doi:10.1016/j.jmb.2004.08.098 J. Mol. Biol. (2004) 343, 1095–1109

JMB

Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER

Protein Misfolding: Optional Barriers, Misfolded Intermediates, and Pathway Heterogeneity

Mallela M. G. Krishna*, Yan Lin and S. Walter Englander

Johnson Research Foundation Department of Biochemistry and Biophysics, University of Pennsylvania School of Medicine, Philadelphia, PA 19104-6059, USA

*Corresponding author

To investigate the character and role of misfolded intermediates in protein folding, a recombinant cytochrome *c* without the normally blocking histidine to heme misligation was studied. Folding remains heterogeneous as in the wild-type protein. Half of the population folds relatively rapidly to the native state in a two-state manner. The other half collapses (fluorescence quenching) and forms a full complement of helix (CD) with the same rate and denaturant dependence as the fast folding fraction but then is blocked and reaches the native structure (695 nm absorbance) much more slowly. The factors that transiently block folding are not intrinsic to the folding process but depend on ambient conditions, including protein aggregation (f(concentration)), N terminus to heme misligation (f(pH)), and proline mis-isomerization (f(U state equilibration time)). The misfolded intermediate populated by the slowly folding fraction was characterized by hydrogen exchange pulse labeling. It is very advanced with all of the native-like elements fairly stably formed but not the final Met80-S to heme iron ligation, similar to a previously studied molten globule form induced by low pH. To complete final native state acquisition, some small back unfolding is required (error repair) but the misfolded intermediate does not revisit the U state before proceeding to N. These properties show that the intermediate is a normal on-pathway form that contains, in addition, adventitious misfolding errors that transiently block its forward progress. Related observations for other proteins (partially misfolded intermediates, pathway heterogeneity) might be similarly explained in terms of the optional insertion of error-dependent barriers into a classical folding pathway.

Keywords: folding barriers; hydrogen exchange; pulse labeling; cytochrome *c*; EX1

Introduction

What is the role of intermediates and barriers in protein folding? Early protein folding experiments were commonly interpreted in terms of classical pathways formed by constructive, on-pathway, native-like intermediates,[1–4] as represented in Figure 1(a). An alternative multi-dimensional energy landscape view, often conceptualized as in Figure 1(b), is based on statistical–mechanical ideas and computer simulations used to explore the energy landscape. The common perception of these simulations is that proteins fold through many parallel trajectories at rates determined by the slope (protein stability) and roughness of the energy landscape and the trapping of misfolded unproductive intermediates, although many theoreticians hold a more nuanced view.[5–12]

The effort to resolve these central issues by experimental studies has proven exceedingly difficult. One finds a variety of contradictory behaviors. Some proteins fold in a two-state manner without obvious intermediate formation. Others fold in a multi-state manner with the transient accumulation of intermediates, which often seem to be misfolded in some way. Many proteins fold heterogeneously with both two-state and multi-state fractions. These

Abbreviations used: Cyt *c*, cytochrome *c*; WT, wild-type Cyt *c*; pWT, pseudo-wild-type recombinant equine Cyt *c* (H26N, H33N; N terminus not blocked); HX, hydrogen exchange; foldon, cooperative folding/unfolding unit; PUF, partially unfolded form; CD, circular dichroism; GdmCl, guanidinium chloride.

E-mail address of the corresponding author: mmg@hx2.med.upenn.edu

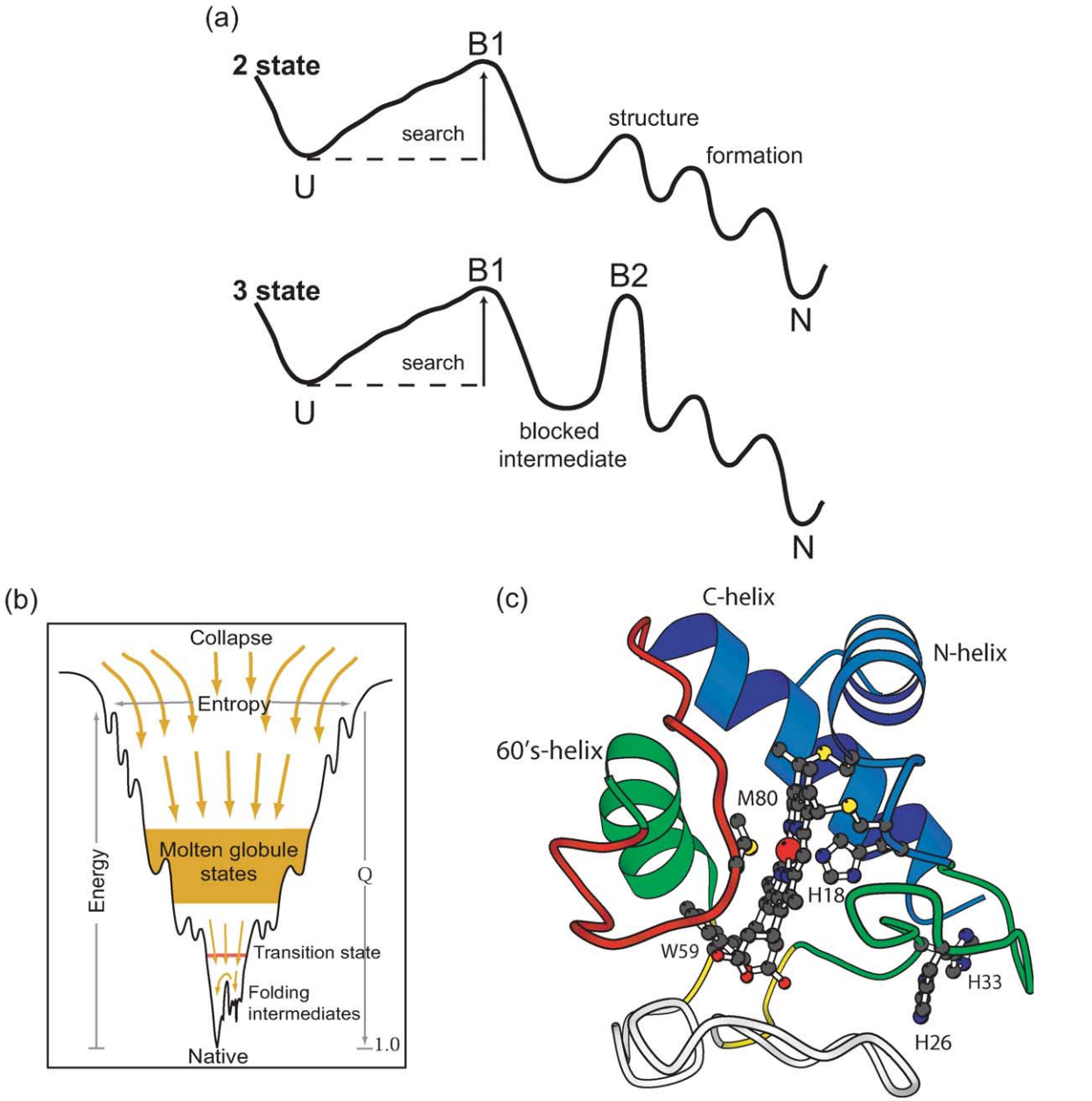

Figure 1. (a) A modified classical pathway view. Folding often appears to be two-state (upper) because it is rate-limited by an initial large scale search barrier (B1).[21,91] The chance insertion of a misfolding error (B2) (lower) can slow some population fraction and transiently populate the preceding intermediate, producing what appears to be a misfolded intermediate, alternative parallel pathways, and asynchronous intermediate formation. (b) Standard representation of a multi-dimensional funnel-shaped energy landscape.[5,7] (c) Cyt *c* structure showing residues important in the present work (1HRC.pdb[98] and MOLSCRIPT[99]). The Cyt *c* diagram is color-coded to indicate the folding units previously identified by HX experiments and ranked in spectral order of decreasing ΔG_{op} value.[24,35,61,63,66,67,69,100] The five foldons are the blue bihelical unit (N and C-terminal helices), the green unit (60's helix plus the 20's–30's Ω-loop), the outer yellow unit (a short antiparallel β-sheet; 37–39, 58–61), the nested-yellow Ω-loop (40–57), and the red Ω-loop (71–85). The blue foldon forms in an initial step, followed by the templated stepwise formation of the other structural elements, leading to the native protein.

observations have been variously interpreted to indicate that distinctly structured intermediates do or do not exist, that populated intermediates are constructive on-pathway or obstructive off-pathway forms, and that proteins fold through single linear paths or multiple parallel paths. The spectroscopic probes widely used to observe fast kinetic folding events do not provide the detailed structural information necessary to clearly distinguish native-like from non-native-like intermediates and the relationships that connect kinetically different pathways.

It first became possible to define kinetic folding intermediates in some structural detail with the development of the hydrogen exchange (HX) pulse labeling experiment.[2,13–15] Intermediates that

accumulate in the three-state folding of ribonuclease A[14] (RNase A) and cytochrome *c*[15] (Cyt *c*) appeared to be distinct partially folded native-like forms, matching the classical expectation. However, intermediate accumulation depends on proline misisomerization in RNase A[16] and histidine to heme misligation in Cyt *c*[17–20] (see Figure 1(c)). Sosnick *et al*.[18,21] pointed out that Cyt *c* and other proteins are also able to fold in a much faster two-state way, without the population of intermediates, suggesting that intermediate accumulation results from the insertion of optional barriers into a normal linear folding pathway (Figure 1(a), lower).

