Quarterly Reviews of Biophysics, Page 1 of 40. © 2008 Cambridge University Press
doi:10.1017/S0033583508004654 Printed in the United Kingdom
Vol 40(4), 287-326 (2008)

Protein folding and misfolding: mechanism and principles

S. Walter Englander[1]*, Leland Mayne[1] and Mallela M. G. Krishna[1,2]

[1] The Johnson Research Foundation, Department of Biochemistry and Biophysics, University of Pennsylvania, Philadelphia, USA
[2] Department of Pharmaceutical Sciences and Biomolecular Structure Program, University of Colorado Health Sciences Center, Denver, CO, USA

Abstract. Two fundamentally different views of how proteins fold are now being debated. Do proteins fold through multiple unpredictable routes directed only by the energetically downhill nature of the folding landscape or do they fold through specific intermediates in a defined pathway that systematically puts predetermined pieces of the target native protein into place? It has now become possible to determine the structure of protein folding intermediates, evaluate their equilibrium and kinetic parameters, and establish their pathway relationships. Results obtained for many proteins have serendipitously revealed a new dimension of protein structure. Cooperative structural units of the native protein, called foldons, unfold and refold repeatedly even under native conditions. Much evidence obtained by hydrogen exchange and other methods now indicates that cooperative foldon units and not individual amino acids account for the unit steps in protein folding pathways. The formation of foldons and their ordered pathway assembly systematically puts native-like foldon building blocks into place, guided by a sequential stabilization mechanism in which prior native-like structure templates the formation of incoming foldons with complementary structure. Thus the same propensities and interactions that specify the final native state, encoded in the amino-acid sequence of every protein, determine the pathway for getting there. Experimental observations that have been interpreted differently, in terms of multiple independent pathways, appear to be due to chance misfolding errors that cause different population fractions to block at different pathway points, populate different pathway intermediates, and fold at different rates. This paper summarizes the experimental basis for these three determining principles and their consequences. Cooperative native-like *foldon units* and the *sequential stabilization* process together generate predetermined stepwise pathways. O*ptional misfolding* errors are responsible for 3-state and heterogeneous kinetic folding.

* Author for correspondence: Dr S. W. Englander, Department of Biochemistry and Biophysics, University of Pennsylvania, Philadelphia, PA 19104-6059, USA.
Tel.: 215-898-4509; Fax: 215-898-2415; Email: engl@mail.med.upenn.edu

1. The protein folding problem

The search for protein folding pathways and the principles that guide them has proven to be one of the most difficult problems in all of structural biology. Biochemical pathways have almost universally been solved by isolating the pathway intermediates and determining their structures. This approach fails for protein folding pathways. Folding intermediates only live for <1 s and they cannot be isolated and studied by the usual structural methods.

The protein folding problem has been attacked by the entire range of fast spectroscopic methods which are able to track folding in real time. Much valuable information has been obtained but mainly of a kinetic nature. Kinetic methods do not describe the structure of the pathway intermediates that they access. In parallel, major efforts have been mounted to decipher the pathways and principles of protein folding by theoretical calculations. The problem of predicting the structure of native proteins from amino-acid sequence has not been solved (Moult *et al.* 2007). Computing a sequence of intermediate structures on the pathway for getting there is an even more daunting challenge.

It has become possible to determine by direct experiment the structures and the thermodynamic and kinetic properties of ephemeral protein folding intermediates in other ways, especially by hydrogen exchange (HX) methods. HX methods depend on the fact that main chain amide group hydrogens exchange naturally with the hydrogens in solvent water, providing nonperturbing structure-sensitive probes at every amino acid (except proline) in every protein molecule. Folding intermediates that become significantly populated during kinetic folding only live for milliseconds but they can be characterized by HX methods in favorable cases. Having reached the native state, proteins repeatedly unfold, revisit their higher free-energy forms, and refold again. Under native conditions the high free-energy pathway intermediates are infinitesimally populated and normally invisible but they can in favorable cases be studied in some detail by HX methods.

The most favorable model protein found so far is equine Cytochrome *c* (Cyt *c*). The most complete information has been obtained for it but similar results have been obtained for many

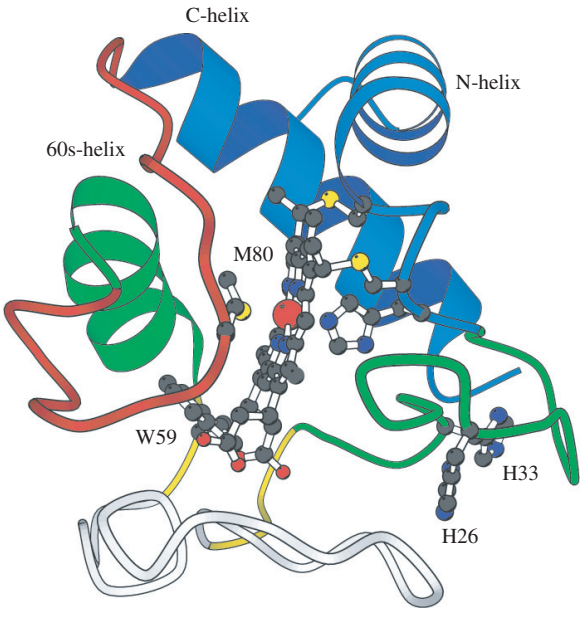


Fig. 1. Ribbon diagram of cytochrome *c*. Cyt *c* has 104 residues, three major α-helices, three major Ω-loops, and six cooperative foldon units shown in color, namely the N/C bi-helix (blue), the 60s helix and loop (green) which can be separated at low pH, the small β-sheet (yellow), and two Ω-loops (red and infrared). The two peripheral histidines that can misligate to the heme iron and block folding in a pH-sensitive way are in the green loop. Also shown are the Met80 ligand and Trp59 which serves as a FRET donor in kinetic folding experiments.

other proteins. The results provide a coherent picture of the structure of folding intermediates and their pathway relationships. Three major principles emerge. Proteins act like accretions of cooperative unfolding/refolding units called foldons which account for the unit steps in folding pathways. The foldon substructure of Cyt *c* is shown in Fig. 1. During folding the formation and stepwise addition of each foldon unit is guided by pre-existing structure in a sequential stabilization process. The native-like conformation of foldon units and their sequential stabilization by native-like interactions together produce predetermined stepwise pathways that progressively assemble the native protein.

However, for some proteins different population fractions accumulate different intermediates and fold with different kinetics, which has been widely interpreted in terms of different pathways. Information now available indicates that heterogeneous folding behavior is due to chance misfolding errors at various points in a determinate pathway and not to different tracks through the folding energy landscape.

Our thesis is that the integration of these three concepts – foldon units, sequential stabilization, and optional errors – now provides a coherent and well supported explanation for the vast literature on the macroscopic folding behavior of protein molecules.

2. A little history

The modern history of the protein folding problem began almost 50 years ago with the demonstration by Anfinsen and co-workers that ribonuclease A (RNase A) can fold with no help

from other biological machinery (Anfinsen *et al.* 1961; Anfinsen, 1973). Anfinsen showed that, as for any chemical reaction, the folding of RNase A proceeds spontaneously downhill to the lowest free-energy polypeptide conformation, the predestined functional native state. The generalization of this demonstration is that however complex the folding process might be, all of the information needed for getting there is wholly encoded in each protein's own amino-acid sequence. More recent work has uncovered networks of helper molecules that edit and secure the folding process, but Anfinsen's original thermodynamic hypothesis remains unchallenged. Biology's helper machinery functions either to delay final folding or to avoid or overcome folding errors, but not to actively *direct* folding. It is interesting that the first intimation of that complex support machinery was found by Anfinsen and co-workers (De Lorenzo *et al.* 1966).

Soon thereafter Cyrus Levinthal pointed out that the native state is unlikely to be found in any reasonable time by a random search through conformational space, the so-called Levinthal paradox (Levinthal, 1969; Dill, 1985). Levinthal's point was that some distinct predetermined pathway must be in play, and also that the native state might therefore not be the lowest free-energy state. It now appears that Levinthal was close to being correct in both conjectures. As described here, it appears that proteins do fold by way of predetermined intermediates in predetermined pathways. Moreover, the native functional state of some proteins is not necesssarily their lowest free-energy form. Many proteins can spontaneously adopt a still lower energy conformation, albeit as a macromolecular aggregate, namely the amyloid form.

Levinthal's work stimulated the successful search for indications of specific folding intermediates, pursued most notably by the schools of Baldwin (Kim & Baldwin, 1982a, 1990) and Creighton (Creighton, 1986). Creighton and co-workers set out to isolate folding intermediates that could be trapped by and defined in terms of naturally occurring covalent disulfide bond formation. Baldwin and co-workers pioneered a range of methodologies, especially spectroscopic methods, to observe folding in real time. They were also the first to test the use of HX to tag transiently populated intermediates that might be analyzed later (Schmid & Baldwin, 1979). These and many other efforts modeled on them did find evidence for the presence of distinct intermediates.

Another wing of the folding field has attempted to bypass the experimental difficulties altogether by the application of theoretical and computer simulation methods. Also this effort was foreshadowed by Levinthal who initially attempted to compute protein folding pathways (Fine *et al.* 1991). Zwanzig later showed, contrary to the Levinthal conjecture, that the search through conformational space might still be random if it is energetically biased to be downhill (Zwanzig *et al.* 1992). No specific pathway would then be necessary. Continuing theoretical investigations have led to the picture of multiple energetically downhill paths through a funnel-shaped folding energy landscape (Leopold *et al.* 1992; Bryngelson *et al.* 1995; Wolynes *et al.* 1995; Plotkin & Onuchic, 2002a, b). The landscape has been explored in many theoretical and experimental efforts, which have generally endorsed the multi-pathway view.

The investigations recounted here have been able to explore by direct experiment the relevant intermediate structures in the free-energy landscape between the native and the unfolded state and to decode their pathway relationships. This article organizes information that has unveiled and characterized cooperative foldon units in a number of protein molecules and explored the role of this new dimension of protein structure in protein folding pathways. A quantity of information affirms that the folding energy landscape is dominated by a small number of discrete, partially formed but distinctly native-like structures that are constructed, one after another,

by the sequential assembly of cooperative foldon units of the target native protein. This folding mechanism can be seen as wholly consistent with the bulk of experimental and theoretical evidence that has been interpreted differently before.

3. Hydrogen exchange

HX studies focus most usefully on the main chain amide hydrogen (—C(O)—NH—) which is placed at every amino acid except for proline in every protein molecule and participates in the ubiquitous hydrogen-bonding interactions that mark protein structural elements. Polar side-chain hydrogens also exchange but are much less informative because they are not usually involved in H-bonded structure.

3.1 HX measurement

A sensitive but difficult density-dependent H–D exchange measurement, developed by Linderstrøm-Lang and co-workers (Linderstrøm-Lang & Schellman, 1959; Hvidt & Nielsen, 1966) in their groundbreaking HX studies, was soon displaced by H–T exchange measured by gel filtration methods (Englander, 1963), which has in turn given way to extensively used 2D NMR (Wagner & Wüthrich, 1982; Wand & Englander, 1996), making it easily possible to obtain residue-resolved HX rate data. More recently a proteolytic fragmentation method (Englander *et al.* 1985) together with analysis by mass spectrometry (Zhang & Smith, 1993; Eyles & Kaltashov, 2004) makes it possible to obtain segment-resolved HX information for proteins that are too large for NMR and even to move toward amino-acid resolution in favorable cases (Englander *et al.* 2003; Englander, 2006).

3.2 HX chemistry

The factors that affect the rate of exchange between amide and solvent hydrogens at the primary chemical level are pH, temperature, inductive and steric effects of nearest-neighbor side-chains, and kinetic, equilibrium, and solvent isotope effects. These various factors have been accurately calibrated in small structureless amide models (Molday *et al.* 1972; Bai *et al.* 1993; Connelly *et al.* 1993). From these calibrations, the second-order HX rate constant that these factors produce under any given conditions, known as the intrinsic rate constant (k_{int}), can be computed, and the expected chemical HX rate (k_{ch}) can be obtained as k_{int} multiplied by the catalyst concentration [$k_{ch} = k_{int}$ (cat)]. The calculation is conveniently accessible at HX2.Med.UPenn.Edu/download.html and elsewhere. Because amides have higher basicity than hydroxide ion, effective catalysts are limited to hydroxide ion itself in the normal pH range, and hydroxide to amide collision must occur hundreds of times on average before a successful proton transfer event (Englander & Kallenbach, 1983). Specific acid catalysis by hydronium ion becomes important at pH 3 and below.

The HX rates of amide hydrogens are remarkably well placed for protein studies. At pH 7 and 0 °C, an average HX rate for unprotected amides is about 1 s^{-1}. In structured proteins rates can be greatly slowed and are spread over many orders of magnitude. The rate increases by a factor of 10 per pH unit and per ~22 °C in temperature, which makes possible the experimental adjustment of HX rates over a range eight orders of magnitude wide between pH 4, 0 °C and pH 10, 40 °C. HX lifetimes between milliseconds and months can be conveniently

measured. Accordingly, the entire range of information-rich protein HX behavior can be moved into the laboratory time window.

3.3 HX analysis

Protecting protein structure, almost always involving an H-bonding interaction, makes HX slower than the baseline chemical rate. Over the years many workers have considered modes by which HX catalyst might reach and remove exchangeable hydrogens. It now appears, as inferred by Linderstrøm-Lang, that hydrogens protected by H-bonded structure can exchange only when their protecting H-bond is transiently broken and they are brought into contact with external solvent, which requires some significant structural distortion or 'opening' reaction, as in Eq. (1). At steady state, the continuing HX rate is then given by Eq. (2), which can be reduced as shown when structure is stable ($k_{cl} \gg k_{op}$).

$$\text{Closed} \underset{k_{cl}}{\overset{k_{op}}{\longleftrightarrow}} \text{open} \xrightarrow{k_{ch}} \text{exchanged}, \tag{1}$$

$$k_{ex} = k_{op}\, k_{ch}/(k_{op} + k_{cl} + k_{ch}) \approx k_{op}k_{ch}/(k_{cl} + k_{ch}), \tag{2}$$

When a hydrogen bond transiently separates and the hydrogen is made accessible to attack by exchange catalyst, a kinetic competition ensues between k_{cl} and k_{ch}. If reclosing is faster than exchange ($k_{cl} > k_{ch}$), then opening and reclosing will occur repeatedly before a successful HX event. The measured exchange rate will then be given by k_{ch} multiplied down by the fraction of time that the hydrogen is accessible, essentially the pre-equilibrium opening constant, K_{op}, as in Eq. (3). This is known as the EX2 regime (bimolecular exchange). In this case, measured k_{ex} together with the predictable value of k_{int} [$k_{ch} = k_{int}$ (cat)] leads to the equilibrium constant for the opening reaction and its free energy. Alternatively, if $k_{ch} > k_{cl}$, for example at increased catalyst concentration (high pH) or decreased structural stability (e.g. mild denaturant), then exchange will occur upon each opening. Measured HX rate then rises to an upper limit equal to the structural opening rate [Eq. (4)], the so-called EX1 limit (monomolecular exchange).

$$k_{ex} = k_{op}k_{ch}/k_{cl} = K_{op}k_{ch}; \quad \Delta G_{ex} = -RT\ \ln\ K_{op} = -RT \ln (k_{ex}/k_{ch}), \tag{3}$$

$$k_{ex} = k_{op}. \tag{4}$$

These equations were first given by Kai U. Linderstrøm-Lang (1958) (Hvidt & Nielsen, 1966; Englander & Kallenbach, 1983). The more complex non-steady state solutions required by some HX experiments were given by Hvidt and Schellman (Hvidt, 1964; Krishna *et al.* 2004a).

3.4 HX structural physics

The exchange of structurally protected hydrogens proceeds through 'open' HX competent states that exist only a fraction of the time [Eq. (1)]. The structural information available depends on the kind of opening that dominates exchange. We distinguish cooperative segmental unfolding reactions and more local distortional fluctuations. Cooperative unfolding reactions have a recognizable signature. They cause multiple neighboring hydrogens to exchange with very similar ΔG_{ex} [EX2 exchange, Eq. (3)] and k_{op} [EX1 exchange, Eq. (4)], and they have a sizeable dependence on destabilants that promote cooperative unfoldings (denaturant, temperature, pH, pressure). In contrast local structural fluctuations are marked by disparate HX rates for neighboring residues and near-zero sensitivity to destabilants.

