**, *380*, 350.
(126) Kaya, H.; Chan, H. S. *Proteins* **2000**, *40*, 637.
(127) Kaya, H.; Chan, H. S. *J. Mol. Biol.* **2002**, *315*, 899.
(128) Kaya, H.; Chan, H. S. *Protein Sci.* **2003**, *52*, 524.
(129) Kaya, H.; Chan, H. S. *Proteins* **2003**, *52*, 510.
(130) Schonbrun, J.; Dill, K. A. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 12678.
(131) Kaya, H.; Chan, H. S. *Proteins* **2005**, *58*, 31.
(132) Kuhlman, B.; Baker, D. *Curr. Opin. Struct. Biol.* **2004**, *14*, 89.
(133) Walsh, S. T.; Cheng, H.; Bryson, J. W.; Roder, H.; DeGrado, W. F. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 5486.
(134) Malakauskas, S. M.; Mayo, S. L. *Nat. Struct. Biol.* **1998**, *5*, 470.
(135) Walsh, S. T.; Lee, A. L.; DeGrado, W. F.; Wand, A. T. *Biochemistry* **2001**, *40*, 9560.
(136) Bai, Y.; Feng, H. *Eur. J. Biochem.* **2004**, *271*, 1609.
(137) Nauli, S.; Kuhlman, B.; Trong, I. L.; Stenkamp, R. E.; Teller, D.; Baker, D. *Protein Sci.* **2002**, *11*, 2924.
(138) Kuhlman, B.; Dantas, G.; Ireton, G.; Varani, G.; Stoddard, B.; Baker, D. *Science* **2003**, *302*, 1364.
(139) Scalley-Kim, M.; Baker, D. *J. Mol. Biol.* **2004**, *338*, 573.
(140) Gillespie, B.; Vu, D. M.; Shah, P. S.; Marshall, S. A.; Dyer, R. B.; Mayo, S. L.; Plaxco, K. W. *J. Mol. Biol.* **2003**, *330*, 813−819.
(141) Makarov, D.; Plaxco, K. W. *Protein Sci.* **2003**, *12*, 17.
(142) Bai, Y.; Zhou, H. Y.; Zhou, Y. Q. *Protein Sci.* **2004**, *13*, 1173.
(143) Kuwajima, K.; Mitani, M.; Sugai, S. *J. Mol. Biol.* **1989**, *206*, 547.
(144) Kamagata, K.; Arai, M.; Kuwajima, K. *J. Mol. Biol.* **2004**, *339*, 951.
(145) Inaba, K.; Kobayashi, N.; Fersht, A. R. *J. Mol. Biol.* **2000**, *302*, 219.
(146) Arrington, C. B.; Teesch, L. M.; Robertson, A. D. *J. Mol. Biol.* **1999**, *285*, 1265.
(147) Arrington, C. B.; Robertson, A. D. *J. Mol. Biol.* **2000**, *300*, 221.
(148) Hoang, L.; Bedard, S.; Krishna, M. M.; Lin, Y.; Englander, S. W. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 12173.
(149) Cliff, M. J.; Higgins, L. D.; Sessions, R. B.; Clarke, A. R. *J. Mol. Biol.* **2004**, *336*, 497.
(150) Ferraro, D. M.; Lazo, N. D.; Robertson, A. D. *Biochemistry* **2004**, *43*, 587.

CR040432I