An influential review article[22] cast a different light on the situation. The populated Cyt *c* intermediate is misfolded. Cyt *c* folding can be heterogeneous; some population fraction folds rapidly in an apparent two-state manner while another forms the misfolded intermediate and is transiently blocked. Intermediate formation itself is asynchronous; different population fractions form given structures at different rates. Baldwin[22] pointed out that these behaviors matched predictions of the newly emerging theoretically based view of protein folding (Figure 1(b)), which pictures visibly populated intermediates as misfolded off-pathway forms that act to slow rather than promote folding. Accordingly the Cyt *c* result has been widely cited as showing that populated intermediates are misfolded off-pathway forms. Results for many other proteins have been similarly interpreted.[23]

The kinetic intermediate populated by wild-type (WT) Cyt *c* at neutral pH has more recently been characterized in detail by an improved HX pulse labeling technique.[24] It strikingly resembles an emerging native-like structure. Native-like N-terminal and C-terminal helices are fully formed and docked together in a native-like way. The rest of the protein is not yet protected except for some minimally stable native-like β-turns. It is true, however, that the intermediate is misfolded in the sense that a histidine to heme misligation dislocates and traps the not yet folded histidine-containing segment (green loop in Figure 1(c)). Also, folding is heterogeneous; a fraction of the population misligates, accumulates the intermediate, and folds slowly whereas another fraction folds rapidly without apparent intermediates. Nevertheless, all of the Cyt *c* molecules pass through this same native-like N/C bi-helical intermediate structure whether they fold in a fast two-state manner or are transiently blocked by the misligation error and fold more slowly.[24] These properties are consistent with a constructive intermediate on a single discrete pathway, with some probability of being transiently blocked by an optionally inserted error-dependent barrier, as shown in Figure 1(a). However, other interpretations (parallel pathways, off-pathway intermediates) are possible.

To understand these issues more deeply, one wants to study the detailed structure of other folding intermediates and how they function to construct folding pathways. Here we used a recombinant pseudo-wild-type (pWT) equine Cyt *c* lacking the two peripheral histidine residues (His26 and His33) that cause misfolding and intermediate accumulation in WT Cyt *c*[25] (Figure 1(c)). Surprisingly, folding remains heterogeneous. About half of the population folds rapidly; the other half folds slowly and transiently accumulates an observable intermediate. The pathway behavior and the structure and properties of the intermediate were characterized in some detail. The results portray a very advanced native-like intermediate, demonstrate its on-pathway nature, and define the misfolding errors that cause it to accumulate. This adds to a great deal of other information on the folding behavior of Cyt *c*, all of which consistently points to a classical stepwise folding pathway that can be interrupted by optionally inserted errors.

Results

Kinetic folding

Cyt *c* is one of the most highly studied folding models.[17–21,26–32] We studied a recombinant pWT equine Cyt *c* in which the two histidine residues (His26 and His33) that cause misfolding in the wild-type protein have been changed to asparagine. Its expression and purification were described.[25] The native structure of recombinant pWT Cyt *c* is identical with the WT protein as indicated by identical chemical shifts in 2D NMR spectra,[25,33] except for small changes in the immediate vicinity of the His to Asn mutations and at the N terminus, which is acetylated in the WT but not in the pWT protein. Also the pWT protein shows the same foldon substructure determined by native state HX as found for WT Cyt *c*.[34,35]

For kinetic refolding experiments, pWT Cyt *c* was initially unfolded in concentrated denaturant in ^2H_2O at pD_{read} 7.5. Folding was initiated by a large dilution into H_2O buffer at pH 6, and monitored by several spectral probes. Tryptophan fluorescence (Figure 2(a); relative to U) is strongly quenched when the single Trp59 is brought close to the heme, either upon native state formation or some prior molecular collapse. CD_{222} (Figure 2(b)) tracks formation of the three major and two minor helices. Absorbance at 695 nm (Figure 2(c)) registers the ligation of the Met80-S to the heme iron, which provides a marker for final native state acquisition.

Almost all of the expected fluorescence quenching and helical content is recovered in a faster phase ($\tau \sim 27$ ms) but only a fraction of the population (39%) reaches the native state at this rate (695 nm signal), folding in a two-state way with no apparent intermediate accumulation. The rest of the population folds to an intermediate at this same fast rate (Figure 2(a) and (b)) but then is blocked and reaches the native state more slowly with multi-exponential

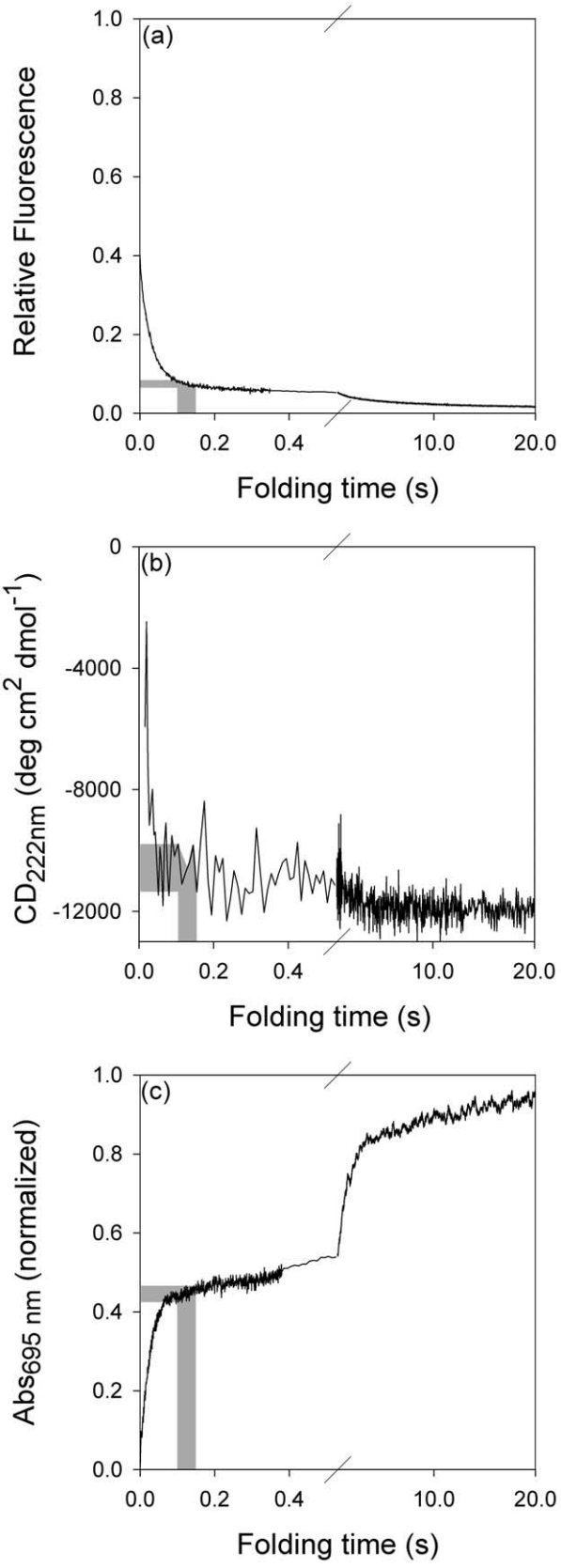

Figure 2. Multi-phasic stopped-flow folding kinetics of pWT Cyt *c* followed by three spectroscopic probes. Unfolded Cyt *c* (4.2 M GdmCl and pD_{read} 7.5) was diluted into the folding buffer (pH 6.0 at 10 °C) to a final GdmCl concentration of 0.23 M GdmCl. Refolding was measured by (a) Trp59 fluorescence (chain contraction/collapse), (b) circular dichroism at 222 nm (helix), and (c) absorbance at the 695 nm Met80-S to heme Fe charge transfer band (native state probe). The gray shading shows where the HX labeling pulse was applied in pulse labeling experiments.

kinetics (695 nm absorbance, Figure 2(c); fit by τ of 1.1 seconds (42%) and 17 seconds (19%)).

Spectroscopic probes (Figure 2) do not reveal what specific structural distortion in the kinetic intermediate blocks formation of the Met80-S to heme iron ligation (695 nm absorbance). The ligation is very weak and requires essentially the fully stable native structure of Cyt to hold it in place. For example, the 695 nm signal is lost at low pH even though the fluorescence and CD signals are not affected (Figure 3). These same properties, retention of the helical content and loss of the 695 nm band, mark both the mildly destabilized acid Cyt *c* molten globule characterized at equilibrium[36,37] and the intermediate characterized here during kinetic folding.

Blocking factors

Several extrinsic factors act to slow folding. The slowest folding phase seen by 695 nm absorbance (τ = 17 seconds, 19% population) depends on the mis-isomerization of Pro76 (near the Met80 ligand). This was shown before in mutational[38] and double jump[39] studies and was checked for the present case by us in double jump experiments (data not shown). This factor adds complexity to folding measurements but may not significantly affect folding itself. The dominant factor is protein aggregation, demonstrated by the decrease of the amplitude of the slow folding phase with protein concentration (Figure 4(a)). Aggregation behavior was found before for WT Cyt *c*.[40,41] The slow folding fraction due to aggregation can be completely removed at lower protein and higher GdmCl concentration.[40] One does not know what part of the protein is involved in the intermolecular association. Some misligation of the N terminus to the heme, seen before by Hammack *et al.*,[42] is also present. It increases the slow phase amplitude. At low pH where protonation of the free N-terminal amino group inhibits its misligation to the heme iron (pK_a for misligation ~6),[42] a fraction of the slow phase is moved into the faster phase (Figure 4(a)). An analogous effect is seen for the WT protein, dependent on the titration of histidine residues.[21] The fast phase rate itself is modestly accelerated, from 27 ms to 12 ms, when the N-terminal amino group is protonated at low pH and cannot misligate. The same faster rate is found for WT equine Cyt *c* when the histidine to heme misligation is removed at low pH[18,21,24] and for recombinant yeast Cyt *c* when N terminus misligation is chemically blocked by glyoxylation.[42] This is the rate in the chevron rollover region at low denaturant which has been attributed to intermolecular

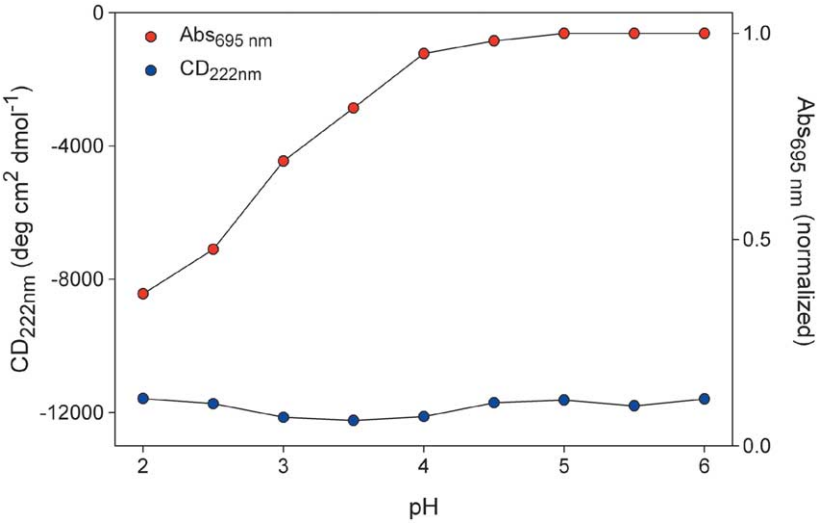

Figure 3. Sensitivity of the 695 nm absorbance probe for the final stable native state. Under mildly destabilizing conditions, the weak Met80-S to heme iron bond is lost with no effect on CD; for example at low pH and high salt in the acid molten globule as shown, as well as in the advanced kinetic folding intermediate studied here.

association.[43] These observations suggest a dominant role for some generalized aggregation, especially at the high concentration used for the 695 nm and pulse labeling experiments.