3.4.1 Global unfolding

That the slowest protein hydrogens might exchange by way of global unfolding was first suggested by Rosenberg and co-workers (Rosenberg & Chakravarti, 1968), was pursued most actively by Woodward and co-workers (Woodward & Hilton, 1979; Woodward *et al.* 1982; Woodward, 1994) based mainly on studies of temperature dependence or mutationally induced changes, and was convincingly demonstrated (Loh *et al.* 1993; Bai *et al.* 1994) when the availability of the calibrations just noted made it possible to calculate absolute ΔG_{ex} values. A survey of the HX literature found many proteins for which the slowest exchanging hydrogens when processed through Eq. (3) yield ΔG_{ex} values that closely match ΔG_{unf} values obtained by standard protein melting experiments (Huyghues-Despointes *et al.* 2001).

3.4.2 Local fluctuations

Many protein hydrogens exchange by way of local fluctuations. Their study can provide amino-acid-resolved information on protein dynamical flexibility and motions. Accordingly it is impressive how little is known about them. Local fluctuational motions that render amides exchange competent extend over a very small number of residues (Maity *et al.* 2003), probably conditioned by the type of secondary structure involved. Exchange rate depends on the density of local interactions (Bahar *et al.* 1998; Vendrusculo *et al.* 2003) and secondarily on the depth of burial (Milne *et al.* 1998). Values of ΔH_{ex} [$\Delta H_{ex} = -\partial R \ln K_{op}/\partial(1/T)$] (Milne *et al.* 1999; Hernandez *et al.* 2000), m [$m = \partial \Delta G_{ex}/\partial$(denaturant)] (Bai *et al.* 1995a) and ΔV_{ex} [$\Delta V = \partial \Delta G_{ex}/\partial$(pressure)] (Fuentes & Wand, 1998b) are close to zero. EX2 exchange is always seen because reclosing is so fast, measured as being faster than microseconds (Hernandez *et al.* 2000). Exchange may occur from a partially blocked state, slower than expected from the usual calibrated values for k_{int}, making the calculated ΔG_{HX} [Eq. (3)] misleadingly high (Maity *et al.* 2003).

For present purposes local fluctuational motions, although exceptionally interesting in respect to protein dynamics, have a negative impact. They often dominate measured exchange and hide the behavior of the larger unfolding reactions that we wish to access. This makes it necessary to develop tactics that artificially promote the larger unfolding pathways in order to make them visible.

3.4.3 Subglobal unfolding

Specially designed methods termed equilibrium and kinetic native state HX allow the direct experimental study of the high free-energy states accessed by transient unfolding reactions, their identification, and the characterization of their thermodynamic and kinetic properties according to Eqs (1)–(4). The application of these methods has led to the concept that proteins are composed of cooperative foldon units that engage in repeated subglobal unfolding/refolding reactions, even under native conditions. It now appears that protein molecules exploit this dynamic dimension of protein structure for many functional purposes.

This article focuses on subglobal unfolding reactions that turn out to be relevant to protein folding intermediates and mechanism. However, it is important to note that not all unfolding reactions necessarily fall into this category. Small functionally important protein unfolding reactions, earlier referred to as local unfolding, were first seen in mechanistic studies of allosteric function in hemoglobin (Englander *et al.* 1998a, 2003). It seems questionable whether these lower

level unfolding reactions are pertinent to protein folding intermediates. It is necessary in any given case to establish the relevance of partially unfolded states for protein folding mechanism by experiment rather than by assumption.

3.5 Summary

Exchangable amide hydrogens provide universal probes for protein structure and for some of its dynamic and thermodynamic properties at amino-acid resolution. The chemistry of the HX process, the measurement of protein HX, and its analysis in terms of structural dynamics are well understood. In structured proteins, opening reactions that determine HX behavior range from local fluctuations that break as little as one protecting H-bond at a time, through larger partial unfolding reactions, up through whole molecule global unfolding. The measurement of HX behavior and its analysis in these terms can provide profound insight into protein structure, dynamics, design, and function as well as the protein folding process.

4. Foldons

Much experimental work on protein folding has focussed on the pathway intermediates that carry initially unfolded proteins to their native state. The entire range of available spectroscopic methods has been exploited to follow folding in real time, verify the presence or absence of populated intermediates, and attempt to understand their pathway relationships. However, these methods do not provide the structural detail necessary to understand the intermediates that are detected. Detailed structural information has come mostly from HX methods. These studies have led to the serendipitous discovery of cooperative foldon units, which turn out to play a key role in protein folding.

4.1 Foldons in kinetic folding

4.1.1 The HX pulse-labeling experiment

Folding intermediates that transiently accumulate during kinetic folding typically live for <1 s and are therefore difficult to study in any structural detail. It first became possible to determine their structure by use of a HX pulse-labeling method (Roder *et al.* 1988; Udgaonkar & Baldwin, 1988), following earlier considerations due to Kim & Baldwin (1982b, 1990). The idea is to label a transiently populated intermediate in a structure-sensitive way by H/D exchange during its brief sub-second lifetime and then analyze the labeling pattern later in the refolded native protein.

An unfolded protein fully deuterium-exchanged in D_2O is induced to begin folding, usually by rapid dilution from high denaturant concentration into H_2O under slow HX conditions in a stopped-flow or continuous-flow apparatus. If the refolding protein encounters a large on-pathway barrier, a folding intermediate transiently accumulates. The intermediate is exposed to a short intense burst of HX labeling by mixing briefly into high pH. For example, at pH 9 and 20 °C the HX lifetime of a freely exposed amide is about 1 ms. The brief pulse selectively labels amides of the intermediate according to their degree of protection (EX2 region) or their rate of transient unfolding (EX1 region). The protein then completes folding to the native state and is placed into slow HX conditions which freezes the H/D labeling profile imposed during the

subsecond pulse. Later sample preparation and 2D NMR analysis to read out the H/D profile may take hours, but the results can identify the structure of the subsecond intermediate, its stability, and its time-course for formation and loss, all at amino-acid resolution (Krishna *et al.* 2003a, 2004a).

4.1.2 Structure of a kinetic folding intermediate

The most detailed HX pulse-labeling results have been obtained for the particularly favorable case of a Cyt *c* intermediate that can be made to accumulate to a very high level (85% of the population) (Krishna *et al.* 2003a). Figure 2*a* shows results obtained when the intensity of the labeling pulse was varied over a 300-fold range (50 ms pulse between pH 7·5 and pH 10). This range accesses both the EX2 regime at lower pH which measures structural stability [Eq. (3); labeling rises with pH], and the EX1 regime at higher pH which provides rate information [Eq. (4); labeling becomes independent of pH], although a more complex analysis is required in this non-steady-state situation (Krishna *et al.* 2003a, 2004a). The results specify the structure of the trapped intermediate (Fig. 2*a*) and its biophysical properties (Fig. 2*b*).

Major protection is found for all of the amides that are H-bonded in the N and C helices of the native protein (blue in Figs 1, 2*a*), indicating that the entire length of both segments is protected. Stopped-flow circular dichroism shows the formation of an equivalent amount of helix confirming that the two terminal segments exist as helices in the intermediate. Little other H-bonded structure is found. Some core hydrophobic residues of the 60s helix (Met65, Leu68) have small protection ($K_{op} \sim 1$) suggesting burial but not helix formation. Some minimal protection is seen for His33 and Thr78 which form main chain to side-chain bends in the native protein and retain some protection even in the unfolded protein.

The site-resolved stability for both helices is maximal at the position where they dock in the native protein (Fig. 2*b*), indicating their native-like configuration (see also Colon *et al.* 1996). They exhibit close to the same stability ($\sim 1{\cdot}4$ kcal mol^{-1} at 0·23 m GdmCl, 10 °C) and the same unfolding (~ 10 s^{-1}) and refolding (~ 100 s^{-1}) rates, confirming their cooperative single foldon nature. As might be expected, there is a fraying-like gradient in rates and stability toward the helix termini (Fig. 2*b*). The distribution of stability mirrors but is much greater than the value expected for isolated helices (AGADIR-predicted stability $+3$ to 4 kcal mol^{-1}) consistent with mutual stabilization.

The closing rate calculated from the HX data is the same as the rate measured by stopped-flow fluorescence for 2-state Cyt *c* folding (in absence of the later blocking barrier), indicating the on-pathway nature of the bihelical intermediate. All of the molecular population folds through this same homogeneously structured form (85% trapped at the time of the pulse with the other 15% having already moved through and reached N).

In short, the kinetically populated folding intermediate resembles in detail a partially constructed native protein. The unstructured regions of the partially folded intermediate undoubtedly represent a broad dynamic ensemble as emphasized by theoretical studies but the same extensive native-like elements are importantly present in all of the molecules. The HX pulse-labeling experiment has now been repeated for transiently blocked intermediates in many proteins. A consistent observation is that structured regions resemble structure within the native protein. The same is true for partially structured proteins more generally, for example for equilibrium molten globules (Baum *et al.* 1989; Jeng *et al.* 1990).

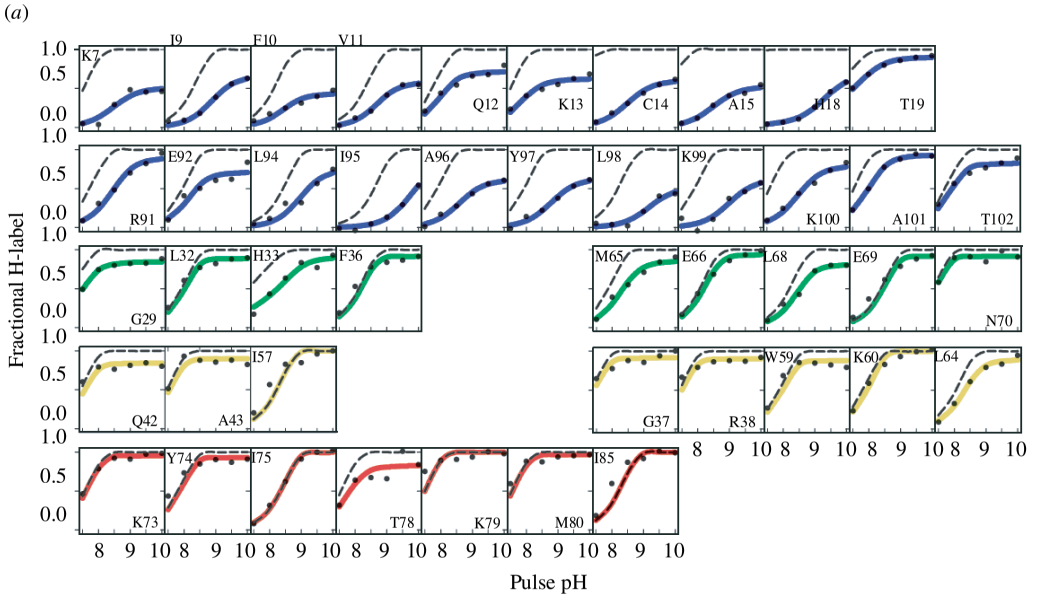


Fig. 2. HX pulse-labeling results for the blocked initial intermediate in Cyt *c* kinetic folding. (*a*) The degree of labeling imposed at each measurable amide during the pulse (colored curves), analyzed by 2D NMR of the refolded native protein, can be compared with the black dashed curve expected for the case of no protection, scaled to 0·85, the fraction of the population in the intermediate state at the time of the HX pulse. The horizontal offset in the EX2 region relates to the stability against labeling. The plateau at higher pH (EX1 region) indicates the limiting k_{op} rate. (*b*) Parameters of the trapped N/C bi-helical intermediate.

4.2 Foldons at equilibrium

4.2.1 The native state HX experiment

The HX pulse-labeling experiment is limited to the study of well populated kinetic intermediates. Most folding intermediates occur after the intrinsic initial rate-limiting step and are invisible to all kinetically based observations. A more powerful method known as native state hydrogen exchange (NHX) can make it possible to study several, perhaps all, of the intermediates that together construct a folding pathway, whether they accumulate to a high level in kinetic folding or not.

The NHX experiment depends on a fundamental concept that has received little attention before (Englander & Kallenbach, 1983). It is true, as in Anfinsen's thermodynamic hypothesis, that proteins are driven to their final native conformation because that is their lowest free-energy state. However, thermodynamic principle also requires that native proteins must continue to populate all of their higher free-energy forms, each according to its Boltzmann factor, and over time the population must cycle through all of the high-energy manifold. Thus protein molecules naturally unfold and refold even under native conditions, repeatedly revisiting the same intermediate forms that carried them down to their native state initially (Bai & Englander, 1996; Englander *et al*. 1997).

Unfortunately, this behavior is normally invisible. The high-energy forms are only infinitesimally populated and the usual methodologies are swamped by signals from the overwhelmingly populated native state. The opposite is true for HX. Measured HX rates receive no contribution from hydrogens protected in the native state but are wholly determined by the transiently populated higher energy forms. Therefore the continual unfolding and refolding behavior of proteins, and parts of proteins, is in principle accessible to HX measurement under native conditions.

The equilibrium version of the NHX experiment can be understood by analogy to the common protein melting experiment. Figure 3*a* illustrates the standard 2-state denaturant melting curve and its analysis by the usual linear extrapolation method (Pace, 1975). The calculated dependence of $\Delta G^{\circ}_{\mathrm{unf}}$ on denaturant is characterized by a large m value ($m = \partial(\Delta G^{\circ})/(\partial[\mathrm{Den}])$, which proportions to the size of the unfolding structure (Myers *et al*. 1995). The usual extrapolation implies that proteins unfold reversibly, albeit invisibly, even at zero denaturant. The value of $\Delta G^{\circ}_{\mathrm{unf}}$ at low denaturant, found by long extrapolation in Fig. 3*a*, can be measured directly by the NHX experiment. This is true because the opening reaction in Eq. (1) that governs the slowest exchanging amide hydrogens in many proteins is often the transient global unfolding reaction itself (Huyghues-Despointes *et al*. 2001).

Figure 3*b* shows the HX behavior of all of the amide hydrogens in the N- and C-terminal helices of Cyt *c*, plotted as ΔG_{HX} [from Eq. (3)] against denaturant concentration far below the melting transition ($C_{\mathrm{mid}} = 2{\cdot}75$ M GdmCl). At low denaturant, many well protected hydrogens normally exchange by way of local fluctuations. These opening reactions are very small as shown by their near-zero m value and the fact that immediately neighboring residues can exchange with very different rates. As denaturant is increased, the large global unfolding, with large m value, is sharply promoted, just as in Fig. 3*a*, and soon comes to be the dominant exposure pathway. One then sees a merging of the separate local fluctuational HX curves for different amino acids into a common HX isotherm. The HX isotherm in Fig. 3*b* is generated by the reversible global unfolding of Cyt *c* under native conditions and can reveal its parameters [e.g. Eq. (3)]. Also the residue-resolved HX data identify the amino-acid amides that

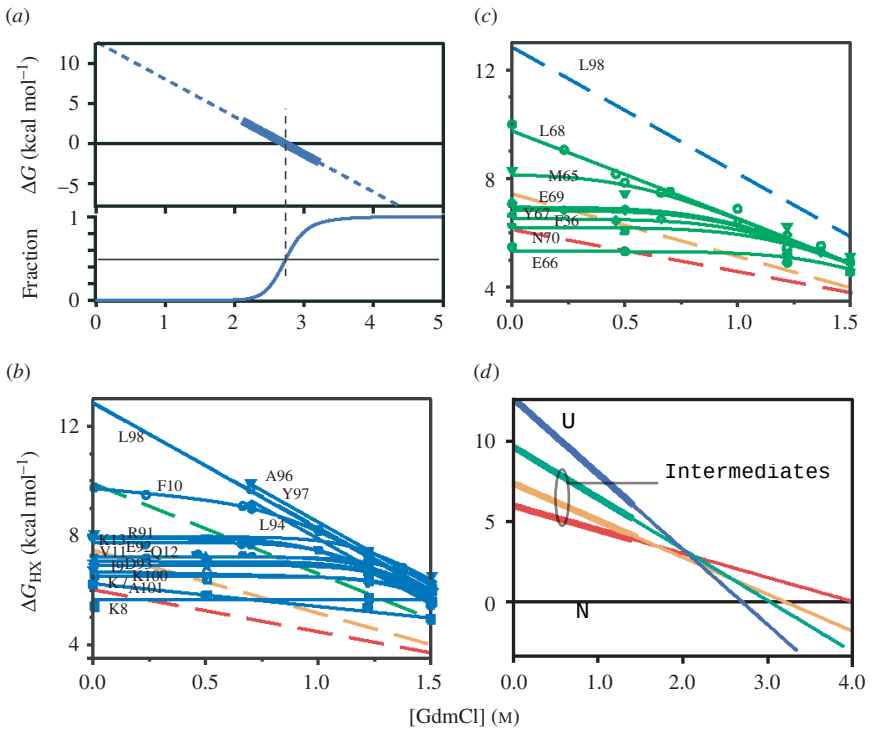


Fig. 3. Native state hydrogen exchange. (*a*) Illustration of the 2-state denaturant melting curve and its analysis by the usual linear extrapolation method (Pace, 1975), showing the sharp dependence of global stability on denaturant concentration even far below the visible melting transition where NHX is measured. (*b*, *c*) The analogous measurement by NHX of transient unfolding reactions of the blue and green Cyt *c* foldons at low denaturant concentration. These large unfolding reactions, initially hidden by local fluctuational HX pathways ($m \sim 0$), are promoted and revealed by their sharp dependence on denaturant concentration. Residues exposed to HX in each unfolding reaction join the HX isotherm and thus identify the unfolding segments. Each HX isotherm, summarized in panel (*d*), shows the denaturant-dependent free-energy level (relative to N) for the Cyt *c* partially unfolded form that first exposes each foldon unit to exchange (color coding connects to Fig. 1). Line thickening shows where HX was actually measured. The overall foldon composition of each partially unfolded form was measured in stability labeling experiments (Section 5, Fig. 5). The infrared foldon was not measured at these conditions (pDr 7, 30 °C).

are exposed in this last unfolding step. They are all of the residues in the N- and C-terminal helices.