These blocking effects can be considered in terms of the insertion of adventitious error-dependent barriers into a classical folding pathway[18] (Figure 1(a)). One possible test for an error repair barrier is provided by the effect of denaturant on folding rate. Increasing denaturant normally slows protein folding because denaturant disfavors surface burial. When folding is blocked by an error-repair barrier, the folding rate may show a reverse denaturant dependence, indicating that the rate is promoted by some back-unfolding or disaggregation step[18,21,24,44,45] (Figure 4(b)). The small denaturant dependence ($m = 0.4(\pm 0.1)$ kcal/mol per molar GdmCl compared to $m_{U-N} = 4.5$ kcal/mol per molar GdmCl) and especially the rate relationships observed ($k_{IN} > 10\ k_{IU}$; $k_{IN} = 0.91\ s^{-1}$; $k_{IU} < 0.09\ s^{-1}$ calculated from pulse labeling measurements described below) show[44] that the intermediate does not unfold to U before reaching N. Reverse denaturant effects have been seen for other proteins.[46–48]

The present study should especially focus attention on the under-appreciated role of transient protein aggregation as a cause for slow folding and multi-state kinetics. A number of prior observations for other proteins support the same conclusion, including the U1A protein,[49,50] CI2,[50] bovine growth harmone,[51] interleukin 1β,[52] lysozyme,[53] α-lactalbumin,[54] phosphoglycerate kinase,[55] myoglobin,[56] barnase[57,58] and ubiquitin.[59]

The pulse labeling experiment

HX pulse labeling experiments were done to characterize the trapped intermediate populated by the slow folding fraction. Unfolded Cyt *c* initially deuterated in ^2H_2O was diluted into H_2O buffer (pH 6) and allowed to fold for 100 ms. At this time almost all of the native-like CD and fluorescence quenching have been recovered, 42% of the population has reached the native state, and the slower misfolded intermediate is maximally populated (Figure 2(a)–(c)). During this period very little ^2H to H exchange occurs ($\tau_{HX} \sim 3$ seconds for freely exposed amides at pH 6 and 10 °C).

A labeling pulse to higher pH was applied for 50 ms (gray shading in Figure 2), then terminated in a third mix by dropping to pH 5.3 and reducing the heme iron (ascorbate). The pulse-labeled intermediate completes folding within one minute. The stable native structure in the heme-reduced form together with the low pH makes further exchange very slow, preserving the H–^{2}H labeling profile imprinted during the sub-second labeling pulse. Samples were concentrated (centrifugal filtration) and the H–^{2}H exchange profile was analyzed by 2D NMR. Control experiments calibrated the uptake of H-label by the native fraction present during the pulse and by all forms of the protein at times other than the pulse (especially by the native form during later sample workup).

Pulse labeling was done at pH values from pH 7.5 to 10. In this range, the intermediate population trapped due to proline mis-isomerization, aggregation and/or N terminus misligation ($pK_a \sim 6$) does not depend on pH. The data points in Figure 5 show the H-labeling measured for 50 amides that could be followed with good accuracy over the whole pH range (H_{obs} in equation (2)). The color-coding identifies the different sub-global foldon units found in previous experiments (Figure 1(c)). The continuous colored curves in Figure 5 show the fit to the data of the equations given in Materials and Methods (equations (3), (4) and (7)–(10)), which provides the residue-resolved fitting parameters k_{op} and k_{cl}, and thus K_{op} and ΔG_{op}. The broken curves in Figure 5 compare the labeling expected for each amide if it were wholly unprotected (black; equation (6)) and the corrected labeling computed for the intermediate alone (colored; equation (5)).

At low pulse pH, exchange is in the pre-equilibrium-opening EX2 mode ($k_{ch} < k_{cl}$; $k_{ex} = K_{op}k_{ch}$). Accordingly, the horizontal displacement of the

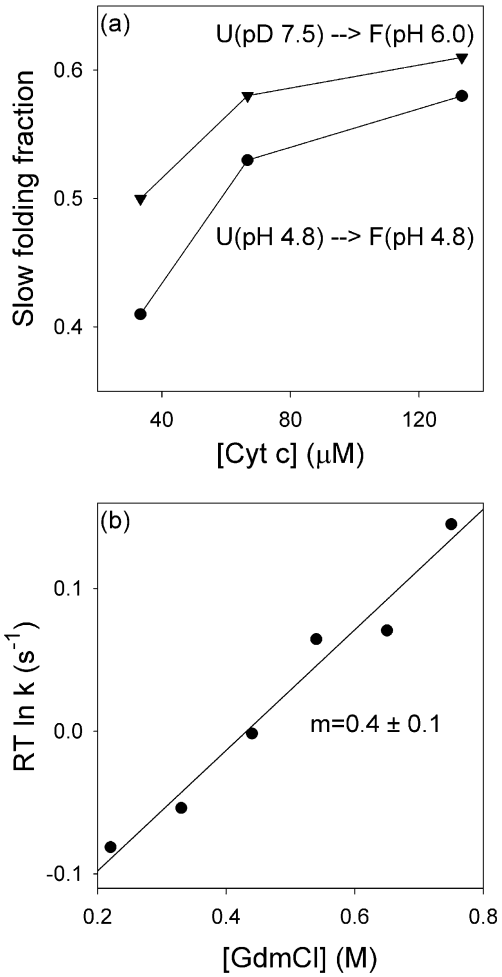

Figure 4. Optional barriers in Cyt *c* folding monitored by 695 nm absorbance. (a) Slow folding molecules move into the faster two-state phase when the protein concentration is reduced (aggregation) and at lower pH (N-terminal to heme misligation). The slowest phase (17 seconds) is removed in double-jump experiments (proline misisomerization). (b) A small but definite reverse denaturant effect on the slow phase folding rate, indicating that the barrier process limiting the slow phase rate includes some unfolding and/or disaggregation (error repair).

colored curves from the black curves reveals the stability of protecting structure. At higher pulse pH, labeling reaches a pH-independent plateau (EX1 region; $k_{ch} > k_{cl}$) where the measured HX labeling rate becomes equal to the rate for transient structural opening ($k_{ex} = k_{op}$). When the pulse time is appropriate the plateau level sensitively indicates the k_{op} parameter. A detailed illustration of these relationships is available.[60]

The intermediate: identity, stability, folding and unfolding rates

The residue-resolved equilibrium stability obtained for the trapped intermediate and kinetic opening and closing rates are displayed in Figure 6.

In the early intermediate of WT Cyt *c* studied before by the same methods,[24] only the N and C helices and a few other residues were protected, all by less than 1.5 kcal/mol. The intermediate studied here is much more advanced, nearly native. All of the measurable amide hydrogen atoms in all of the helical and Ω loop elements are fairly stably protected (H-bonded). For most of the amides in the three major helices (N, C and 60's helix), no significant labeling was observed over the pulse pH range studied. Therefore, one can only obtain lower limits for their residue-resolved stability. The stability measurable in the frayed terminal residues is 3.4 kcal/mol in the blue bi-helical unit and 2.1 kcal/mol in the green helix (color-coded horizontal lines and upward arrow heads in Figure 6(a)). Most of the residues in the three Ω loops show an average ΔG_{op} of 1.0(±0.5) kcal/mol. For comparison, the blue and green helices in the native protein have a ΔG_{HX} value of 12.6 kcal/mol and 9.8 kcal/mol, respectively, and the Ω loops between 5.0 kcal/mol and 7.5 kcal/mol. Although the global stability of the trapped intermediate could not be quantitatively measured, it must be decreased by approximately 6 kcal/mol, since previous work shows that stabilization[61] or destabilization[35] of the lower foldons carries through to the global stability.

Stability of the foldons within the intermediate is ordered according to their native state stabilities. The measurable kinetic unfolding rates of the three Ω-loops (Figure 6(b)) appear to be similarly ordered although the scale is compressed, presumably because stability is low and each unfolding step is promoted by the previous one. The opening rate constant, k_{op}, is about 35 s^{-1} for the lowest lying nested-yellow loop, 20 s^{-1} in the red loop and outer yellow neck, and 10 s^{-1} in the green loop. The value of k_{op} is given directly by the plateau level of the HX rate at high pulse pH in Figure 5. The values for k_{cl} (Figure 6(c)) show a more dispersed range, as seen previously in other EX1 HX studies.[24,60,62–64] The calculation for k_{cl} depends on the less sensitive rising portion of the HX labeling curve (EX2 region) and on the HX rates experienced by amides when they transiently "open" which may not be well represented by the values calibrated in small dipeptide models.[65]

In summary, the intermediate has essentially all of the native-like structural elements more or less in place. Equilibrium protection and kinetic parameters follow the same order, blue, green and then the less stable Ω-loops, as found before in multiple studies of the stability and kinetic ordering of foldon units in WT and pWT Cyt c[24,34,35,44,61,63,66–70] (see Discussion). The intermediate is an on-pathway form, transiently trapped in mid-flight by several non-native factors, and differs from the final native state by its greatly reduced stability, faster foldon unfolding rates, and the absence of the Met80-S to heme ligation.

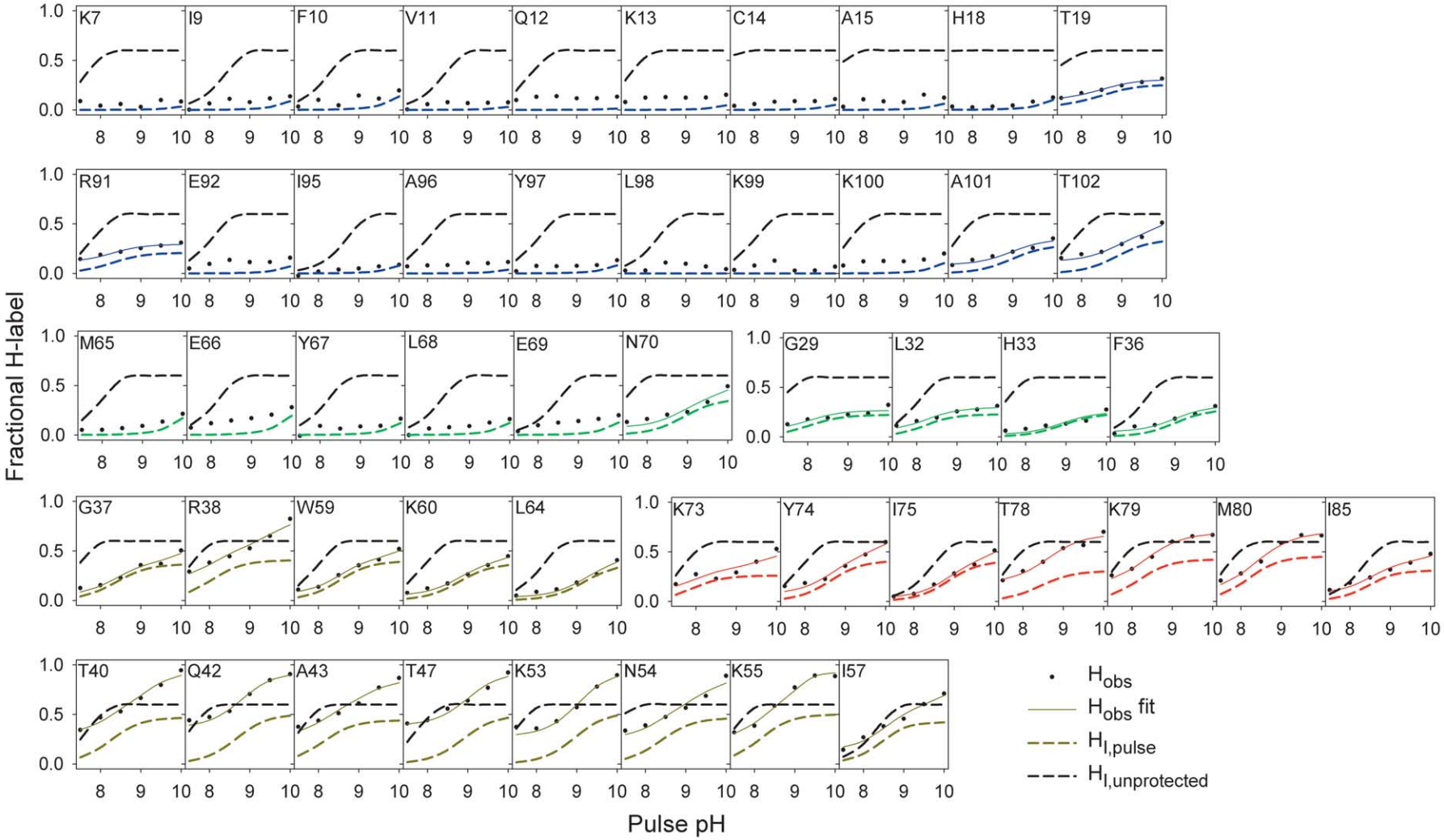

Figure 5. Residue-resolved H-labeling results for the trapped intermediate. The color-coding identifies amides in the foldon units shown in Figure 1(c). Data points show the observed pH-dependent labeling (H_{obs} in Materials and Methods; equation (2)). The continuous lines fit the equations in Materials and Methods (equations (3), (4) and (8)–(10)) to the measured H-labeling data (two free parameters, k_{op} and k_{cl}). The black broken curves show the labeling that would be expected for each amide if it were wholly unprotected in the intermediate ($[1-f_N]H_{I,unprotected}$; equation (6)), calculated from previous calibrations[65,94] (http://hx2.med.upenn.edu/download.html) and scaled to 0.58, the fraction of the protein population present in the misfolded blocked intermediate state at the time of the pulse. The colored broken curves compare the labeling actually obtained within the trapped intermediate alone ($[1-f_N]H_{I,pulse}$; equation (5)) after the measured data (H_{obs}) are corrected for extraneous contributions (H_{bkgd} and $H_{N,pulse}$). The fitting of the equations to the original data implicitly includes these corrections. For residues that show incomplete labeling, only the curve for $(1-f_N)H_{I,pulse}$ is shown.

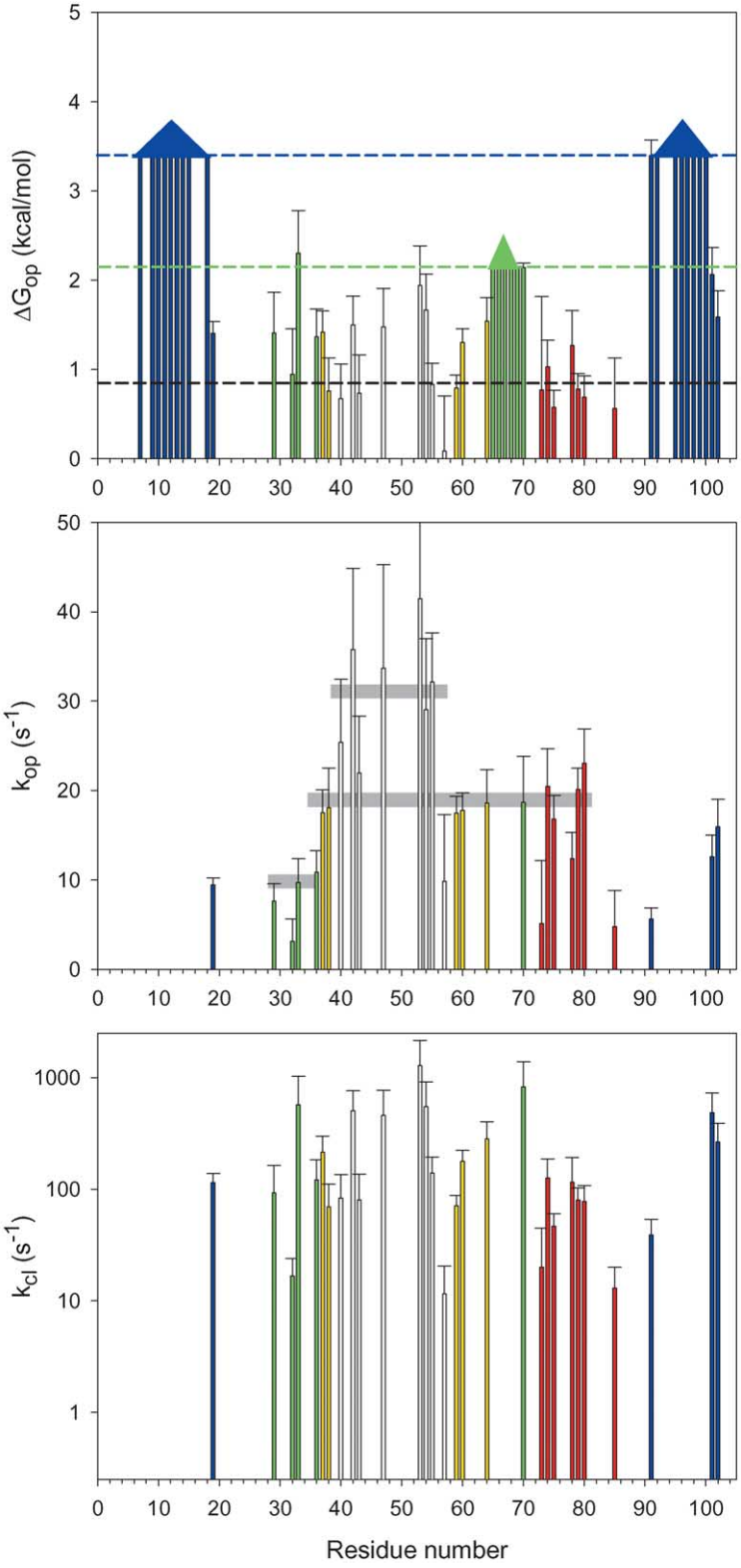

Figure 6. Residue-resolved equilibrium and kinetic parameters in the trapped intermediate obtained by fitting the data in Figure 5. ΔG_{op} in (a) is the residue-resolved stabilization free energy of the unfolding reaction that allows exchange within the trapped intermediate. Blue and green broken horizontal lines and upward arrowheads indicate the lowest limits for the inner residues in the blue and green helices. The black broken line represents the average ΔG_{op} value observed over the three Ω-loops. (b) and (c) Residue-resolved rates for opening (unfolding) and closing (folding). The horizontal bars in (b) indicate the approximately constant k_{op} values measured for individual Ω-loops.

Discussion

In spite of a great deal of work on protein folding, the character and role of misfolded intermediates and heterogeneous pathways remains unclear. These uncertainties have led to very different suggestions about how proteins might fold, ranging from a classical stepwise pathway with distinct intermediates (Figure 1(a)) to multiple indefinite trajectories in which intermediates are ill-defined and obstructive (Figure 1(b)). The following sections briefly summarize the information now

available for the folding behavior of WT and pWT Cyt *c* and then consider the implications for protein folding behavior more generally.