In summary, the NHX experiment is able to directly measure the global unfolding reaction as it occurs under fully native conditions. The same N/C foldon unit found by HX pulse labeling to fold first is found by NHX to unfold last. A similar strategy can reveal structural units that engage in lower free energy more probable cooperative unfolding reactions.

4.2.2 Foldons found by equilibrium NHX

As denaturant is raised further and the global unfolding is sharply promoted, it will ultimately come to dominate the exchange of all of the protein's hydrogens. In favorable cases, it may first become possible to observe lower free-energy unfolding reactions that involve only some part of the native structure. As an example, Fig. 3*c* shows NHX data for all of the measurable hydrogens

in the green helix and loop segments of Cyt *c*. At low denaturant almost all of the hydrogens exchange by way of local fluctuations, with $m = 0$. Exchange of the Leu68 amide NH depends on a larger unfolding reaction. Its unfolding free energy and *m* value are large but less than for the global unfolding, indicating that it is exposed to exchange by a subglobal unfolding. As GdmCl concentration is increased, the large but still subglobal unfolding is selectively promoted and comes to dominate the exchange of all of the hydrogens that it exposes. This is seen as a merging of various local fluctuational HX curves into a single HX isotherm. The amino-acid residues that join the HX isotherm show that this particular unfolding reaction includes the entire green helix and green loop. It is useful to notice that the Leu68 amide hydrogen can serve as a marker for the unfolding of the green foldon because it can only exchange when the green foldon opens. In the same way the amide protons of residues 95–98 act as markers for the global unfolding.

Similar NHX results have identified and characterized cooperative foldon units that account for the entire Cyt *c* molecule. Figure 3*d* shows their HX isotherms. Figure 1 illustrates these foldon units in the Cyt *c* model. They correspond closely to recognizable secondary structural elements in the native protein. When ranked by a color code in order of increasing unfolding free energy (at native conditions) and increasing *m* value, the foldons are called infrared (I), red (R), yellow (Y), green (G) and blue (B). Recent work has shown that the green helix and green loop can be made to unfold separately at low pH where the green loop is selectively destabilized (Krishna *et al.* 2007). Thus the 104 residue Cyt *c* protein is made up of six distinct foldon units.

These same foldon units have been repeatedly confirmed in multiple experiments under different conditions by HX pulse labeling (Krishna *et al.* 2003a, 2004b), by equilibrium NHX (Bai *et al.* 1995b; Bai & Englander, 1996; Englander *et al.* 1997; Krishna *et al.* 2003b, 2007; Maity *et al.* 2005), and by a related kinetic NHX method that distinguishes concerted foldon units by virtue of their different unfolding rates rather than their different stabilities and *m* values (Hoang *et al.* 2002; Krishna *et al.* 2004a).

4.3 Limitations in foldon detection

Cooperative subglobal unfolding reactions might have been found as long ago as the original HX studies pursued by Linderstrøm-Lang and co-workers in the 1950s but they escaped detection for 40 years. A major problem has been that the exchange of many hydrogens is often dominated by local fluctuational openings (Fig. 3*b*, *c*) which hide the possible participation of these residues in larger unfoldings. The NHX strategy selectively promotes the larger unfoldings so that they come to dominate the measured exchange of the hydrogens that they expose and can then be detected, identified, and characterized.

Amides that continue to exchange preferentially by way of local fluctuations, often toward the ends of secondary structural elements, cannot be assigned to their correct foldon. The same is true for residue NMR cross-peaks that escape detection. The general tendency will be to incorrectly limit foldon extent to the most protected, most easily measured residues. A rigorous definition of the foldon construction of any protein would require a complete accounting of all of its residues. In most cases this is unlikely to be accomplished. Accordingly it seems likely that foldons entrain more complete secondary structural segments than is definitively observed. It is hard to picture how a major mid-region of well structured helices or β-strands could concertedly unfold without also entraining their more distal residues in the same unfolding.

Another limit occurs at the other extreme. If a protein is not sufficiently stable, some subglobal units may independently unfold only at a free-energy level that is higher than the U state.

In this case HX of the hydrogens that could reveal foldon behavior will be dominated instead by the global unfolding reaction. The protein then appears to be more 2-state cooperative than is the case. The problem is exacerbated by the fact that subglobal unfoldings tend to occur at surprisingly high ΔG_{HX} values so that high global stability, perhaps 9 kcal mol^{-1} or more, is desirable for a successful analysis (although see Korzhnev *et al.* 2007).

Even when the presence of large unfolding reactions can be demonstrated, the separation of different foldons may not be obvious because foldon hydrogens tend to exhibit a spread of ΔG_{HX} values, on the order of 1 kcal mol^{-1}. If the spread between foldons is not greater than the spread within, their discrimination will be difficult. The spread can be due to fraying mechanisms as seen definitively for the Cyt *c* N/C intermediate by HX pulse labeling (Fig. 2*b*). Further, in incompletely unfolded forms HX may be partially blocked and slower than has been calibrated for freely exposed small molecule models (Maity *et al.* 2003), contributing an artifactual spread to the calculated ΔG_{HX} values.

Finally, the presence of non-native interactions in unfolded forms can confuse foldon identification. In any incompletely native form, some native interactions are absent and energy-minimizing non-native interactions may be present. This has been seen directly (see Feng *et al.* 2005a).

Attempts have been made to create structure-based algorithms that might find protein foldons by computation (Panchenko *et al.* 1996; Fischer & Marqusee, 2000; Weinkam *et al.* 2005; Hilser *et al.* 2006). These important efforts need to be pursued.

4.4 Foldons in many proteins

In spite of the severe limitations just noted, concerted subglobal unfolding reactions have been detected in many proteins. Native state HX experiments have found foldon units in RNase H (Chamberlain *et al.* 1996), barnase (Vu *et al.* 2004), T4 lysozyme (Cellitti *et al.* 2007; Kato *et al.* 2007), PDZ-3 domain (Feng *et al.* 2005b), focal adhesion target domain (Zhou *et al.* 2006), OspA (Yan *et al.* 2002) and Cyt b_{562} (Fuentes & Wand, 1998a, b; Chu *et al.* 2002; Takei *et al.* 2002). HX pulse-labeling studies of kinetic folding intermediates in many proteins have similarly found specific native-like partially folded forms.

Foldons have been shown in the triosephosphate isomerase dimer by a SH reactivity method that is analogous to the NHX method (Silverman & Harbury, 2002a). Kay and co-workers have found unfolding units in three small proteins (Fyn SH3, Abp1p, and FF domains) by an NMR relaxation dispersion method (Korzhnev *et al.* 2004, 2006, 2007). The manipulation of ambient conditions and protein engineering have been able to produce molten globules (Ptitsyn, 1995; Raschke & Marqusee, 1997) and autonomous folding units (Oas & Kim, 1988; Dabora *et al.* 1996; Feng *et al.* 2005a) where again one finds that native-like elements are preserved (Baum *et al.* 1989; Jeng *et al.* 1990). Wolynes and co-workers developed a theoretical analysis that found foldon units in Cyt *c* very similar to those found by HX methods (Pletneva *et al.* 2005; Weinkam *et al.* 2005).

4.5 Summary

Structural elements that engage in subglobal cooperative unfolding and refolding reactions have now been demonstrated in many proteins by many methods. It appears that proteins may generally be considered to be composed of foldon units.

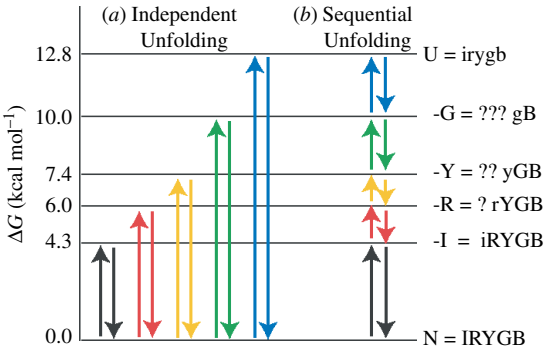


Fig. 4. The two extreme possibilities for foldon unfolding behavior. (*a*) Independent foldon unfolding. (*b*) Sequential foldon unfolding. Information on the foldon composition of the different partially unfolded forms revealed by the NHX results and the foldon status not yet specified by the NHX results (shown as question marks) is listed at the right.

5. Foldons to partially unfolded forms (PUFs)

In any given PUF, one or more foldon units remain folded and others are unfolded. It is crucial to understand the composition of the PUFs in these terms. When foldons are represented by their one-letter color code, native Cyt *c* can be notated as IRYGB and the fully unfolded state as irygb. The pulse-labeling result in Fig. 2 defines a PUF populated during kinetic folding with composition irygB. What is the condition of the various foldons in other partially unfolded states? Figure 4 illustrates two extreme alternatives. The individual foldon units may unfold independently, as in Fig. 4*a*, or more interestingly they may unfold in a pathway manner, as in Fig. 4*b*. Other combinations might also exist.

The time-dependent NMR spectra used to measure H–D exchange in an NHX experiment make it clear which amides exchange in each foldon unfolding reaction (e.g. Fig. 3*b*, *c*), and also some that are still protected. The still protonated amides are seen in the NMR analysis at longer times. For example, in the NMR spectra where one observes the time-course of green foldon exchange (indicated by the Leu68 marker proton in Fig. 3*b*) one can see that the blue foldon marker protons remain unexchanged. Therefore, when the green foldon first unfolds (G to g), the blue foldon is still in the folded condition (B). In the case of independent foldon unfolding (Fig. 4*a*), the foldon composition of this PUF would be IRYgB. In the sequential unfolding case (Fig. 4*b*), it would be irygB. Here we are concerned with exchange by way of cooperative foldon unfolding reactions which can be discerned (over and above exchange through local fluctuations) from the converged HX isotherm or, more conveniently, from available marker protons.

The NHX results do not specify the condition of the amino acids that have already exchanged (H to D) in the lower-lying more populated open states. Accordingly, when the green foldon unfolds, the condition of the lower-lying infrared, red, and yellow foldons – whether they are folded (Fig. 4*a*) or unfolded (Fig. 4*b*) – cannot be determined. Their amide protons have already exchanged by way of lower free-energy unfolding pathways and the cross-peaks that represent them are not seen in the NMR analysis. They may be protected or unprotected in the green foldon unfolded state. At this stage of the analysis, the composition of the green unfolded PUF can only be represented as ???gB, as notated in Fig. 4, where '?' refers to the condition of the infrared, red, and yellow foldons. Similar unknowns occur for all of the PUFs as noted in Fig. 4.

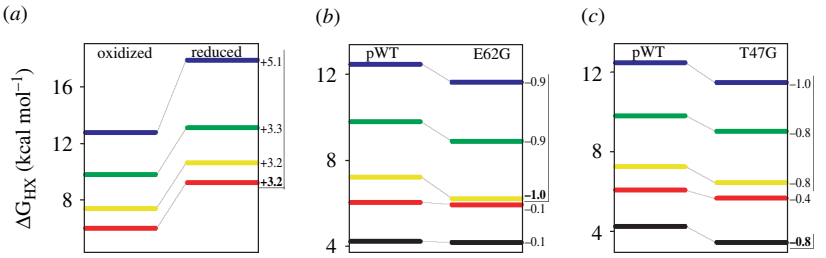


Fig. 5. Stability labeling experiments to determine partially unfolded forms composition. The modifications indicated directly alter the stability of individual foldons. The measured effect on the stability of the altered foldon (in boldface) and on the other foldons answer the question marks in Figure 4. The red and infrared foldons can unfold separately, in either order (*e*).

These uncertainties obscure the choice between independent and sequential unfolding. An experiment referred to as HX stability labeling can resolve the question.

5.1 The stability labeling experiment

Stability labeling experiments interrogate the condition (folded *vs.* unfolded) of any given foldon within the various PUFs by selectively perturbing its stability, for example by a specific mutation. The effect of the mutation on the targeted foldon itself and on the other foldons is then measured by the equilibrium NHX experiment. If foldons unfold independently as in Fig. 4*a*, a destabilizing mutation within one, for example within the yellow foldon, will promote yellow foldon unfolding but it will have no effect on the unfolding equilibrium of the others. If unfolding occurs dependently, e.g. in a sequential unfolding manner as in Fig. 4*b*, then a destabilizing mutation within the yellow foldon will equally promote the unfolding of all higher lying PUFs. They contain the yellow unfolding and will be destabilized by the same mutation. The lower-lying PUFs will not sense the presence of the mutation because the altered yellow foldon remains folded when they unfold. Partial effects can be interpreted in terms of mutual (de)stabilization rather than obligatory inclusion. The various possible interaction alternatives have been formalized (Krishna *et al.* 2007).

5.2 Stability labeling results

Figure 5*a* illustrates the results of a stability labeling experiment in which the red foldon in Cyt *c* was selectively perturbed by reducing the heme iron. This change stabilizes the heme iron bond to its Met80-S ligand in the red loop by 3·2 kcal mol^{-1}, measured independently. NHX results show that the stability of the red loop against unfolding is increased by the same amount (Xu *et al.* 1998; Maity *et al.* 2004). This is expected because the red loop unfolding breaks the Fe–S ligation. The higher-lying unfolding reactions are stabilized equally as expected for sequential unfolding but not for independent unfolding or for non-foldon models. A quantitative exception occurs for the final global blue unfolding which is stabilized by an additional 1·9 kcal mol^{-1}. This occurs because heme iron reduction imposes also a second stabilizing effect. It removes the destabilizing +1 charge on the buried oxidized heme iron. Only the last unfolding step (blue unfolding) senses this additional stabilization, evidently because it finally exposes the heme iron to solvent. (The infrared foldon was not measured in this experiment.)

In confirmation, another stability labeling experiment studied a Glu66Ala variant (Maity *et al.* 2005). This surface mutation does not perturb the native structure but it does remove a stabilizing salt link between the green helix and the red loop. Results show that the red loop and all higher-lying unfolding reactions, including in this case the blue unfolding, are equally destabilized by 0·8 kcal mol^{-1}. Thus the red foldon is unfolded in all of the higher-lying PUFs, as for sequential unfolding (Fig. 4*b*).

Another experiment (Fig. 5*b*) studied a destabilizing surface mutation within the yellow foldon, Glu62Gly, which removes the Glu62 to Lys60 salt link (Maity *et al.* 2004; Krishna *et al.* 2006). The yellow foldon itself and all higher-lying unfolding reactions are destabilized equally, by about 0·9 kcal mol^{-1}, but the lower-lying red and infrared foldons are unaffected. Thus the yellow foldon is unfolded in all of the higher-lying PUFs but not in the lower ones, as for sequential unfolding (Fig. 4*b*).