Foldon intermediates in a stepwise pathway

Thermodynamic principles require that each high energy state that is present in the space between U and N must be occupied according to its equilibrium free energy level, and that all of the molecules must cycle continually, unfolding and refolding, through all available states. In favorable cases native state HX experiments can measure this behavior. These experiments have revealed that the Cyt *c* protein is composed of five concerted folding units, called foldons, shown in Figure 1(c). The foldons unfold and refold as concerted units forming a number of high energy partially unfolded forms (PUFs).[60,71,72] The blue bi-helical unit unfolds with the highest free energy. It marks the transient global unfolding.[66,73] At lower free energy are, in order, the green helix plus loop,[66] then the short yellow neck,[66,69] the red loop,[63,66] and the nested-yellow loop.[69] The same behavior is seen for the pWT protein used here.[34,35]

A great deal of work shows that the foldons unfold and refold in a classical sequential pathway manner as in equation (1), where the letters relate to the color code used in Figure 1(c) (upper case folded; lower case unfolded). The residue-resolved identity of each of the individual intermediate forms has been determined in equilibrium[66,67,69] and kinetic[24,63] modes:

$$\text{NRYGB (native)} \leftrightarrow \text{nRYGB} \leftrightarrow \text{nrYGB} \leftrightarrow \text{nryGB} \leftrightarrow \text{nrygB} \leftrightarrow \text{nrygb (unfolded)} \quad (1)$$

HX pulse labeling experiments done during kinetic folding show that the blue bihelical unit is the first to form.[15,24] Kinetic native state HX experiments demonstrate that the red unit is one of the first to unfold and that further unfolding then proceeds sequentially in the order shown.[63,69,70] The subsequent identification of the nested-yellow foldon placed it even earlier in the unfolding sequence[69] (although an alternative, not wholly ruled out, is that the nested-yellow and red units may unfold as alternative initial steps and then join to form nrYGB[69]). Stability labeling HX experiments[35,61] show that each successive higher-energy PUF includes the unfolding of all of the lower energy foldons but not the higher ones, independently documenting their sequential unfolding nature. Since the native state HX experiments are done under equilibrium native conditions, each unfolding reaction actually measured by HX must be matched by an equal and opposite refolding reaction, as in equation (1), else equilibrium would not be maintained. Finally, it seems compelling that the unfolding–refolding pathway sequence matches the way that the different foldons interact in the native context.[68] In native Cyt *c* the initially formed blue unit only interacts with the two green segments, therefore can only guide and stabilize their formation, and similarly down through the folding sequence in a sequential stabilization process (Figure 1(c)).

In summary, multiple independent equilibrium and kinetic experiments consistently indicate that Cyt *c* comprises five foldon units, that these units form a stepwise unfolding–refolding pathway under fully native conditions (equation (1); Figure 1(a) and (c)), and that all of the molecules pass through all of the same structurally native-like intermediates. The supposed pathway recapitulates the way that the foldons are organized in the native protein. All of these conclusions are based on detailed site-resolved structural data. It seems very difficult to explain these results by any other model.

Analogous experiments have uncovered analogous foldon substructures and led to similar conclusions concerning folding pathways for other proteins,[71,72] including apomyoglobin,[74–77] RNase H,[78–82] apo Cyt *b*562,[83–87] OspA,[88] and triose phosphate isomerase.[89]

The barriers: initial search-collapse-nucleation (B1) and later error repair (B2)

WT Cyt *c* can be made to fold in either a two-state or a three-state manner, or a heterogeneous mixture of the two, even when refolding is done under the very same ambient conditions.[18,21] The controlling condition is the pH value of the initial unfolding solution. When the initially unfolded protein is held at low pH where histidine residues are charged and cannot misligate to the heme, subsequent refolding appears largely two-state. At higher pH, the neutral peripheral histidine residues can misligate to the exposed heme iron.[18,19] Refolding is then slow and three-state; an intermediate with the N and C helices formed (plus the histidine misligation) transiently accumulates. When the initially unfolded protein is held at intermediate pH, subsequent refolding is heterogeneous; some of the Cyt *c* molecules encounter the misligation-dependent barrier and others do not. In fact, the misfolding need not pre-exist in the unfolded state. These very same results are equally obtained when the initially unfolded protein is held at low pH and then mixed into refolding conditions at various pH values.[18]

Under two-state folding conditions, the rate-determining step includes a major chain condensation that buries a great deal of surface (large dependence of rate on denaturant concentration), necessarily including many polar groups. This behavior is characteristic for proteins in general.[90] Sosnick *et al.*[21] suggested that the rate-limiting barrier represents an initial conformational search to find a collapsed transition state that mimics the native-like topology (B1 in Figure 1(a)). This view seems to be validated by the success of the contact order correlation found by Plaxco and co-workers[91,92] and similar correlations described by others, all of

which point to the conclusion that two-state folding is rate-limited by an initial whole molecule conformational search to find the native-like topology. When B1 is limiting, folding appears to be two-state because subsequently formed pathway intermediates are invisible. They form after the initial rate-limiting step and depend on smaller, therefore faster searches (Figure 1(a), upper).

Under three-state folding conditions, a new later barrier intervenes, causing an intermediate to accumulate (Figure 1(a), lower). The WT Cyt *c* intermediate, characterized in detail by HX pulse labeling, has the N and C-terminal helices fully formed and little else.[24] The barrier that determines the rate for formation of the N/C intermediate in three-state folding is identical with the barrier that rate-limits two-state folding (B1). They have the same thermodynamic ($\Delta G^{\ddagger}$, $\Delta H^{\ddagger}$, $\Delta S^{\ddagger}$) and structural ($m^{\ddagger}$) parameters.[21] The rate-limiting two-state barrier is obviously on-pathway, therefore the N/C bi-helical intermediate that it leads to is also. In fact the N and C helices appear to form in the B1 transition state.[90,93] Another conclusion is that all of the refolding molecules traverse the same B1 barrier and all pass through the same N/C bi-helical intermediate whether kinetic folding appears to be two-state or multi-state.[24]

When the N/C intermediate accumulates and becomes visible in three-state kinetic folding, it contains in addition a non-native misligation ("misfolding"). However, the intermediate does not depend on the misfolding; the very same N/C bi-helical form is seen as an unblocked intermediate by native state HX. The misfolding acts as if it inserts a new kinetic barrier (B2) into the folding sequence, which transiently blocks the next step and causes the pre-existing N/C intermediate to accumulate. The next step, formation of the green unit (equation (1) and Figure 1(c)) is blocked because the peripheral histidine residues that can misligate to the heme are placed in the green loop. The misligation holds the green loop out of place (see Figure 1(a) and (c)), unable to take its turn in the folding sequence (see equation (1)) until the error is repaired.

These results illustrate two kinds of rate-determining kinetic barriers, an initial search-collapse-nucleation[21] barrier (B1) that is intrinsic to the folding process and an error-dependent barrier (B2) that is optional rather than intrinsic.[18]

Alternative parallel pathways *versus* alternative optional barriers

To further study the character and role of misfolded kinetic intermediates, the present experiments used a recombinant pWT Cyt *c* protein (H26N, H33N) in which the previously studied histidine to heme misligation error is eliminated. The results obtained are parallel to the WT Cyt *c* results just reviewed.

In the kinetic folding of pWT and WT Cyt *c*, molecular collapse (Figure 2(a)) and helix formation (Figure 2(b)) occur in the same faster phase for all of the protein population. The two-state population and the faster fraction of the three-state population show this same rate and other properties, evidently determined by the same initial barrier (B1). B1 appears to represent an initial conformational search that is intrinsic to the folding process.[21]

About half of the population proceeds to the native state in a two-state manner at the B1 rate. The other half, after passing B1, encounters some later block (B2) and folds to the native state more slowly in a multi-state manner (Figure 2(c)). This behavior makes folding appear to be heterogeneous.

The factors that impose the later block and produce multi-state behavior are all optional (protein aggregation, N-terminal misligation, proline mis-isomerization). They can be removed by altering external conditions (protein concentration, solution pH, the time spent in the unfolded state before folding is initiated). These factors and the B2 blockage that they produce are clearly not intrinsic to the folding process.

The late intermediate that then accumulates in the multi-state pWT fraction is "misfolded" in some way, corrupted by one or more of the extrinsic blocking factors. Still, it's structure is exceedingly native-like, it forms at the same rate as the two-state fraction, and it does not unfold to the U state before progressing to N. Thus, the structured part of the intermediate appears to represent an obligatory form intrinsic to the normal folding pathway, although the version seen to accumulate in multi-state folding necessarily contains in addition one or more slowly repaired errors.

These observations can all be understood in terms of obligatory intermediates and barriers experienced by all of the molecular population moving through a common folding pathway, which can however be interrupted by optionally encountered errors. In these terms, the time-consuming error-repair process seems to be well represented as an optionally inserted barrier (B2).

Conclusions

Heterogeneous folding behavior together with the accumulation of multiple alternative intermediates has been seen for many proteins and simulated folding models. In the absence of detailed structural information, these observations have been widely interpreted in terms of multiple optional intermediates and multiple fundamentally different trajectories spread through different regions of a folding landscape like that diagrammed in Figure 1(b). Detailed structural information now available for Cyt *c* and some other proteins suggests a different view.

A great deal of information supports the conclusion that Cyt *c* is made up of five concerted folding units. As a result, it folds through a distinct set of intermediate structures, apparently over a distinct set of barriers, in a classical

pathway that progressively constructs the native protein.[15,18,21,24,35,61,63,66,67,69] The pathway is essentially linear because of the way that the different foldons are arranged in the native protein.[68]

It appears that the entire population passes through the same sequence of intermediates whether the measured folding kinetics appear to be two-state or multi-state. In two-state kinetic folding, the Cyt *c* intermediates are invisible because the initial barrier (B1) is rate-limiting.[21,91,92] In three-state folding, the entire protein population traverses the same earlier barriers (this work and Sosnick *et al*.[21]) and intermediates,[15,24] but some fraction then encounters a newly inserted error-dependent barrier (B2). Folding pauses and the corrupted intermediate accumulates (this work and Sosnick *et al*.[18]). External or internal factors can determine the error-dependent barrier that is inserted, therefore the intermediate that populates, and the rate of final folding. Because B2 is optional, not intrinsic to the folding process, the probability of encountering a folding error can be zero (two-state) or unity (multi-state) or in between (heterogeneous folding).

We suggest that similar effects may explain similar behavior in other proteins. A protein population that folds through a single pathway but encounters probabilistic errors misleadingly appears to experience alternative parallel pathways. Known condition-dependent errors in protein folding include proline mis-isomerization, heme misligation, alternative disulfide bond formation, alternative domain docking modes, formation of non-native interactions, and aggregation.[23,24] The thesis of this paper is that all of these factors might be appropriately viewed in terms of the optional insertion of non-intrinsic barriers into an otherwise classical folding pathway, as shown in Figure 1(a).