When a destabilizing mutation, Thr47Gly, was placed in the infrared foldon, the infrared foldon itself and all but one of the higher-lying foldons were destabilized equally, by $\sim$0·9 kcal mol^{-1} (Fig. 5*c*). Again as for sequential unfolding, the infrared foldon is unfolded in most of the higher-lying PUFs. The one exception, the red foldon, was destabilized only partially, by 0·4 kcal mol^{-1}. The analogous effect occurs in the other direction. When the red foldon was selectively stabilized, the infrared loop was stabilized partially (Krishna *et al.* 2006). These results show that the red and infrared loops can unfold separately but they are mutually stabilizing as might be expected from their large interaction surface. When one is unfolded, this promotes but does not require unfolding of the other. Thus the states iRYGB, IrYGB, and irYGB all occur.

5.3 The identity of Cyt *c* PUFs

These results remove the question marks noted before and specify the foldon composition of the dominant Cyt *c* PUFs as follows, written in the order of increased unfolding: IRYGB(native state), (iR or Ir)YGB, irYGB, iryGB, irygB, irygb(unfolded state). All other high-energy states must also exist but they must be less populated or accessed more slowly than the HX time scale.

6. PUFs to pathways

The size (*m* value) of the measured Cyt *c* unfolding reactions increases along with their unfolding free energy (ΔG_{HX}) (Fig. 3*d*). This correspondence, noted in the initial discovery of foldons, led Bai *et al.* (1995b) to suggest that the Cyt *c* foldons might unfold in a ladder-like way with each higher free-energy unfolded form containing all previous unfoldings, as shown in Fig. 4*b* and formalized in reaction Scheme 1 (the infrared foldon was found later). This interesting possibility, that progressively slower exchanging sets of amide hydrogens might represent a sequential unfolding pathway, had been suggested much earlier on other grounds (Englander & Kallenbach, 1983; Kim *et al.* 1993).

$$\text{RYGB(native)} \leftrightarrow \text{rYGB} \leftrightarrow \text{ryGB} \leftrightarrow \text{rygB} \leftrightarrow \text{rygb(unfolded)}.$$

Scheme 1

Importantly, if Scheme 1 represents the major unfolding sequence, then the reverse sequence must represent the major refolding pathway. This is true because the NHX experiments are done

under native conditions at equilibrium. Each unfolding step must be matched by an equal flow in the folding direction, otherwise detailed equilibrium could not be maintained.

6.1 Evidence from stability labeling

The stability labeling experiments described previously show that the foldon composition of the dominant partially unfolded states are just the ones suggested in reaction Scheme 1. In spite of the dictum that one cannot derive a kinetic mechanism from equilibrium data (Clarke & Fersht, 1996), these equilibrium-derived PUFs look remarkably like a sequence of kinetic pathway intermediates in which each PUF is produced, one from the other, by the unfolding or refolding of one additional foldon. Additional evidence supports this same reaction sequence, as follows.

6.2 Pathway order by kinetic NHX

A variant of the equilibrium NHX experiment known as kinetic NHX can determine the order of the reversible unfolding reactions as they proceed from the native state (Hoang *et al.* 2002). The experiment exploits EX1 HX behavior at increasing pH [Eq. (4)]. Results show that the red foldon unfolds first, followed in order by the yellow, then the green, and finally the blue foldon. Later results indicated that the infrared loop tends to unfold even before the red loop although they are able to unfold in either order, completing the foldon inventory (Krishna *et al.* 2004a, 2006).

This is just the time sequence for unfolding expected from Scheme 1.

6.3 Red foldon unfolds first

In some cases it has been possible to directly compare HX results with data from standard kinetic folding and unfolding experiments. The global Cyt *c* unfolding rate measured by stopped-flow as a function of high denaturant concentration can be extrapolated to zero denaturant. The rate obtained (10 s^{-1}) is equal to the rate for red foldon unfolding measured at zero denaturant by the kinetic NHX experiment (Hoang *et al.* 2002). Apparently both are rate-limited by the same early unfolding barrier, consistent with Scheme 1. At zero denaturant the other foldons unfold much more slowly, whereas they are all rate-limited by an early unfolding barrier at high denaturant. This behavior is the kinetic expression of the equilibrium free-energy ladder in Fig. 3*d* and the structurally determined sequential unfolding/refolding process. (The infrared foldon was not separately measured in these experiments.)

6.4 Blue foldon folds first

The HX pulse-labeling experiment described before identifies the blue N/C bi-helical foldon as the first to form during whole molecule kinetic folding (Fig. 2*a*). In the kinetically populated intermediate only the blue foldon is formed; the others join later. In agreement the equilibrium NHX experiment identifies the blue foldon as the last to unfold (at low denaturant; Fig. 3*b*). The other foldons are seen to unfold earlier at lower free energy, leaving the blue foldon to mark the final step to the globally unfolded state, as expected for Scheme 1.

Further information confirms the on-pathway nature of the blue foldon. The N/C intermediate forms in 3-state folding with the same rate ($\Delta G^{\ddagger}$), temperature dependence ($\Delta H^{\ddagger}$), and

structural ($m^{\ddagger}$) parameters as is seen for the rate-limiting step in 2-state folding (stopped-flow) (Sosnick *et al.* 1996). The barrier that determines the 2-state folding rate is obviously on-pathway, and it is the first pathway step, otherwise folding would not be 2-state. Therefore the N/C intermediate that it directly leads to in both 2-state and 3-state folding must be the first on-pathway intermediate in both cases, as in Scheme 1. Similar logic affirms the on-pathway nature of the red foldon unfolding just discussed.

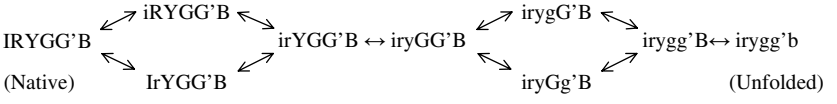
6.5 Green foldon folds next

The blue N/C bi-helical intermediate accumulates in 3-state Cyt *c* folding because it encounters a large barrier due to the misligation of a peripheral histidine to the heme iron. Both equilibrium and kinetic NHX experiments place the folding of the green foldon next. This sequence explains why the histidine misligation barrier blocks folding after the N/C bi-helix is formed. The peripheral histidine that is responsible for the heme misligation barrier is in the green loop (Fig. 1). In the initial globally unfolded state at neutral pH, the histidine is already misligated, holding the green loop on the wrong side of the heme. When folding is initiated, the N/C bi-helix forms normally but subsequent folding is blocked because the next foldon in line is the green unit. It cannot move into place until the misligation is released in a slow time-requiring error-repair process.

6.6 Pathway branching

Further experiments have shown that the Cyt *c* folding/unfolding pathway is not perfectly captured by Scheme 1 as initially conjectured. Low pH experiments designed to study the exchange of hydrogens that were otherwise too fast to measure show that the large bottom Ω-loop, initially thought to be part of the yellow foldon, is a separately cooperative least stable unit, now called the infrared foldon (Krishna *et al.* 2003b). Stability labeling experiments performed as a function of decreasing pH or mutation show that the infrared and red foldons can unfold in either order. Either one can unfold first and the other then joins (Krishna *et al.* 2006). Similarly, the two green segments initially thought to represent a single foldon can unfold separately (at low pH) and their unfolding order is optional (Krishna *et al.* 2007). Either the green loop or the green helix can unfold first, and is then joined by the other.

These results add some minimal branching to the linear sequence as in reaction Scheme 2. The Cyt *c* pathway that emerges, when read in either direction, exhibits four obligate steps and two optional branch points.



Scheme 2

6.7 Summary

Detailed structural, thermodynamic, and kinetic information for Cyt *c* has been obtained from HX pulse labeling, kinetic and equilibrium NHX, stability labeling, stopped-flow folding and unfolding, equilibrium melting, and misligation barrier insertion experiments. All results

consistently support a defined pathway that builds the final native protein by putting native-like foldon units into place in a stepwise manner. Some parallelism in the form of optional branching can occur, but in a way that is strictly limited by the cooperative foldon principle which specifies a limited number of folding units, and by the sequential stabilization principle which specifies the order of folding and unfolding events as described below.

7. Pathway construction by sequential stabilization

When the experimentally determined folding sequence of Cyt *c* (Scheme 2) is compared with its native structure (Fig. 1), a compelling observation emerges. Once the first native-like intermediate has been formed, the continuing sequence of obligatory steps and optional branching steps is determined by the way that the foldons fit together in the native protein.

7.1 The first pathway step

Why does the N-terminal to C-terminal bi-helical PUF form first? Basic polymer principles dictate that closure of the whole molecule loop by N-terminal to C-terminal docking is the least likely event from both kinetic and equilibrium points of view (Ozkan *et al.* 2007). Nevertheless N/C closure occurs as an initial step in kinetic folding, and the reverse is the final step in unfolding. Apparently the protein is designed to make the N/C bi-helix the most stable foldon unit and the first to fold. A similar result was found in HX labeling studies of apomyoglobin (Hughson *et al.* 1990; Jennings & Wright, 1993; Nishimura *et al.* 2006). An initial intermediate forms with the native-like N-terminal helix (helix A) and a C-terminal helical pair (helices G and H) docked against each other approximately as in the native protein.

A literature survey provoked by these results (Krishna & Englander, 2005) found that almost half of known proteins have their N-terminal and C-terminal elements docked together. Many appear to involve their terminal elements in a first folding step as indicated by φ-analysis and ψ-analysis. No folding model predicts this most common folding mode. Evidently evolutionary biology has overruled polymer physics, perhaps due to some selected advantage for early N/C closure or perhaps due to the history of protein evolutionary development which, one might speculate, may be reprised by the folding pathways of contemporary proteins.

7.2 Pathway sequence follows the native structure

Subsequent steps (Scheme 2) follow a more understandable pattern (Rumbley *et al.* 2001). The pathway order of foldon formation appears to be dictated by the native pattern. Where structure already in place is able to guide the formation of only one particular foldon, that step is found. Where prior structure can stabilize either of two incoming foldons, optional branching is seen.

In native Cyt *c* the N/C bi-helical unit contacts only the green helix and loop (Fig. 1). If folding proceeds in the native context, the initially formed blue unit can guide and stabilize only the two green segments. In fact, both green foldon segments are seen to fold next. They can form in either order (Scheme 2), just as suggested by native Cyt *c* structure, but both must be in place in order to support the next pathway step which brings the two sequentially distant yellow segments together into a minimal β-sheet. The two yellow segments are appended to two ends of the green segments. In turn, the pre-formed BGYri matrix is necessary to support the final steps,

formation of the least stable, mutually supportive red and infrared loops. As might be expected from the native structure, the experiment suggests that they can form in either order (Scheme 2).

We refer to this systematic behavior as sequential stabilization and take it as further evidence that the pathway progressively builds the native protein one step at a time in the native context.

7.3 Templating in biochemistry

The templating behavior seen here in terms of a protein folding pathway can be understood as a unimolecular version of familiar observations in bimolecular interactions. For example, intrinsically disordered peptides and proteins are induced to form complementary structure when they encounter their natural partners (Dyson & Wright, 2002). Segments isolated from the interaction region of binding proteins, although unstructured in solution, adopt the appropriate docking structure when they encounter their natural target protein (Sugase *et al.* 2007). In domain swapping, segments of one protein are induced to conform with the complementary structure of a sister protein (Schlunegger *et al.* 1997). Many proteins can be induced to join pre-formed amyloids by adopting closely complementary structures. Complementary nucleic acid strands are mutually stabilizing (Watson & Crick, 1953).

Biology is rooted in binding processes that depend on the interaction of small and large molecules with complementary protein structures. The so-called sequential stabilization process simply reframes in these terms the intramolecular foldon-dependent interactions that determine the protein folding process.

7.4 Summary

The foldon principle together with the phenomenon of complementary structure stabilization accounts for the stepwise nature and the experimentally observed sequence of steps in the Cyt *c* folding pathway. How the first pathway step is predetermined remains to be explained (Plaxco *et al.* 1998; Fersht & Daggett, 2007). Sosnick has suggested that the initial search for the nucleating transition state may similarly be directed by the sequential stabilization principle (Krantz *et al.* 2004; Sosnick *et al.* 2006).

8. Other proteins, other methods, similar results

Advances in science have most often depended on the exploitation of some particularly favorable model system. The question then arises whether the particular model or the methods used make the results atypical. The work just described depends heavily on the Cyt *c* model. Are the Cyt *c* results misleading, orchestrated somehow by its covalently bound heme group? Do the results reflect artifacts that depend on the chemical denaturants used or the HX methods themselves?

Although the bound heme group is unusual, Cyt *c* is a typical protein with the same primary, secondary, and tertiary structural elements and interactions common to all proteins. It may be that the presence of the bound heme group somehow aids the experimental discrimination of the different Cyt *c* foldons but it is hard to see how it could create foldons and orchestrate their sequential incorporation *de novo*.

In any case, analogous results have now been obtained in many laboratories for many proteins with and without prosthetic groups, using a variety of destabilants (urea, GdmCl, pH, temperature, pressure), or none at all, and even in experiments that did not use HX. Some examples can be noted.

8.1 Apomyoglobin (apoMb, heme removed)

The structure of an intermediate in the folding pathway of apoMb (eight helices, A–H) has been extensively studied by HX pulse labeling.

Initial study of a populated kinetic folding intermediate found that the A, G, and H helices are formed (Jennings & Wright, 1993). Helix B may be formed in part. The intermediate is native-like, obligatory, and on the folding pathway, although somewhat malleable. Individual helices can be added or removed by mutational manipulation of their relative stability (Cavagnero *et al.* 1999; Garcia *et al.* 2000). As for the blocked Cyt *c* intermediate, some misfolding is present (Jamin *et al.* 1999; Nishimura *et al.* 2006).

In agreement, HX protection studies of a pH 4 equilibrium molten globule found that when helices A, G, and H are formed, they dock against each other as in the native protein (Hughson *et al.* 1990; Eliezer *et al.* 2000; Bertagna & Barrick, 2004), and further stabilization by added trichloracetate induces helix B formation (Loh *et al.* 1995).

These results are much like the Cyt *c* results described previously. A discrete on-pathway intermediate with native-like interacting N-terminal and C-terminal elements is formed as an initial step. Subsequent folding is slow, suggesting a barrier limited by the need to first repair a non-native misfolding that is seen to be present (Nishimura *et al.* 2006) (see Section 9).

8.2 Ribonuclease H1 (RNase H)

Marqusee and co-workers extensively analyzed the folding behavior of RNase H, which consists of five helices (A–E) and five β-strands (I–V).

NHX results (Chamberlain *et al.* 1996; Chamberlain & Marqusee, 2000) for the native protein at equilibrium found two subglobal PUFs. In the lowest free-energy PUF strands I, II, III, V, and helix E are unfolded. The next higher PUF additionally unfolds helix B and strand IV, leaving helices A and D as the final unfolding unit. A similar ladder with the same order but slightly different groupings was found in a thermophilic homolog (Hollien & Marqusee, 1999).

In agreement, folding experiments (HX pulse labeling, CD, mutational effects) connect the earliest kinetic phase to formation of helices A and D and strand IV and a second phase largely to β-sheet formation, placing the PUFs found by NHX on the folding pathway (Raschke & Marqusee, 1997; Raschke *et al.* 1999). An HX protection study of the equilibrium acid molten globule shows that only the highest free-energy PUF with helices A and D and also B are well formed (Dabora *et al.* 1996), just as for apoMb. Related work found that a large synthetic fragment containing the contiguous helices A–D and strand IV displays independent stability (Chamberlain *et al.* 1999; Cecconi *et al.* 2005). In the different kinds of experiment, analogous PUFs were found to have similar stability.

These kinetic and thermodynamic results reveal cooperative foldon units in RNase H, describe their composition in terms of the secondary structural elements of the native protein, and show their role in determining a stepwise folding pathway. As for apoMb, additional structural units can be added to or subtracted from any given PUF by adjusting

condition-dependent local stabilities, but it appears that the rank order of PUF formation is robust, as expected from the sequential stabilization principle.

8.3 Apocytochrome b_{562} (apoCyt b_{562})

ApoCyt b_{562} is a four-helix bundle protein (I–IV) with the non-covalently bound heme group removed.

Fuentes & Wand (1998a, b) studied apoCyt b_{562} by native state HX using both GdmCl and pressure as destabilants. The results imply a sequence in which the core bi-helices II + III fold first followed in unspecified order by helices I and IV. Bai and co-workers stabilized apoCyt b_{562} by multiple mutations which provided the additional dynamic range in ΔG_{HX} needed to more definitively distinguish the different foldons by NHX (Chu *et al.* 2002). They also performed kinetic folding and phi analysis studies. All of the results consistently show that the core helices II + III fold first followed by helix IV and then helix I.

Bai and co-workers then created different constructs with the native state mutationally destabilized so that other states normally at higher free energy became dominantly populated, and solved their solution structures by NMR. Well-folded structures were found, nearly identical to the first and second PUFs found by NHX (Takei *et al.* 2002; Feng *et al.* 2003a, b, 2005a). The helices maintain their near-native main-chain conformation but the newly exposed apolar side-chains energy minimize by significant non-native repacking.

These results independently document the foldon units found by NHX and demonstrate a sequential foldon-dependent folding pathway. They also yield the important demonstration that incomplete native forms (intermediates, transition states) are likely to include non-native interactions and illustrate one case in atomic detail.

8.4 Outer surface protein A (OspA)

The 28 kDa outer surface protein from *B. burgdorferi* consists of 21 adjacent antiparallel β-strands and one α-helix arranged as a flat nine-stranded β-sheet capped on both ends by globular domains.

Koide and co-workers used NHX in EX1 and EX2 modes to dissect OspA into five co-operative foldon units (Yan *et al.* 2002). The NHX results and mutational phi analysis showed that at least two PUFs constructed from these foldons account for sequential intermediates in the folding pathway (Yan *et al.* 2004).

8.5 Triosephosphate isomerase (TIM)

Triosephosphate isomerase is the archetypal TIM barrel protein, with eight alternating α/β pairs. Silverman and Harbury studied the TIM homodimer by a technology that does not depend on HX (Silverman & Harbury, 2002b). They measured the accessibility to attack of 47 Cys residues inserted at various buried sites, so that reactivity depends on transient exposure by dynamic structural opening reactions (Silverman & Harbury, 2002a). Results found were exactly analogous to NHX results. With increasing denaturant, local fluctuational reaction pathways with zero *m* values merge into larger subglobal unfolding reactions that define three distinct PUFs. Further mutational studies analogous to the stability labeling experiments in Section 5 showed that the foldons unfold in a sequential pathway manner.

These results do not depend on any possible artifacts based on HX behavior. They demonstrate subglobal foldon units and their role in constructing a stepwise folding pathway.

8.6 Summary

Amino-acid-resolved information obtained by HX and non-HX methods under native conditions for many proteins of all structural types indicate that proteins may be viewed quite generally as accretions of cooperative foldon units. Where sufficiently detailed information is available, the results show that the steps in folding pathways are accounted for by the formation and association of native-like foldon units, and that this occurs in a sequential stabilization process that progressively grows the native protein. The two motivating concepts, cooperative foldon units and sequential stabilization, can be seen to represent straightforward expressions of highly documented protein behaviors.

9. Foldons and PUFs: principles and implications

The discovery of foldons and the high free-energy PUFs that they create has many implications for protein structure and function.

9.1 Foldon structure

One wants to understand the structural determinants of foldon units. The well documented foldons in Cyt *c* closely mimic entire native-like secondary structural elements or pairs thereof. This might be expected. When regular secondary structures are stable in isolation, they are seen to unfold and refold in a cooperative way. When they are integrated into larger protein molecules, the cooperative property may be modified but it is unlikely to be lost. The concerted unfolding of a major region of well structured helices or β-strands seems likely to entrain the entire structured element and terminate only at more flexible end residues.

The cooperativity relationship captured in the Zimm–Bragg (Zimm & Bragg, 1959) and Lifson–Roig (Lifson & Roig, 1961) formulations applies most obviously to helices. Some known foldons do encompass entire helical lengths (Bai *et al.* 1995b; Fuentes & Wand, 1998a; Krishna *et al.* 2003a; Feng *et al.* 2005a) but others may not (Chamberlain *et al.* 1996). Moreover, entire Ω-loops act as concerted unfolding units (Hoang *et al.* 2002; Krishna *et al.* 2003b), unsurprisingly so, because they are internally packed self-contained structures (Leszczynski & Rose, 1986). β-structures tend to break up into smaller separately cooperative units (Chamberlain *et al.* 1996; Yan *et al.* 2002, 2004; Bédard *et al.* 2008).

The component foldon units of PUFs need not precisely mimic the native structure. In unfolded forms, some of the interactions that delimit structural elements in the native protein are absent and other non-native energy-minimizing interactions may be present. Accordingly, one often finds evidence in partially folded intermediates for non-native interactions in addition to strikingly native-like conformation (Radford *et al.* 1992; Capaldi *et al.* 2002; Krishna *et al.* 2004b; Feng *et al.* 2005a; Bollen *et al.* 2006; Neudecker *et al.* 2007). Most noteworthy, Bai and colleagues solved the NMR solution structures of two apoCyt b_{562} PUFs (Feng *et al.* 2003a, 2004, 2005a). They are largely identical to the two partly unfolded forms found previously by NHX (Fuentes & Wand, 1998a, b; Chu *et al.* 2002; Takei *et al.* 2002) but distinct non-native interactions are also present, evidently because they help to energy minimize the PUFs.

In summary, foldon units tend to mimic the cooperative secondary structural units of the native protein from which they derive, or groupings thereof, but they may also incorporate energy-minimizing non-native interactions. Non-native interactions formed during kinetic folding may or may not act to impose large rate-limiting kinetic barriers, depending on the repair-reorganization rate.

9.2 The multi-state nature of protein molecules

The discovery of PUFs modifies the conviction, ingrained in protein chemical thinking by a half century of protein melting experiments (Fig. 3*a*), that protein molecules are rigorously 2-state objects. A more insightful view of the multi-state nature of protein molecules is illustrated in Fig. 3*d*. The figure shows the free-energy level relative to the native state of each Cyt *c* PUF as a function of increasing denaturant concentration (Bai & Englander, 1996).

The native and fully unfolded states are the lowest free energy most populated forms only in the melting transition region. Equilibrium unfolding, measured in this region, will tend to appear 2-state with only the native and fully unfolded states significantly populated (although see Mayne & Englander, 2000). Under native conditions individual PUFs may be more stable than the unfolded state, as in Fig. 3*d*. Here multi-state unfolding and refolding can be observed and dissected by NHX and related experiments.

We find that fully and partially unfolded proteins under folding conditions reach the native state by stepping down the ladder of intervening states. This can be directly observed by NHX methods when the intervening PUFs are at lower free energy than the U state. One can imagine that a refolding protein may well experience the same pathway progression, driven by the same structural determinants, even when intermediate PUFs are at higher free energy than U. This track, although uphill initially, provides an easy solution to the Levinthal paradox and may be pertinent to the search for transition states in apparent 2-state folding (Krantz *et al*. 2004; Sosnick *et al*. 2006).

One can also imagine that a jump from native into denaturing conditions may cause the protein to descend the now inverted ladder (Fig. 3*d*) to the lowest free-energy unfolded state.

9.3 The folding energy landscape

Results for many proteins now indicate that a small number of distinct native-like PUFs occupy the dominant local minima in the high free-energy folding landscape under native conditions. These PUFs are thermodynamically favored due to the intrinsically cooperative nature of foldon units and their stabilizing inter-foldon interactions. When foldons are stable relative to U, partially formed cooperative foldons are at higher free energy, essentially by definition.

The cooperative nature of PUFs also determines the stepwise progressive nature of the folding process. The PUFs that are observed by HX dominate the HX measurement precisely because they are the major intermediates visited by the continual unfolding and refolding behavior of the protein. All other possible forms implied in energy landscape diagrams must also exist but it appears that they are less represented or are kinetically isolated and accessed more slowly than the HX time scale. A corollary is that these other forms cannot carry significant flux in the folding/unfolding pathway.

Specifically structured foldons and PUFs have not appeared prominently in theoretical investigations of the landscape, perhaps because the calculations have so far not captured the full importance of local cooperativity effects which depend on multi-body interactions and local geometric considerations.

9.4 Foldons and function

In respect to protein function, it is interesting that the more stable protein foldons seem to provide a supporting scaffold while less stable regions, including loops and turns, often account for functional behaviors. This distribution of responsibility is generally obvious but the biophysical and functional ramifications of this distribution have not been much studied.

In the best worked-out case of Cyt *c*, the least stable red and infrared foldons correlate with known physical and functional attributes. Physical correlates include the observations that the lowest lying infrared segment is the most dynamic as seen by MD simulations (Garcia & Hummer, 1999), NMR relaxation (Fetrow & Baxter, 1999), and susceptibility to proteolysis (Wang & Kallenbach, 1998; Spolaore *et al.* 2001). Functionally, the lowest lying infrared loop unfolds as a first step in Cyt *c*-mediated apoptosis and necrosis, and it is seen to unfold when Cyt *c* binds to lipid membranes (Jemmerson *et al.* 1999). Decrease in stability of the infrared foldon correlates with decreased Cyt *c* expression (Fetrow *et al.* 1998). It is the least conserved segment in evolution, being deleted as a unit in some species (Pettigrew & Moore, 1987).

Both the equilibrium and the rate of a Cyt *c* structural transition have been found to be controlled by the properties of the two lowest-lying foldons. In the Cyt *c* akaline transition, the residue that ligands the heme iron changes by one position, from Met80 to Lys79 (and minimally also Lys73). These residues are all in the red loop foldon. The equilibrium stability of native Cyt *c* against the transition is quantitatively equal to the equilibrium stability of the red foldon against unfolding over a wide stability range, which can be manipulated by mutations and by denaturant addition (Maity *et al.* 2006). When Cyt *c* is jumped to high pH, the transition to the alkaline form is kinetically bi-exponential (Kihara *et al.* 1976). In the first phase, the deprotonation of an as yet unidentified buried group is rate-limited by the reversible unfolding of the infrared foldon, which allows access by OH^- ion (or buffer base). The slower phase in which the Met80 ligand is replaced by the neighboring Lys79 is ultimately rate-limited by the unfolding of the red foldon (Hoang *et al.* 2003). These observations indicate that, rather than simply sliding over one residue, the transition waits for the unfolding of the entire red foldon, then refolds in either the old or the new position (Hoang *et al.* 2003; Maity *et al.* 2006).

A dysfunctionality found for many proteins, their ability to be drawn into growing amyloid fibrils, undoubtedly requires the spontaneous unfolding of parts of the protein, like the reversible foldon unfolding behavior described above. For example, the tendency of transthyretin to form amyloid is known to be controlled by the stability of the tetramer against dissociation and transient unfolding (Kelly *et al.* 1997). The next step, recruitment into an amyloid fibril, involves an encounter with the growing end of a compatible amyloid fibril by the partially unfolded protein and its adoption of complementary beta structure driven by a templating mechanism. This intermolecular step mimics the intramolecular sequential stabilization process, which guides the analogous recruitment of foldons into the growing protein during kinetic folding.

9.5 Summary

In addition to the protein folding process, the structure and behavior of foldon units and the PUFs that they construct have significance for protein behavior and properties more broadly including stability, cooperativity, dynamics, design, function, and malfunction.

10. Folding models

10.1 Two fundamentally different views

The integrated foldon construction of proteins together with the sequential stabilization principle can be expected to produce well defined folding pathways. All of the protein molecules in a given population should fold identically, through the same set of predetermined intermediates, although with optional branching when pre-existing structure is able to template the formation of more than one foldon unit.

Theoretical and experimental results have often been interpreted very differently (Wallace & Matthews, 2002). Computer simulations find many alternative pathways in which proteins are constructed one amino acid at a time in no particular order (Plotkin & Onuchic, 2002a, b). Experimental results show that different population fractions can accumulate different kinetic intermediates and fold to the native state at different rates. These results have been interpreted in terms of independent unrelated pathway schemes that emulate the theoretically derived multi-track view, although with many fewer tracks.

This uncertainty concerns the most fundamental conceptualization of the folding process. Do proteins fold by the stepwise incorporation of foldon units, predetermined by native-like structure and interactions, or do different population fractions fold through unrelated intermediates in independent pathways? These clearly opposed interpretations of clearly correct experimental observations can be reconciled by the realization that refolding proteins tend to make errors in a probabilistic way. Different population fractions may experience blocking errors at different points in the pathway, and therefore populate different intermediates and fold at different rates. This behavior mimics independent unrelated pathways even though all of the molecules may use the same native-like foldons and fold through the same PUFs templated by the same native-like interactions.

We first consider whether an optional error hypothesis can account for known examples of heterogeneous folding kinetics. It can. We then consider structural information that can distinguish the two different folding models.

10.2 The independent unrelated pathways (IUP) model

Heterogeneous folding, seen by theory and experiment, is commonly taken to indicate that different population fractions fold by way of different intermediates in different pathways. Intermediates in the different tracks are not considered to be built from common native-like units. We call this view the IUP model to emphasize that intermediate structures in one track have no particular relationship to those in another.

IUP models drawn to fit heterogeneous kinetic folding results are of the form in Scheme 3, with the added possibility of cross-pathway interchanges. The reaction scheme is written to symbolize that different population fractions show different behaviors and different rates because they flow through different intermediates, or none at all, in different tracks through the energy

landscape. The experimental observation of additional folding phases is explained by the *ad hoc* postulation of additional pathways or intermediates. An inherent assumption is that the specified intermediates (I_i) will accumulate and produce 3-state kinetic folding.

$$U \leftrightarrow N$$

$$U \leftrightarrow I_1 \leftrightarrow \cdots N$$

$$U \leftrightarrow I_2 \leftrightarrow \cdots N$$

Scheme 3

10.3 The predetermined pathway – optional error (PPOE) model

In contrast, a quantity of structural information detailed before indicates that all of a protein population folds through a pathway that is predetermined by cooperative native-like foldon units and their native-like interactions. In order to explain heterogeneous folding, it is necessary to consider in addition an optional error hypothesis based on probabilistic misfolding, which leads to a PPOE model. The PPOE model assumes that all of a protein population does fold through the same productive on-pathway intermediates but that any one may be taken off the productive pathway by the addition of an optional structure-based error. For example Scheme 4 shows a simplified PPOE reaction scheme in a double T configuration (see Krishna & Englander, 2007 for a more general version).

$$\begin{array}{ccccccc} U & \leftrightarrow & I_1 & \leftrightarrow & I_2 & \leftrightarrow & N \\ & & \updownarrow & & \updownarrow & & \\ & & I_1^X & & I_2^X & & \end{array}$$

Scheme 4

Misfolding errors and the barriers that they impose are probabilistic, not intrinsically coded into the pathway. Therefore they may affect only some fraction of a refolding population. Scheme 4 sets the probability of committing a corrupting error at any pathway point as the I_i to I_i^X rate constant divided by the sum of all rate constants away from I_i. If the return error-repair step is slow, the misfold will act as an inserted kinetic barrier causing the accumulation of I_i^X and slowed folding in that particular population fraction.

Suppose that an optional barrier is encountered with probability between zero and unity. When the error probability is zero, 2-state folding will be seen as often occurs for small proteins. When the probability is high at some pathway point, 3-state folding will predominate. Any other combination will lead to heterogeneous folding with different population fractions blocking at different steps, or none at all, populating different partially corrupted intermediates, and reaching N at different rates. This behavior will give the appearance of multiple independent 2-state and 3-state pathways even though all of the protein population folds through the same productive intermediates in a single predetermined pathway (Sosnick *et al.* 1994; Sosnick *et al.* 1996; Englander *et al.* 1998b; Krishna *et al.* 2004b; Krishna & Englander, 2007).

10.4 Tests of the models

Heterogeneous folding has been observed in a number of well-studied proteins including Cyt *c*, α-tryptophan synthase, hen egg lysozyme, and staphylococcal nuclease. The kinetic data can be

successfully fit by multi-track IUP models, essentially by definition. Can the same results be explained by the PPOE model?

10.4.1 Cytochrome *c*

When a peripheral Cyt *c* histidine is deprotonated, it avidly misligates to the heme iron, inserting a barrier that imposes 3-state kinetic folding and intermediate accumulation. In this case one understands the source of the inserted barrier and why it is that the N/C intermediate is the particular one that accumulates. Both peripheral histidines are on the green loop (Fig. 1). Folding blocks after N/C formation because the green loop is programmed to fold next, but it is held out of place by the misligation.