Materials and Methods

Recombinant pseudo-wild-type equine Cyt *c* (pWT) was expressed and purified as described.[25] It differs from the wild-type protein in that the two native histidine residues are mutated to asparagine (H26N,H33N) and the N terminus is not blocked. The structure of the pWT protein is closely similar to the highly studied parent equine WT Cyt *c*, as indicated by NMR chemical shifts which differ only in the immediate vicinity of the covalent modifications.[25,33]

The pH buffers used were the highest grade available from Sigma. ^2H_2O was from Isotec or Aldrich. Deuterated GdmCl (d-GdmCl) was prepared by dissolving in ^2H_2O and vacuum evaporating three times. Folding kinetics were measured at 10 °C using stopped-flow Biologic SFM-400 apparatus (695 nm absorbance and fluorescence) or an Aviv stopped-flow module attached to an AVIV CD spectrometer (model 202). Buffers, pH values, and solvents were the same as those in the HX pulse labeling experiments.

HX pulse labeling and data analysis were as described.[24,60] Cyt *c* was unfolded in 4.2 M d-GdmCl in 10 mM phosphate, pD_{read} 7.5 ^2H_2O buffer. Refolding was initiated by diluting 18 times into the folding buffer in H_2O (pH 6, 10 mM Mes, 0.23 M GdmCl). Protein concentration during folding was about 130 μM. Folding was continued for 100 ms. A high pH pulse (pH 7.5 to 10) was then applied for 50 ms. Pulse buffers were (after mixing): 50 mM Hepes (pH 7.5), Epps (pH 8), Bicine (pH 8.5), Ches (pH 9 and 9.5) and Caps (pH 10) in H_2O. HX labeling was stopped by mixing with the quench buffer (63 mM citrate, 35 mM ascorbate, H_2O (pH 5.3)). Cyt *c* reduction by ascorbate together with pH 5.3 slows H–^{2}H exchange and minimizes background labeling during sample workup.

Final samples were concentrated (from 15 ml to 0.5 ml, 4 °C, Amicon centriprep YM-10), moved into NMR buffer (Sephadex spin columns pre-equilibrated with 100 mM deuterated acetate, 12 mM ascorbate, ^2H_2O (pD_{read} 5)), transferred to an NMR tube filled with argon, and frozen at −80 °C pending NMR analysis. Typical workup time was two hours. Duplicate samples were run at each pH and the data were averaged.

Data analysis

2D correlated spectroscopy (COSY) NMR spectra were recorded on a 500 MHz Varian INOVA spectrometer in magnitude mode (24 scans of 2048 complex data points for 512 increments; no water suppression, 20 °C). Spectra were processed using Felix 2.3 (Biosym/MSI) on a Silicon Graphics workstation. NH–C$^\alpha$H cross-peak volumes for each amide were used to obtain the fractional H labeling, H_{obs}, as in equation (2):

$$H_{obs} = \frac{V/V^{ref}}{V_C/V_C^{ref}} \quad (2)$$

The cross-peak volume after background subtraction, V, is normalized to the volume for a non-exchanging reference cross-peak (V^{ref}, heme bridge 4). V_C and V_C^{ref} are the analogous values in a ^{2}H/H equilibrated control, representing 100% labeling, obtained by using the native protein equilibrated to the same ^{2}H/H ratio as during the pulse (pH 7.5 at 55 °C for three hours). Note that H_{obs} and the analogous terms below represent fractions from 0 to 1.

H_{obs} includes the labeling that occurs in all of the protein forms (unfolded, intermediate, and native) at all stages of the experiment (before, during, and after the pulse and during the sample workup). Analysis is necessary to remove the extraneous contributions and so extract the structural, equilibrium, and kinetic information that is contained in the way that the intermediate alone becomes labeled during the pulse.

Extraneous contributions, H_{bkgd} and $H_{N,pulse}$, were evaluated in three control experiments in which the pre-deuterated unfolded protein was passed through the entire experiment but without the pulse (H_{bkgd}), and the pre-deuterated native protein was passed through the experiment with ($H_{N,obs}$) and without the pulse ($H_{N,bkgd}$). In all three cases, the stopped-flow mixing and sample workup were done exactly as before and the fractional labeling was calculated as in equation (2).

The no-pulse control for H_{bkgd} determines the summed background labeling at each amide during the entire experiment except during the pulse. H_{obs} can then be expressed as the sum of two terms: the fractional labeling at each site during the pulse (H_{pulse}) and the additional labeling during steps other than the pulse, as in equation (3):

$$H_{obs} = H_{pulse} + (1 - H_{pulse})H_{bkgd} \quad (3)$$

H_{bkgd} comes mainly from labeling during the sample workup. It is ~0.1 for most of the amides in the blue and green units, and ~0.3 for amides in the other units, similar to values found before for WT Cyt *c*.[24] H_{bkgd} is much lower in the nested-yellow Ω-loop of pWT than was found for WT. The pWT protein does not contain His26, which tends to protonate at low pH (in WT) and destabilize the nested-yellow loop.[69] The absence of this destabilizing effect at the low pH of the sample workup allowed us to obtain data for eight amides in the nested-yellow Ω-loop of pWT compared to three in WT.[24]

H_{pulse} includes mainly the fractional labeling of the intermediate during the pulse ($H_{I,pulse}$), which is the experimental parameter of major interest, but it also includes the labeling in the 42% of native protein (f_N) present during the pulse ($H_{N,pulse}$), as expressed in equation (4) (42%=39% fast phase+3% of the slower phase). $H_{N,pulse}$ at each pulse pH was calculated from the two native controls $H_{N,obs}$ and $H_{N,bkgd}$ (like equation (3)). $H_{N,pulse}$ values were close to zero for amides in the three helices at all pulse pH values; some labeling occurred at high pH for amides in the three Ω loops but this correction is multiplied down by f_N (0.42) (equation (4)):

$$H_{pulse} = f_U H_{U,pulse} + f_I H_{I,pulse} + f_N H_{N,pulse} \quad (4)$$

In the present case, $f_U \sim 0$ (Figure 2(a)). Hence f_I in equation (4) becomes $(1-f_N)$.

Equations (3) and (4) together with the independently evaluated parameters, H_{bkgd}, $H_{N,pulse}$, and f_N, connect the measured H_{obs} to the desired $(1-f_N)H_{I,pulse}$:

$$(1-f_N)H_{I,pulse} = \frac{H_{obs} - H_{bkgd}}{1 - H_{bkgd}} - f_N H_{N,pulse} \quad (5)$$

(H_{bkgd} must not be too large.)

As indicated, these various terms can be evaluated from experiments that pass the native protein and the initially unfolded protein through the entire stopped-flow and sample preparation procedure, both with and without the labeling pulse. From this, the desired parameter $H_{I,pulse}$ can be obtained. The equations described below relate $H_{I,pulse}$ to opening and closing (deprotection and reprotection) rates. These relationships allow experimental pulse labeling data for H_{obs} *versus* pH to be fit in order to obtain, in favorable cases, amide resolved opening equilibrium constants and opening and closing rates as they exist in a populated intermediate.

We compare the observed labeling in the intermediate state ($(1-f_N)H_{I,pulse}$) with that expected when there is no protection in the unfolded state, calculated as:

$$(1-f_N)H_{I,unprotected} = (1-f_N)(1-\exp(-k_{ch}t_p)) \quad (6)$$

where t_p (0.05 second) is the pulse time and k_{ch} is the unprotected chemical exchange rate calculated for the ambient conditions from model dipeptide calibrations[65,94]†.

Data fitting

The Linderstrøm–Lang scheme for opening dependent HX was assumed:[95]

$$\mathrm{NH(closed)} \underset{k_{cl}}{\overset{k_{op}}{\rightleftarrows}} \mathrm{NH(open)} \xrightarrow{k_{ch}} \mathrm{exchanged} \quad (7)$$

As usual, k_{op} and k_{cl} are the opening (unfolding) and closing (folding) rates of the protecting structure, and k_{ch} is the unprotected chemical exchange rate[65,94]†.

However, the usual steady-state equation is inadequate for the present situation, wherein a finite fraction of any given amide may be unprotected and not in rapid equilibrium with the protected form (closed form). In this case, the first bracketed term in equation (8) is necessary. When this term is small ($k_{cl} \gg k_{op}$, k_{ch}, and/or the pre-folding time t_f is sufficiently long; see equation (10)), equation (8) reduces to the usual steady-state equation.

The data were analyzed using the complete HX equations adapted from Hvidt[24,60,96] (equations (8)–(10)):

$$H_{I,pulse} = 1 - \left(\frac{k_{ch}[\mathrm{NH(open)}]_{t_f} - \lambda_2}{\lambda_1 - \lambda_2}\right)e^{-\lambda_1 t_p} - \left(\frac{\lambda_1 - k_{ch}[\mathrm{NH(open)}]_{t_f}}{\lambda_1 - \lambda_2}\right)e^{-\lambda_2 t_p} \quad (8)$$

where:

$$\lambda_{1,2} = \frac{k_{op} + k_{cl} + k_{ch} \pm \sqrt{(k_{op} + k_{cl} + k_{ch})^2 - 4k_{op}k_{ch}}}{2} \quad (9)$$

Here $[\mathrm{NH(open)}]_{t_f}$ is the population fraction in the open form during the pulse, after folding for t_f (0.1 second). Equation (10) expresses $[\mathrm{NH(open)}]_{t_f}$ in these terms (assuming no protection in the unfolded state):

$$[\mathrm{NH(open)}]_{t_f} = \frac{k_{op}}{k_{op} + k_{cl}} + \frac{k_{cl}}{k_{op} + k_{cl}} e^{-(k_{op}+k_{cl})t_f} \quad (10)$$

To implement these equations, the basic experimental data for each amide (H_{obs} *versus* pulse pH; Figure 5) were fit to equation (3) using equation (4) for H_{pulse} and equations (7)–(10) for $H_{I,pulse}$, with H_{bkgd} and $H_{N,pulse}$ fixed at the values determined in the control experiments described. Alternatively, one can calculate $H_{I,pulse}$ using equation (5) and then fit to equations (7)–(10). Fitting parameters are k_{op} and k_{cl}; all others are experimentally determined. Data fitting used Sigma Plot 2001.

The unfolding free energy of protecting structure was obtained using equation (11):

$$\Delta G_{HX} = -RT \ln\left(\frac{k_{op}}{k_{cl}}\right) \quad (11)$$

The errors in ΔG_{HX} were calculated from the individual errors in k_{op} and k_{cl} obtained from data fitting using standard error propagation formulae.[97]

† http://hx2.med.upenn.edu/download.html

Acknowledgements

We thank Jon Rumbley for creating the pWT construct, Leland Mayne for helpful discussions, and T. R. Sosnick, N. R. Kallenbach, R. L. Baldwin and C. R. Matthews for comments on the manuscript. This work was supported by research grants from the NIH and the Mathers Foundation.