When the Cyt *c* histidines are protonated at low pH or mutationally removed, misligation does not occur and the error-repair barrier is not inserted. One then observes that Cyt *c* is able to fold rapidly in a kinetically 2-state manner, as for many other proteins. The N/C intermediate is still obligatorily present, stable, and on-pathway, as are subsequent intermediates (shown by NHX), but they do not block and accumulate.

At less acid pH only some fraction of the Cyt *c* population misligates. Heterogeneous folding is then observed. Different population fractions, with and without the misligation, fold differently. In the absence of structural information, the heterogeneous folding kinetics might be interpreted in terms of multiple alternative pathways. Detailed structural evidence described before indicates otherwise. The entire population folds through the same major intermediates as dictated by their intrinsic foldon units and the sequential stabilization principle. The N/C intermediate that accumulates in 3-state folding is nearly identical to an unblocked analog in the 2-state case but it has in addition a slowly repaired misfolding error, as in reaction Scheme 4.

This example – the misligation error, its optional nature, and the kinetic behavior that it determines – provides a model for the basis of heterogeneous folding more generally.

10.4.2 α-Tryptophan synthase and proline isomerization

Spectroscopic and mutational studies have documented the heterogeneous folding of α-tryptophan synthase. Different population fractions accumulate different intermediates and fold at different rates, generating complex folding and unfolding behavior. This behavior was initially interpreted in terms of four independent folding tracks (Bilsel *et al.* 1999; Gualfetti *et al.* 1999). Further studies showed that each of the slow kinetic tracks can be removed by removing critical proline residues one or more at a time (Wu & Matthews, 2002, 2003).

Proline-dependent barriers may be considered the classical optional error example. They have been found for many proteins. When a critical proline residue pre-exists in a non-native isomeric form, folding of that population subfraction proceeds to a point, then may be blocked by a proline-dependent barrier. A partially corrupted intermediate accumulates. Population fractions with different proline barriers or none at all populate different intermediates and fold at different rates. This is precisely the behavior dictated by the PPOE model.

10.4.3 Hen egg-white lysozyme

Radford *et al.* (1992) used HX pulse labeling to show that a populated intermediate in lysozyme folding is heterogeneous and interpreted this result in terms of multiple parallel pathways. In

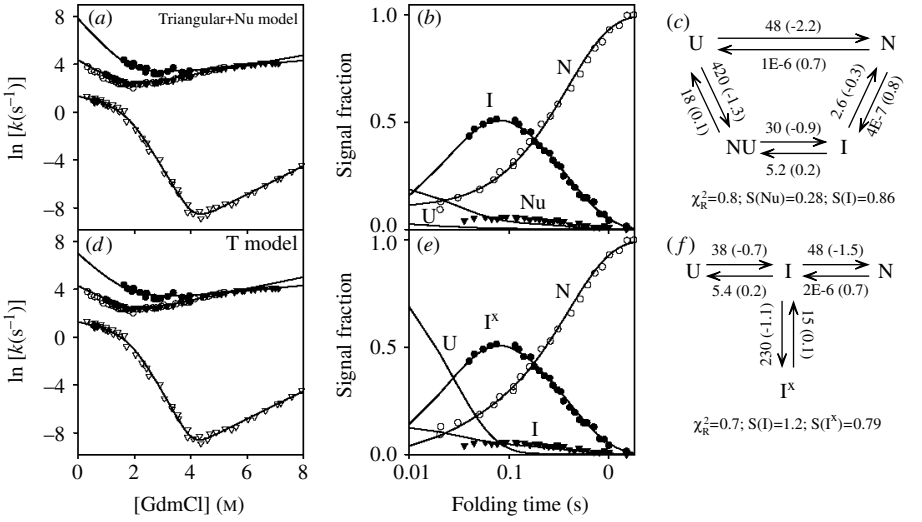


Fig. 6. Kinetic folding and unfolding results for hen egg lysozyme. Kiefhaber and co-workers measured the three kinetic phases and the time-dependent populations by standard stopped-flow fluorescence methods. The data shown are globally fit by a two-pathway IUP model [extended triangular model; panels (*a*)–(*c*)] and by a PPOE model [single T model; panels (*d*)–(*f*)]. Best-fit rate constants are shown, with *m* values in parentheses (boldface indicates well fixed parameters). Other datasets under other conditions with additional kinetic phases were equally well fit by both models but usually with fewer fitting constants for the PPOE model.

extensive experiments with lysozyme, Kiefhaber and co-workers discerned multiple kinetic folding and unfolding phases, and intermediate populations, and their dependence on denaturant concentration under a variety of conditions (Kiefhaber, 1995; Kiefhaber *et al.* 1997; Wildegger & Kiefhaber, 1997; Bieri *et al.* 1999; Bieri & Kiefhaber, 2001). Given the IUP assumption and other constraints, the lysozyme data were thought to require at least two independent pathways, framed as a triangular model (Wildegger & Kiefhaber, 1997). The observation of additional kinetic phases in the presence of elevated salt or pH was explained by additional intermediates and pathways (Bieri *et al.* 1999; Bieri & Kiefhaber, 2001).

We tested the ability of the PPOE model to explain the extensive kinetic results for lysozyme (Krishna & Englander, 2007). Figure 6 shows an example. Kiefhaber *et al.* (1997) measured the three kinetic phases and the time-dependent populations indicated in Fig. 6(*a*, *b*) by direct and interrupted kinetic folding and unfolding experiments (stopped-flow fluorescence). An IUP model minimally requires the two independent pathways in Fig. 6*c*, with eight fitting constants, each with an associated *m* value (in parentheses). The best global fit is shown.

The minimal single-T PPOE model matches all of the qualitative restrictions inferred by the Kiefhaber group to be required by their experimental results and it fits the data quantitatively (Fig. 6*d–f*). The PPOE data fit is at least as good as for the IUP model previously invoked (χ^2 shown) and it requires fewer fitting constants (6 *vs.* 8, plus *m* values for each). Similar results were found for folding at other solution conditions with additional kinetic phases where again the PPOE model was able to fit the data with equivalent χ^2 and equal or fewer fitting constants (Krishna & Englander, 2007).

10.4.4 Staphylococcal nuclease (SNase)

Direct and interrupted folding experiments performed by Kamagata *et al.* (2003) show that the fluorescence signal used to follow SNase folding directly detects native state formation. Different fractions of the refolding SNase population (proline-free) reach the native state at different rates. These observations were thought to require an IUP model with different pathways, intermediates, and rates. However, an analysis like that just described for lysozyme shows that a predetermined pathway model with optional errors fits the data just as well with the same number of fitting constants (Krishna *et al.* unpublished observations).

10.4.5 Summary

Contrary to prior belief, the observation of heterogeneous kinetic folding behavior does not require multi-pathway kinetic schemes. The examples just noted show that the PPOE model can fit heterogeneous folding data at least as well as alternative independent pathways models.

10.5 IUP or PPOE

Proteins may fold in a 2-state or a 3-state way or in a mixed heterogeneous way with 2-state and different 3-state fractions. The IUP and PPOE models are both framed to account for these complexities. Both models can provide a good fit to any kinetic folding data (Krishna & Englander, 2007), yet they represent fundamentally different physical mechanisms. It is essential to distinguish the correct model. The previous exercise shows that this distinction cannot be made on the basis of kinetic fitting alone.

The IUP model explains folding complexity in terms of different population fractions folding by way of wholly different intermediates and pathways, as in Scheme 3, often represented as different tracks in the energy landscape. Formal IUP reaction schemes assume that intermediates that accumulate in 3-state pathways do so simply because they are stable. In 2-state pathways written as U directly to N, the implicit assumption is that intermediates are absent or unstable.

Known information contradicts these assumptions. Stable on-pathway intermediates are not intrinsically slow. As we have seen, they can often be demonstrated and characterized even when folding is kinetically 2-state (NHX, SH reactivity analysis, relaxation dispersion). This can be expected. In order to add additional native-like structure to a native-like intermediate, only a small-scale guided search (sequential stabilization) is required. A small guided search to add additional complementary structure is likely to be much faster than the initial whole molecule unguided search necessary to find a nucleating structure that starts the productive pathway. Thus kinetic folding should be intrinsically 2-state.

Nevertheless, 3-state folding is common. On-pathway intermediates are seen to block and go forward far more slowly than they are formed (AGH in apoMb, N/C intermediate in Cyt *c*, A/D bi-helix plus β-strand in RNase H). In order to accumulate and cause 3-state folding, intermediates must be stable (relative to U) as assumed in IUP models, but they must also encounter a barrier that is larger than all prior barriers. In the PPOE model, it is the additional presence of an inserted error-repair barrier that distinguishes 3-state from apparent 2-state folding and not simply the presence of a stable intermediate.

In agreement, intermediates that accumulate in 3-state folding are often seen to harbor misfolding errors, including Cyt *c* (Elöve *et al.* 1994; Sosnick *et al.* 1994), apoMb (Nishimura *et al.* 2006), Im7 (Capaldi *et al.* 2002), and α-Trp synthase (Wu & Matthews, 2003). This observation

provides a nececessary but not a sufficient proof for the blocking role of misfolding because non-native energy-minimizing interactions are likely to be present in any incompletely native state. In the case of Cyt *c*, it has been possible to compare the populated N/C bi-helical intermediate seen in 3-state folding with a kinetically invisible non-populating intermediate seen by NHX. They are identical (Figs 2, 3*b*) except for the known histidine to heme misligation error. In Cyt *c* it is the misligation error that accounts for 3-state folding. In the absence of errors folding is 2-state. The same is true for other well known errors (proline misisomerization, aggregation, etc.).

Folding errors are known to be ubiquitous, both *in vitro* where they obstruct laboratory and industrial protein production (Chi *et al*. 2003) and *in vivo* where they may result in the loss of more than 30% of synthesized polypeptides (Yewdell, 2005) and account for a significant fraction of human pathologies. Well-known misfolding errors that can transiently block folding include prolyl and non-prolyl peptide bond misisomerization, transient aggregation, non-native hydrophobic clusters, disulfide shuffling, heme misligation, topological misarrangement, and non-native domain docking modes. These errors are optional, not intrinsic to the folding process. They can often be made to appear or not depending on ambient conditions. For example, proline misisomerization barriers depend on the length of time the protein is held unfolded before initiating refolding. Aggregation-dependent barriers can depend on concentration, temperature, and pH. Optional errors will variably affect different population fractions, leading to heterogeneous folding. Intermediates that populate in 3-state folding are often seen to be native-like, especially by HX pulse labeling (e.g. Fig. 2). It is unlikely that they are produced by, rather than rendered kinetically visible by, these non-specific optionally inserted barriers.

The PPOE model formalizes all of these observations in terms of productive, intrinsically fast intermediates in a stepwise pathway that is common to the entire refolding population. In heterogeneous folding it is the optional misfolding barriers that vary and not the pathway or its predetermined on-pathway intermediates.

10.6 Comparison with theoretical results

Theory-based IUP models generally visualize many independent pathways, commonly represented as multiple tracks that descend through a funnel-shaped energy landscape. As noted before, the experimental support for this view is questionable. We suspect that theory-based results do not find concerted foldons and their sequential stabilization behavior because the pairwise potential functions and model geometries commonly used do not adequately capture the fundamental cooperativity of protein structures on which foldon structures and PUF to foldon interactions are based (Chan *et al*. 2004, although see Weinkam *et al*. 2005).

In any case, the conflict between the PPOE model and simulation results seems to be more apparent than real. The multi-track behavior commonly observed by theory focuses a microscopic view that seems more pertinent to the initial unguided whole molecule search step. Experiment is only able to detect macroscopic folding pathway behavior. At this level, detailed structural information supports a much more determinate pathway.

10.7 Summary

IUP models, motivated by kinetic rather than structural information, account for heterogeneous folding by the *ad hoc* postulation of multiple independent pathways. However, the properties of pathway intermediates implicit in the IUP model – their unrelatedness and their intrinsically slow

folding – are contradicted by experiment. Detailed results now available for the structure of folding intermediates and the pathways that they construct show protein folding to be a much more determinate process directed at each step by the same principles and interactions that determine the final native state. The PPOE model is a formal statement of this information. It is able to quantitatively fit observed kinetic folding behavior at least as well as IUP models. It is not inconsistent with the concept of a funneled energy landscape, even though this abstraction is often drawn to represent IUP-like behavior with multiple independent folding tracks.

11. The principles of protein folding

A great deal of experimental work summarized in this article documents ordered protein folding pathways based on native-like foldon units and intermediates. In retrospect this behavior can be seen to emerge as a consequence of three self-obvious principles – the cooperativity of foldon units, their sequential stabilization, and the ubiquity and chance nature of folding errors.

The independently cooperative nature of the foldon building blocks that compose native proteins is well known. Much information now demonstrates their continued unfolding-refolding behavior in structured proteins. The intrinsically cooperative nature of foldon units predisposes them rather than individual amino acids or other fractional units to account for the unit steps in protein unfolding and refolding. Thermodynamic principle requires that their continued unfolding and refolding under native conditions must recapitulate the protein's natural folding pathway.

The sequential formation of native-like foldons in a stepwise pathway manner flows naturally from the obvious fact that the association of complementary structures is energetically favorable. Pre-existing structure guides and stabilizes the formation of complementary structure.

Acting alone these two properties would dictate a common stepwise pathway that sequentially forms and assembles native-like foldon units of the target native protein in a kinetically 2-state manner, rate-limited by the initial pathway step. This kind of pathway behavior is observed for some proteins but others exhibit more complex behavior. The complexities that are observed can be explained by the widely confirmed tendency of proteins to commit chance misfolding errors that insert time-consuming error-repair barriers at unpredictable points in the folding pathway.

The authenticity of each of these three principles is beyond question. Taken together, they specify *how* proteins fold – by way of predetermined native-like intermediates in predetermined native-like pathways, interspersed with chance misfolding errors – and they explain *why* proteins should fold in that way. Precisely this behavior has now been experimentally observed for a number of proteins. It seems unlikely that other proteins behave very differently.

12. Acknowledgements

We acknowledge with gratitude and admiration the efforts and contributions of the many able scientists who have contributed so much to the advancement of the protein folding field, in our own laboratory and in many others. The authors of this article were supported by research grants from the NIH (GM 031847, GM 075105), the Mathers Charitable Foundation, and a faculty seed grant from the School of Pharmacy of the University of Colorado at Denver (M.M.G.K.).

13. References

Anfinsen, C. B. (1973). Principles that govern the folding of protein chains. *Science* **181**, 223–230.

Anfinsen, C. B., Haber, E., Sela, M. & White, F. H. (1961). Kinetics of formation of native ribonuclease during oxidation of reduced polypeptide chain. *Proceedings of the National Academy of Sciences USA* **47**, 1309–1314.

Bahar, I., Wallqvist, A., Covell, D. G. & Jernigan, R. L. (1998). Correlation between native-state hydrogen exchange and cooperative residue fluctuations from a simple model. *Biochemistry* **37**, 1067–1075.

Bai, Y., Englander, J. J., Mayne, L., Milne, J. S. & Englander, S. W. (1995a). Thermodynamic parameters from hydrogen exchange measurements. *Methods in Enzymology* **259**, 344–356.

Bai, Y. & Englander, S. W. (1996). Future directions in folding: the multi-state nature of protein structure. *Proteins: Structure, Function, Genetics* **24**, 145–151.

Bai, Y., Milne, J. S., Mayne, L. & Englander, S. W. (1993). Primary structure effects on peptide group hydrogen exchange. *Proteins: Structure, Function, Genetics* **17**, 75–86.

Bai, Y., Milne, J. S., Mayne, L. & Englander, S. W. (1994). Protein stability parameters measured by hydrogen exchange. *Proteins: Structure, Function, Genetics* **20**, 4–14.

Bai, Y., Sosnick, T. R., Mayne, L. & Englander, S. W. (1995b). Protein folding intermediates: native-state hydrogen exchange. *Science* **269**, 192–197.

Baum, J., Dobson, C. M., Evans, P. A. & Hanley, C. (1989). Characterization of a partly folded protein by NMR methods: studies on the molten globule state of guinea pig alpha-lactalbumin. *Biochemistry* **28**, 7–13.