References

1. Levinthal, C. (1968). Are there pathways for protein folding. *J. Chim. Phys.* **65**, 44–45.

2. Kim, P. S. & Baldwin, R. L. (1982). Specific intermediates in the folding reactions of small proteins and the mechanism of protein folding. *Annu. Rev. Biochem.* **51**, 459–489.
3. Kim, P. S. & Baldwin, R. L. (1990). Intermediates in the folding reactions of small proteins. *Annu. Rev. Biochem.* **59**, 631–660.
4. Ptitsyn, O. B. (1995). Molten globule and protein folding. *Advan. Protein Chem.* **47**, 83–229.
5. Wolynes, P. G., Onuchic, J. N. & Thirumalai, D. (1995). Navigating the folding routes. *Science*, **267**, 1619–1620.
6. 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: Struct. Funct. Genet.* **21**, 167–195.
7. Onuchic, J. N., Wolynes, P. G., Luthey-Schulten, Z. & Socci, N. D. (1995). Toward an outline of the topography of a realistic protein-folding funnel. *Proc. Natl Acad. Sci. USA*, **92**, 3626–3630.
8. Shakhnovich, E. I. (1997). Theoretical studies of protein-folding thermodynamics and kinetics. *Curr. Opin. Struct. Biol.* **7**, 29–40.
9. Dill, K. A. & Chan, H. S. (1997). From Levinthal to pathways to funnels. *Nature Struct. Biol.* **4**, 10–19.
10. Plotkin, S. S. & Onuchic, J. N. (2002). Understanding protein folding with energy landscape theory. Part I: basic concepts. *Quart. Rev. Biophys.* **35**, 111–167.
11. Plotkin, S. S. & Onuchic, J. N. (2002). Understanding protein folding with energy landscape theory. Part II: quantitative concepts. *Quart. Rev. Biophys.* **35**, 205–286.
12. Ozkan, S. B., Dill, K. A. & Bahar, I. (2002). Fast-folding protein kinetics, hidden intermediates, and the sequential stabilization model. *Protein Sci.* **11**, 1958–1970.
13. Kuwajima, K., Kim, P. S. & Baldwin, R. L. (1983). Strategy for trapping intermediates in the folding of ribonuclease A and for using ^{1}H-NMR to determine their structures. *Biopolymers*, **22**, 59–67.
14. 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.
15. Roder, H., Elöve, G. A. & Englander, S. W. (1988). Structural characterization of folding intermediates in cytochrome *c* by H-exchange labeling and proton NMR. *Nature*, **335**, 700–704.
16. Schultz, D. A., Schmid, F. X. & Baldwin, R. L. (1992). *Cis* proline mutants of ribonuclease A. II. Elimination of the slow-folding forms by mutation. *Protein Sci.* **1**, 917–924.
17. Babul, J. & Stellwagen, E. (1972). Participation of the protein ligands in the folding of cytochrome *c*. *Biochemistry*, **11**, 1195–1200.
18. Sosnick, T. R., Mayne, L., Hiller, R. & Englander, S. W. (1994). The barriers in protein folding. *Nature Struct. Biol.* **1**, 149–156.
19. 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.
20. Pierce, M. M. & Nall, B. T. (2000). Coupled kinetic traps in cytochrome *c* folding: his-heme misligation and proline isomerization. *J. Mol. Biol.* **298**, 955–969.
21. Sosnick, T. R., Mayne, L. & Englander, S. W. (1996). Molecular collapse: the rate-limiting step in two-state cytochrome *c* folding. *Proteins: Struct. Funct. Genet.* **24**, 413–426.
22. Baldwin, R. L. (1995). The nature of protein folding pathways: the classical *versus* the new view. *J. Biomol. NMR*, **5**, 103–109.
23. Wallace, L. A. & Matthews, C. R. (2002). Sequential versus parallel protein-folding mechanisms: experimental tests for complex folding reactions. *Biophys. Chem.* **101–102**, 113–131.
24. Krishna, M. M. G., Lin, Y., Mayne, L. & Englander, S. W. (2003). Intimate view of a kinetic protein folding intermediate: residue-resolved structure, interactions, stability, folding and unfolding rates, homogeneity. *J. Mol. Biol.* **334**, 501–513.
25. Rumbley, J. N., Hoang, L. & Englander, S. W. (2002). Recombinant equine cytochrome *c* in *Escherichia coli*: high-level expression, characterization, and folding and assembly mutants. *Biochemistry*, **41**, 13894–13901.
26. Ikai, A., Fish, W. W. & Tanford, C. (1973). Kinetics of unfolding and refolding of proteins. II Results for cytochrome *c*. *J. Mol. Biol.* **73**, 165–184.
27. Englander, S. W., Sosnick, T. R., Mayne, L. C., Shtilerman, M., Qi, P. X. & Bai, Y. (1998). Fast and slow folding in cytochrome *c*. *Acc. Chem. Res.* **31**, 737–744.
28. Yeh, S. R., Han, S. W. & Rousseau, D. L. (1998). Cytochrome *c* folding and unfolding: a biphasic mechanism. *Acc. Chem. Res.* **31**, 727–736.
29. Lyubovitsky, J. G., Gray, H. B. & Winkler, J. R. (2002). Mapping the cytochrome *c* folding landscape. *J. Am. Chem. Soc.* **124**, 5481–5485.
30. Telford, J. R., Wittung-Stafshede, P., Gray, H. B. & Winkler, J. R. (1998). Protein folding triggered by electron transfer. *Acc. Chem. Res.* **31**, 755–763.
31. Russell, B. S., Melenkivitz, R. & Bren, K. L. (2000). NMR investigation of ferricytochrome *c* unfolding: detection of an equilibrium unfolding intermediate and residual structure in the denatured state. *Proc. Natl Acad. Sci. USA*, **97**, 8312–8317.
32. Travaglini-Allocatelli, C., Gianni, S., Morea, V., Tramontano, A., Soulimane, T. & Brunori, M. (2003). Exploring the cytochrome *c* folding mechanism: cytochrome *c*552 from *Thermus thermophilus* folds through an on-pathway intermediate. *J. Biol. Chem.* **278**, 41136–41140.
33. Liu, W., Rumbley, J., Englander, S. W. & Wand, A. J. (2003). Backbone and side-chain heteronuclear resonance assignments and hyperfine NMR shifts in horse cytochrome *c*. *Protein Sci.* **12**, 2104–2108.
34. Hoang, L. D. (2002) Transient intermediate units in protein folding. PhD thesis, University of Pennsylvania.
35. Maity, H., Maity, M. & Englander, S. W. (2004). How cytochrome *c* folds, and why: submolecular foldon units and their stepwise sequential stabilization. *J. Mol. Biol.* **343**, 223–233.
36. Jeng, M.-F. & Englander, S. W. (1991). Stable submolecular folding units in a non-compact form of cytochrome *c*. *J. Mol. Biol.* **221**, 1045–1061.
37. Goto, Y., Hagihara, Y., Hamada, D., Hoshino, M. & Nishii, I. (1993). Acid-induced unfolding and refolding transitions of cytochrome *c*: a three-state mechanism in H_2O and D_2O. *Biochemistry*, **32**, 11878–11885.
38. Wood, L. C., White, T. B., Ramdas, L. & Nall, B. T. (1988). Replacement of a conserved proline eliminates the absorbance-detected slow folding phase of iso-2-cytochrome *c*. *Biochemistry*, **27**, 8562–8568.