Bédard, S., Mayne, L., Peterson, R. W., Wand, A. J. & Englander, S. W. (2008). The foldon substructure of staphylococcal nuclease. *Journal of Molecular Biology* **376**, 1142–1154.

Bertagna, A. M. & Barrick, D. (2004). Nonspecific hydrophobic interactions stabilize an equilibrium intermediate of apomyoglobin at a key position within the AGH region. *Proceedings of the National Academy of Sciences USA* **101**, 12514–12519.

Bieri, O. & Kiefhaber, T. (2001). Origin of apparent fast and non-exponential kinetics of lysozyme folding measured in pulsed hydrogen exchange measurements. *Journal of Molecular Biology* **310**, 919–935.

Bieri, O., Wildegger, G., Bachmann, A., Wagner, C. & Kiefhaber, T. (1999). A salt-induced kinetic intermediate is on a new parallel pathway of lysozyme folding. *Biochemistry* **38**, 12460–12470.

Bilsel, O., Zitzewitz, J. A., Bowers, K. E. & Matthews, C. R. (1999). Folding mechanism of the a-subunit of tryptophan synthase, an a/b barrel protein: global analysis highlights the interconversion of multiple native, intermediate, and unfolded forms through parallel channels. *Biochemistry* **38**, 1018–1029.

Bollen, Y. J. M., Kamphuis, M. B. & van Mierlo, C. P. M. (2006). The folding energy landscape of apoflavodoxin is rugged: hydrogen exchange reveals nonproductive misfolded intermediates. *Proceedings of the National Academy of Sciences USA* **103**, 4095–4100.

Bryngelson, J. D., Onuchic, J. N., Socci, N. D. & Wolynes, P. G. (1995). Funnels, pathways, and the energy landscape of protein folding: a synthesis. *Proteins: Structure, Function, and Genetics* **21**, 167–195.

Capaldi, A. P., Kleanthous, C. & Radford, S. E. (2002). Im7 folding mechanism: misfolding on a path to the native state. *Nature Structural Biology* **9**, 209–216.

Cavagnero, S., Dyson, H. J. & Wright, P. E. (1999). Effect of H helix destabilizing mutations on the kinetic and equilibrium folding of apomyoglobin. *Journal of Molecular Biology* **285**, 269–282.

Cecconi, C., Shank, E. A., Bustamante, C. & Marqusee, S. (2005). Direct observation of the three-state folding of a single protein molecule. *Science* **309**, 2057–2060.

Cellitti, J., Bernstein, R. & Marqusee, S. (2007). Exploring subdomain cooperativity in T4 lysozyme II: uncovering the C-terminal subdomain as a hidden intermediate in the kinetic folding pathway. *Protein Science* **16**, 852–862.

Chamberlain, A. K., Fischer, K. F., Reardon, D., Handel, T. M. & Marqusee, A. S. (1999). Folding of an isolated ribonuclease H core fragment. *Protein Science* **8**, 2251–2257.

Chamberlain, A. K., Handel, T. M. & Marqusee, S. (1996). Detection of rare partially folded molecules in equilibrium with the native conformation of RNase H. *Nature Structural Biology* **3**, 782–787.

Chamberlain, A. K. & Marqusee, S. (2000). Comparison of equilibrium and kinetic approaches for determining protein folding mechanisms. *Advances in Protein Chemistry* **53**, 283–328.

Chan, H. S., Shimizu, S. & Kaya, H. (2004). Cooperativity principles in protein folding. *Methods in Enzymology* **380**, 350–379.

Chi, E. Y., Krishnan, S., Randolph, T. W. & Carpenter, J. F. (2003). Physical stability of proteins in aqueous solution: mechanism and driving forces in nonnative protein aggregation. *Pharmaceutical Research* **20**, 1325–1336.

Chu, R., Pei, W., Takei, J. & Bai, Y. (2002). Relationship between native-state hydrogen exchange and the folding pathway of a four-helix bundle protein. *Biochemistry* **41**, 7998–8003.

Clarke, J. & Fersht, A. R. (1996). An evaluation of the use of hydrogen exchange at equilibrium to probe intermediates on the protein folding pathway. *Folding & Design* **1**, 243–254.

Colon, W., Elove, G. A., Wakem, L. P., Sherman, F. & Roder, H. (1996). Side chain packing of the N- and C-terminal helices plays a critical role in the kinetics of cytochrome *c* folding. *Biochemistry* **35**, 5538–5549.

Connelly, G. P., Bai, Y., Jeng, M.-F., Mayne, L. & Englander, S. W. (1993). Isotope effects in peptide group hydrogen exchange. *Proteins: Structure, Function, Genetics* **17**, 87–92.

Creighton, T. E. (1986). Disulfide bonds as probes of protein folding pathways. *Methods in Enzymology* **131**, 83–106.

Dabora, J. M., Pelton, J. G. & Marqusee, S. (1996). Structure of the acid state of *Escherichia coli* ribonuclease HI. *Biochemistry* **35**, 11951–11958.

De Lorenzo, F., Goldberger, R., Steers, E. J., Givol, D. & Anfinsen, C. B. (1966). Purification and properties of an enzyme from beef liver which catalyzes sulfhydryl-disulfide interchange in proteins. *Journal of Biological Chemistry* **241**, 1562–1567.

Dill, K. A. (1985). Theory for the folding and stability of globular proteins. *Biochemistry* **24**, 1501–1509.

Dyson, H. J. & Wright, P. E. (2002). Coupling of folding and binding for unstructured proteins. *Current Opinion in Structural Biology* **12**, 54–60.

Eliezer, D., Chung, J., Dyson, H. J. & Wright, P. E. (2000). Native and non-native secondary structure and dynamics in the pH 4 intermediate of apomyoglobin. *Biochemistry* **39**, 2894–2901.

Elöve, G. A., Bhuyan, A. K. & Roder, H. (1994). Kinetic mechanism of cytochrome c folding: involvement of the heme and its ligands. *Biochemistry* **33**, 6925–6935.

Englander, J. J., Del Mar, C., Li, W., Englander, S. W., Kim, J. S., Stranz, D. D., Hamuro, Y. & Woods, V. L. J. (2003). Protein structure change studied by hydrogen-deuterium exchange, functional labeling, and mass spectrometry. *Proceedings of the National Academy of Sciences USA* **100**, 7057–7062.

Englander, J. J., Louie, G., McKinnie, R. E. & Englander, S. W. (1998a). Energetic components of the allosteric machinery in hemoglobin measured by hydrogen exchange. *Journal of Molecular Biology* **284**, 1695–1706.

Englander, J. J., Rogero, J. R. & Englander, S. W. (1985). Protein hydrogen exchange studied by the fragment separation method. *Analytical Biochemistry* **147**, 234–244.

Englander, S. W. (1963). A hydrogen exchange method using tritium and Sephadex. Application to ribonuclease. *Biochemistry* **2**, 798–807.

Englander, S. W. (2006). Hydrogen exchange and mass spectrometry: a historical perspective. *Journal of the American Society for Mass Spectrometry* **17**, 1481–1489.

Englander, S. W. & Kallenbach, N. R. (1983). Hydrogen exchange and structural dynamics of proteins and nucleic-acids. *Quarterly Reviews of Biophysics* **16**, 521–655.

Englander, S. W., Mayne, L., Bai, Y. & Sosnick, T. R. (1997). Hydrogen exchange: the modern legacy of Linderstrom-Lang. *Protein Science* **6**, 1101–1109.

Englander, S. W., Sosnick, T. R., Mayne, L. C., Shtilerman, M., Qi, P. X. & Bai, Y. W. (1998b). Fast and slow folding in cytochrome c. *Accounts of Chemical Research* **31**, 737–744.

Eyles, S. J. & Kaltashov, I. A. (2004). Methods to study protein dynamics and folding by mass spectrometry. *Methods* **34**, 88–99.

Feng, H., Takei, J., Lipsitz, R., Tjandra, N. & Bai, Y. (2003a). Specific non-native hydrophobic interactions in a hidden folding intermediate: implications for protein folding. *Biochemistry* **42**, 12461–12465.

Feng, H., Takei, J., Lipsitz, R., Tjandra, N. & Bai, Y. (2004). The high-resolution structure of a protein intermediate state: implications for protein folding. *Protein Science* **13**, 219–220.

Feng, H., Zhou, Z. & Bai, Y. (2005a). A protein folding pathway with multiple folding intermediates at atomic resolution. *Proceedings of the National Academy of Sciences USA* **102**, 5026–5031.

Feng, H. Q., Vu, N. D. & Bai, Y. W. (2005b). Detection of a hidden folding intermediate of the third domain of PDZ. *Journal of Molecular Biology* **346**, 345–353.

Feng, Y., Liu, D. & Wang, J. (2003b). Native-like partially folded conformations and folding process revealed in the N-terminal large fragments of staphylococcal nuclease: a study by NMR spectroscopy. *Journal of Molecular Biology* **330**, 821–837.

Fersht, A. R. & Daggett, V. (2007). Folding and binding: implementing the game plan. *Current Opinion in Structural Biology* **17**, 1–2.

Fetrow, J. S. & Baxter, S. M. (1999). Assignment of ^{15}N chemical shifts and ^{15}N relaxation measurements for oxidized and reduced iso-1-cytochrome c. *Biochemistry* **38**, 4480–4492.

Fetrow, J. S., Dreher, U., Wiland, D. J., Schaak, D. L. & Boose, T. L. (1998). Mutagenesis of histidine 26 demonstrates the importance of loop-loop and loop-protein interactions for the function of iso-1-cytochrome c. *Protein Science* **7**, 994–1005.

Fine, R., Dimmler, G. & Levinthal, C. (1991). Fastrun – a special purpose, hardwired computer for molecular simulation. *Proteins: Structure, Function, Genetics* **11**, 242–253.

Fischer, K. F. & Marqusee, S. (2000). A rapid test for identification of autonomous folding units in proteins. *Journal of Molecular Biology* **302**, 701–712.

Fuentes, E. J. & Wand, A. J. (1998a). Local dynamics and stability of apocytochrome b562 examined by hydrogen exchange. *Biochemistry* **37**, 3687–3698.

Fuentes, E. J. & Wand, A. J. (1998b). Local stability and dynamics of apocytochrome b562 examined by the dependence of hydrogen exchange on hydrostatic pressure. *Biochemistry* **37**, 9877–9883.

Garcia, A. E. & Hummer, G. (1999). Conformational dynamics of cytochrome c: correlation to hydrogen exchange. *Proteins: Structure, Function, Genetics* **36**, 175–191.

Garcia, C., Nishimura, C., Cavagnero, S., Dyson, H. J. & Wright, P. E. (2000). Changes in the apomyoglobin folding pathway caused by mutation of the distal histidine residue. *Biochemistry* **39**, 11227–11237.

Gualfetti, P. J., Bilsel, O. & Matthews, C. R. (1999). The progressive development of structure and stability during the equilibrium folding of the alpha subunit of tryptophan synthase from *Escherichia coli*. *Protein Science* **8**, 1623–1635.

Hernandez, G., Jenney, F. E., Adams, M. W. W. & LeMaster, D. M. (2000). Millisecond time scale conformational flexibility in a hyperthermophile protein at ambient temperature. *Proceedings of the National Academy of Sciences USA* **97**, 3166–3170.

Hilser, V. J., Garcia-Moreno, B. E., Oas, T. G., Kapp, G. & Whitten, S. T. (2006). A statistical thermodynamic model of the protein ensemble. *Chemical Reviews* **106**, 1545–1558.

Hoang, L., Bédard, S., Krishna, M. M. G., Lin, Y. & Englander, S. W. (2002). Cytochrome *c* folding pathway: kinetic native-state hydrogen exchange. *Proceedings of the National Academy of Sciences USA* **99**, 12173–12178.

Hoang, L., Maity, H., Krishna, M. M., Lin, Y. & Englander, S. W. (2003). Folding units govern the cytochrome *c* alkaline transition. *Journal of Molecular Biology* **331**, 37–43.

Hollien, J. & Marqusee, S. (1999). A thermodynamic comparison of mesophilic and thermophilic ribonucleases H. *Biochemistry* **38**, 3831–3836.

Hughson, F. M., Wright, P. E. & Baldwin, R. L. (1990). Structural characterization of a partly folded apomyoglobin intermediate. *Science* **249**, 1544–1548.

Huyghues-Despointes, B. M., Pace, C. N., Englander, S. W. & Scholtz, J. M. (2001). Measuring the conformational stability of a protein by hydrogen exchange. *Methods in Molecular Biology* **168**, 69–92.

Hvidt, A. (1964). A discussion of the pH dependence of the hydrogen-deuterium exchange of proteins. *Comptes Rendus des Travaux du Laboratoire Carlsberg Séries Chimique* **34**, 299–317.

Hvidt, A. & Nielsen, S. O. (1966). Hydrogen exchange in proteins. *Advances in Protein Chemistry* **21**, 287–386.

Jamin, M., Yeh, S. R., Rousseau, D. L. & Baldwin, R. L. (1999). Submillisecond unfolding kinetics of apomyoglobin and its pH 4 intermediate. *Journal of Molecular Biology* **292**, 731–740.

Jemmerson, R., Liu, J., Hausauer, D., Lam, K.-P., Mondino, A. & Nelson, R. D. (1999). A conformational change in cytochrome *c* of apoptotic and necrotic cells is detected by monoclonal antibody binding and mimicked by association of the native antigen with synthetic phospholipid vesicles. *Biochemistry* **38**, 3599–3609.

Jeng, M. F., Englander, S. W., Elöve, G. A., Wand, A. J. & Roder, H. (1990). Structural description of acid-denatured cytochrome c by hydrogen exchange and 2D NMR. *Biochemistry* **29**, 10433–10437.

Jennings, P. A. & Wright, P. E. (1993). Formation of a molten globule intermediate early in the kinetic folding pathway of apomyoglobin. *Science* **262**, 892–896.

Kamagata, K., Sawano, Y., Tanokura, M. & Kuwajima, K. (2003). Multiple parallel-pathway folding of proline-free staphylococcal nuclease. *Journal of Molecular Biology* **332**, 1143–1153.

Kato, H., Vu, N. D., Feng, H. Q., Zhou, Z. & Bai, Y. W. (2007). The folding pathway of T4 lysozyme: an on-pathway hidden folding intermediate. *Journal of Molecular Biology* **365**, 881–891.

Kelly, J. W., Colon, W., Lai, Z., Lashuel, H. A., McCulloch, J., McCutchen, S. L., Miroy, G. J. & Peterson, S. A. (1997). Transthyretin quaternary and tertiary structural changes facilitate misassembly into amyloid. *Advances in Protein Chemistry* **50**, 161–181.

Kiefhaber, T. (1995). Kinetic traps in lysozyme folding. *Proceedings of the National Academy of Sciences USA* **92**, 9029–9033.

Kiefhaber, T., Bachmann, A., Wildegger, G. & Wagner, C. (1997). Direct measurement of nucleation and growth rates in lysozyme folding. *Biochemistry* **36**, 5108–5112.

Kihara, H., Saigo, S., Nakatani, H., Hiromi, K., Ikeda-Saito, M. & Iizuka, T. (1976). Kinetic study of isomerization of ferricytochrome *c* at alkaline pH. *Biochimica et Biophysica Acta* **430**, 225–243.

Kim, K. S., Fuchs, J. A. & Woodward, C. K. (1993). Hydrogen exchange identifies native-state motional domains important in protein folding. *Biochemistry* **32**, 9600–9608.

Kim, P. S. & Baldwin, R. L. (1982a). Influence of charge on the rate of amide proton exchange. *Biochemistry* **21**, 1–5.

Kim, P. S. & Baldwin, R. L. (1982b). Specific intermediates in the folding reactions of small proteins and the mechanism of protein folding. *Annual Review of Biochemistry* **51**, 459–489.

Kim, P. S. & Baldwin, R. L. (1990). Intermediates in the folding reactions of small proteins. *Annual Review of Biochemistry* **59**, 631–660.

Korzhnev, D. M., Neudecker, P., Zarrine-Afsar, A., Davidson, A. R. & Kay, L. E. (2006). Abp1p and Fyn SH3 domains fold through similar low-populated intermediate states. *Biochemistry* **45**, 10175–10183.

Korzhnev, D. M., Religa, T. L., Lundstrom, P., Fersht, A. R. & Kay, L. E. (2007). The folding pathway of an FF domain: characterization of an on-pathway intermediate state under folding conditions by (15)N, (13)C(alpha) and (13)C-methyl relaxation dispersion and (1)H/(2)H-exchange NMR spectroscopy. *Journal of Molecular Biology* **372**, 497–512.

Korzhnev, D. M., Salvatella, X., Vendruscolo, M., Nardo, A. A. D., Davidson, A. R., Dobson, C. M. & Kay, L. E. (2004). Low-populated folding intermediates of Fyn SH3 characterized by relaxation dispersion NMR. *Nature* **430**, 586–590.

Krantz, B. A., Dothager, R. S. & Sosnick, T. R. (2004). Discerning the structure and energy of multiple transition states in protein folding using y-analysis. *Journal of Molecular Biology* **337**, 463–475.

Krishna, M. M. G. & Englander, S. W. (2005). The N-terminal to C-terminal motif in protein folding and function. *Proceedings of the National Academy of Sciences USA* **102**, 1053–1058.

Krishna, M. M. G. & Englander, S. W. (2007). A unified mechanism for protein folding: predetermined pathways with optional errors. *Protein Science* **16**, 449–464.

Krishna, M. M. G., Hoang, L., Lin, Y. & Englander, S. W. (2004a). Hydrogen exchange methods to study protein folding. *Methods* **34**, 51–64.

Krishna, M. M. G., Lin, Y. & Englander, S. W. (2004b). Protein misfolding: optional barriers, misfolded intermediates, and pathway heterogeneity. *Journal of Molecular Biology* **343**, 1095–1109.

Krishna, M. M. G., Lin, Y., Mayne, L. & Walter Englander, S. (2003a). Intimate view of a kinetic protein folding intermediate: residue-resolved structure, interactions, stability, folding and unfolding rates, homogeneity. *Journal of Molecular Biology* **334**, 501–513.

Krishna, M. M. G., Lin, Y., Rumbley, J. N. & Englander, S. W. (2003b). Cooperative omega loops in cytochrome c: role in folding and function. *Journal of Molecular Biology* **331**, 29–36.

Krishna, M. M. G., Maity, H., Rumbley, J. N. & Englander, S. W. (2007). Branching in the sequential folding pathway of cytochrome c. *Protein Science* **16**, 1946–1956.

Krishna, M. M. G., Maity, H., Rumbley, J. N., Lin, Y. & Englander, S. W. (2006). Order of steps in the cytochrome *c* folding pathway: evidence for a sequential stabilization mechanism. *Journal of Molecular Biology* **359**, 1411–1420.

Leopold, P. E., Montal, M. & Onuchic, J. N. (1992). Protein folding funnels: a kinetic approach to the sequence-structure relationship. *Proceedings of the National Academy of Sciences USA* **89**, 8721–8725.

Leszczynski, J. F. & Rose, G. D. (1986). Loops in globular proteins: a novel category of secondary structure. *Science* **234**, 849–855.

Levinthal, C. (1969). How to fold graciously. In *Mossbauer Spectroscopy in Biological Systems. Proceedings, University of Illinois Bulletin*, vol. 67, pp. 22–24. Urbana, IL: University of Illinois Press.

Lifson, S. & Roig, A. (1961). On the theory of the helix-coil transition in polypeptides. *Journal of Chemical Physics* **34**, 1963–1974.

Linderstrøm-Lang, K. (1958). Deuterium exchange and protein structure. In *Symposium on Protein Structure* (ed. A. Neuberger). London: Methuen.

Linderstrøm-Lang, K. U. & Schellman, J. A. (1959). Protein structure and enzyme activity. In *The Enzymes* (eds P. D. Boyer, H. Lardy and K. Myrback), pp. 443–510. New York: Academic Press.

Loh, S. N., Kay, M. S. & Baldwin, R. L. (1995). Structure and stability of a second molten globule intermediate in the apomyoglobin folding pathway. *Proceedings of the National Academy of Sciences USA* **92**, 5446–5450.

Loh, S. N., Prehoda, K. E., Wang, J. & Markley, J. L. (1993). Hydrogen exchange in unligated and ligated staphylococcal nuclease. *Biochemistry* **32**, 11022–11028.

Maity, H., Lim, W. K., Rumbley, J. N. & Englander, S. W. (2003). Protein hydrogen exchange mechanism: local fluctuations. *Protein Science* **12**, 153–160.

Maity, H., Maity, M. & Englander, S. W. (2004). How cytochrome *c* folds, and why: submolecular foldon units and their stepwise sequential stabilization. *Journal of Molecular Biology* **343**, 223–233.

Maity, H., Maity, M., Krishna, M. M. G., Mayne, L. & Englander, S. W. (2005). Protein folding: the stepwise assembly of foldon units. *Proceedings of the National Academy of Sciences USA* **102**, 4741–4746.

Maity, H., Rumbley, J. N. & Englander, S. W. (2006). Functional role of a protein foldon: an W-loop foldon controls the alkaline transition in ferricytochrome *c*. *Proteins: Structure, Function, Genetics* **63**, 349–355.

Mayne, L. & Englander, S. W. (2000). Two-state vs. multistate protein unfolding studied by optical melting and hydrogen exchange. *Protein Science* **9**, 1873–1877.

Milne, J. S., Mayne, L., Roder, H., Wand, A. J. & Englander, S. W. (1998). Determinants of protein hydrogen exchange studied in equine cytochrome c. *Protein Science* **7**, 739–745.

Milne, J. S., Xu, Y., Mayne, L. C. & Englander, S. W. (1999). Experimental study of the protein folding landscape: unfolding reactions in cytochrome c. *Journal of Molecular Biology* **290**, 811–822.

Molday, R. S., Englander, S. W. & Kallen, R. G. (1972). Primary structure effects on peptide group hydrogen exchange. *Biochemistry* **11**, 150–158.

Moult, J., Fidelis, K., Kryshtafovych, A., Rost, B., Hubbard, T. & Tramontano, A. (2007). Critical assessment of methods of protein structure prediction – round VII. *Proteins: Structure Function and Bioinformatics* **69**, 3–9.

Myers, J. K., Pace, C. N. & Scholtz, J. M. (1995). Denaturant *m* values and heat capacity changes: relation to changes in accessible surface areas of protein unfolding. *Protein Science* **4**, 2138–2148.

Neudecker, P., Zarrine-Afsar, A., Davidson, A. R. & Kay, L. E. (2007). Phi-value analysis of a three-state protein folding pathway by NMR relaxation dispersion

spectroscopy. *Proceedings of the National Academy of Sciences USA* **104**, 15717–15722.

Nishimura, C., Dyson, H. J. & Wright, P. E. (2006). Identification of native and non-native structure in kinetic folding intermediates of apomyoglobin. *Journal of Molecular Biology* **355**, 139–156.

Oas, T. G. & Kim, P. S. (1988). A peptide model of a protein folding intermediate. *Nature* **336**, 42–48.

Ozkan, S. B., Wu, G. A., Chodera, J. D. & Dill, K. A. (2007). Protein folding by zipping and assembly. *Proceedings of the National Academy of Sciences USA* **104**, 11987–11992.

Pace, C. N. (1975). The stability of globular proteins. *CRC Critical Reviews in Biochemistry* **3**, 1–43.

Panchenko, A. R., Luthey schulten, Z. & Wolynes, P. G. (1996). Foldons, protein structural modules, and exons. *Proceedings of the National Academy of Sciences USA* **93**, 2008–2013.

Pettigrew, G. W. & Moore, G. R. (1987). *Cytochromes c. Biological Aspects*. Berlin Heidelberg, Germany: Springer-Verlag.

Plaxo, K. W., Simons, K. T. & Baker, D. (1998). Contact order, transition state placement and the refolding rates of single domain proteins. *Journal of Molecular Biology* **277**, 985–994.

Pletneva, E. V., Gray, H. B. & Winkler, J. R. (2005). Snapshots of cytochrome *c* folding. *Proceedings of the National Academy of Sciences USA* **102**, 18397–18402.

Plotkin, S. S. & Onuchic, J. N. (2002a). Understanding protein folding with energy landscape theory. Part I: Basic concepts. *Quarterly Reviews of Biophysics* **35**, 111–167.

Plotkin, S. S. & Onuchic, J. N. (2002b). Understanding protein folding with energy landscape theory. Part II: Quantitative concepts. *Quarterly Reviews of Biophysics* **35**, 205–286.

Ptitsyn, O. B. (1995). Molten globule and protein folding. *Advances in Protein Chemistry* **47**, 83–229.

Radford, S. E., Dobson, C. M. & Evans, P. A. (1992). The folding of hen lysozyme involves partially structured intermediates and multiple pathways. *Nature* **358**, 302–307.

Raschke, T. M., Kho, J. & Marqusee, S. (1999). Confirmation of the hierarchical folding of RNase H: a protein engineering study. *Nature Structural Biology* **6**, 825–831.

Raschke, T. M. & Marqusee, S. (1997). The kinetic folding intermediate of ribonuclease H resembles the acid molten globule and partially unfolded molecules detected under native conditions. *Nature Structural Biology* **4**, 298–304.

Roder, H., Elove, G. A. & Englander, S. W. (1988). Structural characterization of folding intermediates in cytochrome *c* by H-exchange labelling and proton NMR. *Nature* **335**, 700–704.

Rosenberg, A. & Chakravarti, K. (1968). Studies of hydrogen exchange in proteins. I. The exchange kinetics of bovine carbonic anhydrase. *Journal of Biological Chemistry* **243**, 5193–5201.

Rumbley, J., Hoang, L., Mayne, L. C. & Englander, S. W. (2001). An amino acid code for protein folding. *Proceedings of the National Academy of Sciences USA* **105**, 105–112.

Schlunegger, M. P., Bennett, M. J. & Eisenberg, D. (1997). Oligomer formation by 3D domain swapping: a model for protein assembly and misassembly. *Advances in Protein Chemistry* **50**, 61–122.

Schmid, F. X. & Baldwin, R. L. (1979). Detection of an early intermediate in the folding of ribonuclease A by protection of amide protons against exchange. *Journal of Molecular Biology* **135**, 199–215.

Silverman, J. A. & Harbury, P. B. (2002a). The equilibrium unfolding pathway of a $(b/a)_8$ barrel. *Journal of Molecular Biology* **324**, 1031–1040.

Silverman, J. A. & Harbury, P. B. (2002b). Rapid mapping of protein structure, interactions, and ligand binding by misincorporation proton-alkyl exchange. *Journal of Biological Chemistry* **277**, 30968–30975.

Sosnick, T. R., Krantz, B. A., Dothager, R. S. & Baxa, M. (2006). Characterizing the protein folding transition state using psi analysis. *Chemical Reviews* **106**, 1862–1876.

Sosnick, T. R., Mayne, L. & Englander, S. W. (1996). Molecular collapse: the rate-limiting step in two-state cytochrome c folding. *Proteins: Structure, Function, Genetics* **24**, 413–426.

Sosnick, T. R., Mayne, L., Hiller, R. & Englander, S. W. (1994). The barriers in protein folding. *Nature Structural Biology* **1**, 149–156.

Spolaore, B., Bermejo, R., Zambonin, M. & Fontana, A. (2001). Protein interactions leading to conformational changes monitored by limited proteolysis: Apo form and fragments of horse cytochrome c. *Biochemistry* **40**, 9460–9468.

Sugase, K., Dyson, H. J. & Wright, P. E. (2007). Mechanism of coupled folding and binding of an intrinsically disordered protein. *Nature* **447**, 1021–1025.

Takei, J., Pei, W., Vu, D. & Bai, Y. (2002). Populating partially unfolded forms by hydrogen exchange-directed protein engineering. *Biochemistry* **41**, 12308–12312.

Udgaonkar, J. B. & Baldwin, R. L. (1988). NMR evidence for an early framework intermediate on the folding pathway of ribonuclease A. *Nature* **335**, 694–699.

Vendrusculo, M., Paci, E., Dobson, C. M. & Karplus, M. (2003). Rare fluctuations of native proteins sampled by equilibrium hydrogen exchange. *Journal of the American Chemical Society* **125**, 15686–15687.

Vu, N. T., Feng, H. & Bai, Y. (2004). The folding pathway of barnase: the rate-limiting transition state and a hidden intermediate under native conditions. *Protein Science* **13**, 220–220.

Wagner, G. & Wüthrich, K. (1982). Amide proton exchange and surface conformation of BPTI in solution: studies with 2D NMR. *Journal of Molecular Biology* **160**, 343–361.

Wallace, L. A. & Matthews, C. R. (2002). Sequential vs. parallel protein-folding mechanisms: experimental tests for complex folding reactions. *Biophysical Chemistry* **101**, 113–131.

Wand, A. J. & Englander, S. W. (1996). Protein complexes studied by NMR spectroscopy. *Current Opinion in Biotechnology* **7**, 403–408.

Wang, L. & Kallenbach, N. R. (1998). Proteolysis as a measure of the free energy difference between cytochrome *c* and its derivatives. *Protein Science* **7**, 2460–2464.

Watson, J. D. & Crick, F. H. C. (1953). Molecular structure of nucleic acids. A structure for deoxyribose nucleic acid. *Nature* **171**, 737–738.

Weinkam, P., Zong, C. & Wolynes, P. G. (2005). A funneled energy landscape for cytochrome *c* directly predicts the sequential folding route inferred from hydrogen exchange experiments. *Proceedings of the National Academy of Sciences USA* **102**, 12401–12406.

Wildegger, G. & Kiefhaber, T. (1997). Three-state model for lysozyme folding: triangular folding mechanism with an energetically trapped intermediate. *Journal of Molecular Biology* **270**, 294–304.

Wolynes, P. G., Onuchic, J. N. & Thirumalai, D. (1995). Navigating the folding routes. *Science* **267**, 1619–1620.

Woodward, C. K. (1994). Hydrogen exchange rates and protein folding. *Current Opinion in Structural Biology* **4**, 112–116.

Woodward, C. K. & Hilton, B. D. (1979). Hydrogen exchange kinetics and internal motions in proteins and nucleic acids. *Annual Review of Biophysics & Bioengineering* **8**, 99–127.

Woodward, C. K., Hilton, B. D. & Tuchsen, E. (1982). Hydrogen exchange and the dynamic structure of proteins. *Molecular and Cellular Biochemistry* **48**, 135–160.

Wu, Y. & Matthews, C. R. (2002). Parallel channels and rate-limiting steps in complex protein folding reactions: prolyl isomerization and the alpha subunit of Trp synthase, a TIM barrel protein. *Journal of Molecular Biology* **323**, 309–325.

Wu, Y. & Matthews, C. R. (2003). Proline replacements and the simplification of the complex, parallel channel folding mechanism for the alpha subunit of Trp synthase, a TIM barrel protein. *Journal of Molecular Biology* **330**, 1131–1144.

Xu, Y., Mayne, L. & Englander, S. W. (1998). Evidence for an unfolding and refolding pathway in cytochrome c. *Nature Structural Biology* **5**, 774–778.

Yan, S., Gawlak, G., Smith, J., Silver, L., Koide, A. & Koide, S. (2004). Conformational heterogeneity of an equilibrium folding intermediate quantified and mapped by scanning mutagenesis. *Journal of Molecular Biology* **338**, 811–825.

Yan, S., Kennedy, S. D. & Koide, S. (2002). Thermodynamic and kinetic exploration of the energy landscape of *Borrelia budgdorferi* OspA by native-state hydrogen exchange. *Journal of Molecular Biology* **323**, 363–375.

Yewdell, J. W. (2005). Serendipity strikes twice: the discovery and rediscovery of defective ribosomal products (DRiPS). *Cell and Molecular Biology* **51**, 635–641.

Zhang, Z. & Smith, D. L. (1993). Determination of amide hydrogen exchange by mass spectrometry: a new tool for protein structure elucidation. *Protein Science* **2**, 522–531.

Zhou, Z., Feng, H. & Bai, Y. (2006). Detection of a hidden folding intermediate in the focal adhesion target domain: implications for its function and folding. *Proteins: Structure, Function, Genetics* **65**, 259–265.

Zimm, G. H. & Bragg, J. K. (1959). Theory of the phase transition between helix and random coil in polypeptide chains. *Journal of Chemical Physics* **31**, 526–535.

Zwanzig, R., Szabo, A. & Bagchi, B. (1992). Levinthal's paradox. *Proceedings of the National Academy of Sciences USA* **89**, 20–22.