39. Ridge, J. A., Baldwin, R. L. & Labhardt, A. M. (1981). Nature of fast and slow refolding reactions of iron(III) cytochrome *c*. *Biochemistry*, **20**, 1622–1630.
40. Nawrocki, J. P., Chu, R.-A., Pannell, L. K. & Bai, Y. (1999). Intermolecular aggregations are responsible for the slow kinetics observed on the folding of cytochrome *c* at neutral pH. *J. Mol. Biol.* **293**, 991–995.
41. Segel, D. J., Eliezer, D., Uversky, V., Fink, A. L., Hodgson, K. O. & Doniach, S. (1999). Transient dimer in the refolding kinetics of cytochrome *c* characterized by small-angle X-ray scattering. *Biochemistry*, **38**, 15352–15359.
42. Hammack, B., Godbole, S. & Bowler, B. E. (1998). Cytochrome *c* folding traps are not due solely to histidine-heme ligation: direct demonstration of a role for N-terminal amino group-heme ligation. *J. Mol. Biol.* **275**, 719–724.
43. Krantz, B. A., Mayne, L., Rumbley, J., Englander, S. W. & Sosnick, T. R. (2002). Fast and slow intermediate accumulation and the initial barrier mechanism in protein folding. *J. Mol. Biol.* **324**, 359–371.
44. Bai, Y. (1999). Kinetic evidence for an on-pathway intermediate in the folding of cytochrome *c*. *Proc. Natl Acad. Sci. USA*, **96**, 477–480.
45. Bhuyan, A. K. & Udgaonkar, J. B. (2001). Folding of horse cytochrome *c* in the reduced state. *J. Mol. Biol.* **312**, 1135–1160.
46. Weissman, J. S. & Kim, P. S. (1995). A kinetic explanation for the rearrangement pathway of BPTI folding. *Nature Struct. Biol.* **2**, 1123–1130.
47. Bilsel, O., Zitzewitz, J. A., Bowers, K. E. & Matthews, C. R. (1999). Folding mechanism of the α-subunit of tryptophan synthase, an α/β barrel protein: global analysis highlights the interconversion of multiple native, intermediate, and unfolded forms through parallel channels. *Biochemistry*, **38**, 1018–1029.
48. Muñoz, V., Lopez, E. M., Jager, M. & Serrano, L. (1994). Kinetic characterization of the chemotactic protein from *Escherichia coli*, CheY. Kinetic analysis of the inverse hydrophobic effect. *Biochemistry*, **33**, 5858–5866.
49. Silow, M. & Oliveberg, M. (1997). Transient aggregates in protein folding are easily mistaken for folding intermediates. *Proc. Natl Acad. Sci. USA*, **94**, 6084–6086.
50. Oliveberg, M. (1998). Alternative explanations for "multistate" kinetics in protein folding: transient aggregation and changing transition-state ensembles. *Acc. Chem. Res.* **31**, 765–772.
51. Brems, D. N. (1988). Solubility of different folding conformers of bovine growth harmone. *Biochemistry*, **27**, 4541–4546.
52. Finke, J. M., Roy, M., Zimm, B. H. & Jennings, P. A. (2000). Aggregation events occur prior to stable intermediate formation during refolding of interleukin 1β. *Biochemistry*, **39**, 575–583.
53. Goldberg, M. E., Rudolph, R. & Jaenicke, R. (1991). A kinetic study of the competition between renaturation and aggregation during the refolding of denatured-reduced egg white lysozyme. *Biochemistry*, **30**, 2790–2797.
54. Ewbank, J. J. & Creighton, T. E. (1993). Pathway of disulfide-coupled unfolding and refolding of bovine α-lactalbumin. *Biochemistry*, **32**, 3677–3693.
55. Pecorari, F., Minard, P., Desmadril, M. & Yon, J. M. (1996). Occurrence of transient multimeric species during the refolding of a monomeric protein. *J. Biol. Chem.* **271**, 5270–5276.
56. Eliezer, D., Chiba, K., Tsuruta, H., Doniach, S., Hodgson, K. O. & Kihara, H. (1993). Evidence of an associative intermediate on the myoglobin refolding pathway. *Biophys. J.* **65**, 912–917.
57. Chu, R.-A. & Bai, Y. (2002). Lack of definable nucleation sites in the rate-limiting transition state of barnase under native conditions. *J. Mol. Biol.* **315**, 759–770.
58. Fersht, A. R. (2000). A kinetically significant intermediate in the folding of barnase. *Proc. Natl Acad. Sci. USA*, **97**, 14121–14126.
59. Went, H. M., Benitez-Cardoza, C. G. & Jackson, S. E. (2004). Is an intermediate state populated on the folding pathway of ubiquitin? *FEBS Letters*, **567**, 333–338.
60. Krishna, M. M. G., Hoang, L., Lin, Y. & Englander, S. W. (2004). Hydrogen exchange methods to study protein folding. *Methods*, **34**, 51–64.
61. Xu, Y., Mayne, L. & Englander, S. W. (1998). Evidence for an unfolding and refolding pathway in cytochrome *c*. *Nature Struct. Biol.* **5**, 774–778.
62. Arrington, C. B. & Robertson, A. D. (1997). Microsecond protein folding kinetics from native-state hydrogen exchange. *Biochemistry*, **36**, 8686–8691.
63. Hoang, L., Bédard, S., Krishna, M. M. G., Lin, Y. & Englander, S. W. (2002). Cytochrome *c* folding pathway: kinetic native-state hydrogen exchange. *Proc. Natl Acad. Sci. USA*, **99**, 12173–12178.
64. Ferraro, D. M., Lazo, N. D. & Robertson, A. D. (2004). EX1 hydrogen exchange and protein folding. *Biochemistry*, **43**, 587–594.
65. Bai, Y., Milne, J. S., Mayne, L. & Englander, S. W. (1993). Primary structure effects on peptide group hydrogen exchange. *Proteins: Struct. Funct. Genet.* **17**, 75–86.
66. Bai, Y., Sosnick, T. R., Mayne, L. & Englander, S. W. (1995). Protein folding intermediates: native-state hydrogen exchange. *Science*, **269**, 192–197.
67. Milne, J. S., Xu, Y., Mayne, L. C. & Englander, S. W. (1999). Experimental study of the protein folding landscape: unfolding reactions in cytochrome *c*. *J. Mol. Biol.* **290**, 811–822.
68. Rumbley, J., Hoang, L., Mayne, L. & Englander, S. W. (2001). An amino acid code for protein folding. *Proc. Natl Acad. Sci. USA*, **98**, 105–112.
69. Krishna, M. M. G., Lin, Y., Rumbley, J. N. & Englander, S. W. (2003). Cooperative omega loops in cytochrome *c*: role in folding and function. *J. Mol. Biol.* **331**, 29–36.
70. Hoang, L., Maity, H., Krishna, M. M. G., Lin, Y. & Englander, S. W. (2003). Folding units govern the cytochrome *c* alkaline transition. *J. Mol. Biol.* **331**, 37–43.
71. Englander, S. W. (2000). Protein folding intermediates and pathways studied by hydrogen exchange. *Annu. Rev. Biophys. Biomol. Struct.* **29**, 213–238.
72. Chamberlain, A. K. & Marqusee, S. (2000). Comparison of equilibrium and kinetic approaches for determining protein folding mechanisms. *Advan. Protein Chem.* **53**, 283–328.
73. Bai, Y., Milne, J. S., Mayne, L. & Englander, S. W. (1994). Protein stability parameters measured by hydrogen exchange. *Proteins: Struct. Funct. Genet.* **20**, 4–14.
74. Hughson, F. M., Wright, P. E. & Baldwin, R. L. (1990). Structural characterization of a partly folded apomyoglobin intermediate. *Science*, **249**, 1544–1548.

75. Loh, S. N., Kay, M. S. & Baldwin, R. L. (1995). Structure and stability of a second molten globule intermediate in the apomyoglobin folding pathway. *Proc. Natl Acad. Sci. USA*, **92**, 5446–5450.
76. Tsui, V., Garcia, C., Cavagnero, S., Siuzdak, G., Dyson, H. J. & Wright, P. E. (1999). Quench-flow experiments combined with mass spectrometry show apomyoglobin folds through an obligatory intermediate. *Protein Sci.* **8**, 45–49.
77. Nishimura, C., Wright, P. W. & Dyson, H. J. (2003). Role of B-helix in early folding events in apomyoglobin: evidence from site-directed mutagenesis for native-like long range interactions. *J. Mol. Biol.* **334**, 293–307.
78. Hollien, J. & Marqusee, S. (1999). Structural distribution of stability in a thermophilic enzyme. *Proc. Natl Acad. Sci USA*, **96**, 13674–13678.
79. 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 Struct. Biol.* **4**, 298–304.
80. Raschke, T. M., Kho, J. & Marqusee, S. (1999). Confirmation of the hierarchical folding of RNase H: a protein engineering study. *Nature Struct. Biol.* **6**, 825–831.
81. Dabora, J. M., Pelton, J. G. & Marqusee, S. (1996). Structure of the acid state of *Escherichia coli* ribonuclease HI. *Biochemistry*, **35**, 11951–11958.
82. Chamberlain, A. K., Fischer, K. F., Reardon, D., Handel, T. M. & Marqusee, A. S. (1999). Folding of an isolated ribonuclease H core fragment. *Protein Sci.* **8**, 2251–2257.
83. Fuentes, E. J. & Wand, A. J. (1998). Local stability and dynamics of apocytochrome *b*562 examined by the dependence of hydrogen exchange on hydrostatic pressure. *Biochemistry*, **37**, 9877–9883.
84. Fuentes, E. J. & Wand, A. J. (1998). Local dynamics and stability of apocytochrome *b*562 examined by hydrogen exchange. *Biochemistry*, **37**, 3687–3698.
85. Chu, R., Pei, W., Takei, J. & Bai, Y. (2002). Relationship between the native-state hydrogen exchange and folding pathways of a four-helix bundle protein. *Biochemistry*, **41**, 7998–8003.
86. Takei, J., Pei, W., Vu, D. & Bai, Y. (2002). Populating partially unfolded forms by hydrogen exchange-directed protein engineering. *Biochemistry*, **41**, 12308–12312.
87. Feng, H., Takei, J., Lipsitz, R., Tjandra, N. & Bai, Y. (2003). Specific non-native hydrophobic interactions in a hidden folding intermediate: implications for protein folding. *Biochemistry*, **42**, 12461–12465.
88. 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. *J. Mol. Biol.* **323**, 363–375.
89. Silverman, J. A. & Harbury, P. B. (2002). The equilibrium unfolding pathway of a $(\beta/\alpha)_8$ barrel. *J. Mol. Biol.* **324**, 1031–1040.
90. Krantz, B. A., Srivastava, A. K., Nauli, S., Baker, D., Sauer, R. T. & Sosnick, T. R. (2002). Understanding protein hydrogen bond formation with kinetic H/D amide isotope effects. *Nature Struct. Biol.* **9**, 458–463.
91. Plaxco, K. W., Simons, K. T. & Baker, D. (1998). Contact order, transition state placement and the refolding rates of single domain proteins. *J. Mol. Biol.* **277**, 985–994.
92. Makarov, D. E. & Plaxco, K. W. (2003). The topomer search model: a simple, quantitative theory of two-state protein folding kinetics. *Protein Sci.* **12**, 17–26.
93. Krantz, B. A., Moran, L. B., Kentsis, A. & Sosnick, T. R. (2000). D/H amide kinetic isotope effects reveal when hydrogen bonds form during protein folding. *Nature Struct. Biol.* **7**, 62–71.
94. Connelly, G. P., Bai, Y., Jeng, M.-F. & Englander, S. W. (1993). Isotope effects in peptide group hydrogen exchange. *Proteins: Struct. Funct. Genet.* **17**, 87–92.
95. Linderstrøm-Lang, K. (1958). Deuterium exchange and protein structure. In *Symposium on Protein Structure* (Neuberger, A., ed), pp. 23–24, Methuen, London.
96. Hvidt, A. (1964). A discussion of the pH dependence of the hydrogen-deuterium exchange of proteins. *C.R. Trav. Lab. Carlsberg*, **34**, 299–317.
97. Bevington, P. R. & Robinson, D. K. (1994). *Data Reduction and Error Analysis for the Physical Sciences* (2nd edit.). McGraw-Hill, New York.
98. Bushnell, G. W., Louie, G. V. & Brayer, G. D. (1990). High-resolution three-dimensional structure of horse heart cytochrome *c*. *J. Mol. Biol.* **214**, 585–595.
99. Kraulis, P. J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallog.* **24**, 945–949.
100. Bai, Y. & Englander, S. W. (1996). Future directions in folding: the multi-state nature of protein structure. *Proteins: Struct. Funct. Genet.* **24**, 145–151.

Edited by C. R. Matthews

(Received 7 July 2004; received in revised form 20 August 2004; accepted 26 August 2004)