doi:10.1016/j.jmb.2003.09.070 J. Mol. Biol. (2003) 334, 501–513

JMB

Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER

Intimate View of a Kinetic Protein Folding Intermediate: Residue-resolved Structure, Interactions, Stability, Folding and Unfolding Rates, Homogeneity

Mallela M. G. Krishna*, Yan Lin, Leland Mayne 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

A cytochrome *c* kinetic folding intermediate was studied by hydrogen exchange (HX) pulse labeling. Advances in the technique and analysis made it possible to define the structured and unstructured regions, equilibrium stability, and kinetic opening and closing rates, all at an amino acid-resolved level. The entire N-terminal and C-terminal helices are formed and docked together at their normal native positions. They fray in both directions from the interaction region, due to a progression in both unfolding and refolding rates, leading to the surprising suggestion that helix propagation may proceed very slowly in the condensed milieu. Several native-like beta turns are formed. Some residues in the segment that will form the native 60s helix are protected but others are not, suggesting energy minimization to some locally non-native conformation in the transient intermediate. All other regions are unprotected, presumably dynamically disordered. The intermediate resembles a partially constructed native state. It is early, on-pathway, and all of the refolding molecules pass through it. These and related results consistently point to distinct, homogeneous, native-like intermediates in a stepwise sequential pathway, guided by the same factors that determine the native structure. Previous pulse labeling efforts have always assumed EX2 exchange during the labeling pulse, often leading to the suggestion of heterogeneous intermediates in alternative parallel pathways. The present work reveals a dominant role for EX1 exchange in the high pH labeling pulse, which will mimic heterogeneous behavior when EX2 exchange is assumed. The general problem of homogeneous *versus* heterogeneous intermediates and pathways is discussed.

© 2003 Elsevier Ltd. All rights reserved.

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

Introduction

Do proteins fold through a multiplicity of alternative intermediates and parallel pathways, as envisioned by many theoretical[1–3] and experimental studies,[4] or through a small number of distinct intermediates on discrete pathway(s) to the native state?[5–7] To understand how proteins fold, and why, it will be necessary to obtain detailed information on the structure of kinetic folding intermediates. This has been difficult. Kinetic intermediates exist only briefly (< seconds) and cannot be isolated for structural studies. Spectroscopic methods fast enough to follow kinetic folding yield essentially one-parameter data that provide no structural definition and can confuse intermediate formation with events such as chain compaction and transient aggregation. Theoretical methods are not yet able to deal with the complex structures, potential functions, and folding time-scale of real proteins.

Detailed structural information has come mainly from hydrogen exchange (HX) methods.[7] An HX pulse labeling method, applied during kinetic folding, can provide residue-resolved information

Abbreviations used: Cyt c, cytochrome *c*; HX, hydrogen exchange; NHX, native state hydrogen exchange; foldon, cooperative folding/unfolding unit; CD, circular dichroism; GdmCl, guanidinium chloride.

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

0022-2836/$ - see front matter © 2003 Elsevier Ltd. All rights reserved.

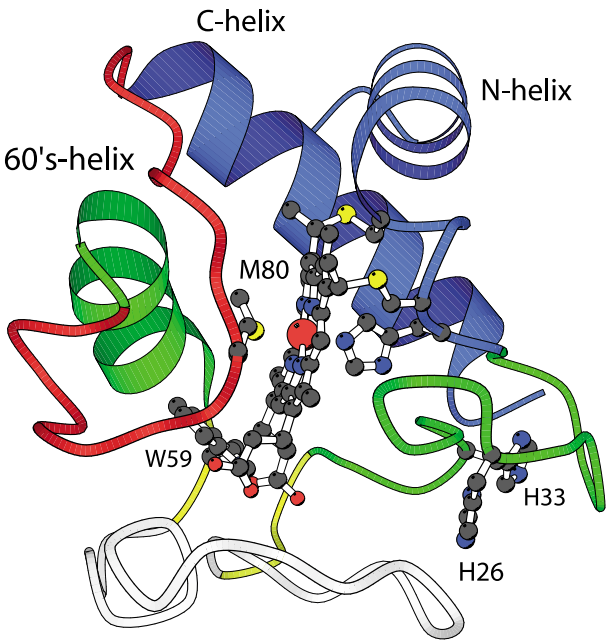

Figure 1. Molecular Cyt c structure (1HRC.pdb[80] and MOLSCRIPT[81]) color coded to indicate the folding units identified by NHX experiments and ranked spectrally in order of decreasing stability and apparent folding sequence (blue to green to yellow to red to nested yellow).[19,20,22,23,27] The five foldons are the blue unit (N and C-terminal helices), the green unit (60s helix and 20s–30s Ω-loop), the outer yellow unit (a short anti-parallel β-sheet; residues 37–39, 58–61), the nested yellow Ω-loop (residues 40–57), and the red unit (71–85 Ω-loop). The histidine residues in the green loop can misligate to the Met80 side of the heme Fe in the unfolded protein. This introduces an incipient barrier that causes transient accumulation of the initial N/C bi-helical intermediate (blue) by blocking the subsequent folding of the green unit.

on intermediates that accumulate transiently in three-state kinetic folding.[7–18] A native state HX method avoids the kinetic problem altogether by studying the high-energy, partially folded forms that proteins experience as they reversibly unfold and refold under native conditions.[7,19–21] Extensive HX results for cytochrome *c* (Cyt c; 104 residues; Figure 1) define a small number of discrete intermediates that form in a sequential pathway by the stepwise addition of native-like structural units.[19–27] However, HX pulse labeling results now available for many proteins, including Cyt c, have most often been interpreted somewhat differently, in terms of multiple alternative intermediates in alternative parallel pathways, similar to results from computer simulations of simplified folding models.

The present work reconfigured the pulse labeling experiment and the data analysis, taking into account all contributing factors and the principles of HX processes, especially at the high pH normally used in the HX labeling pulse. The corrected pulse labeling experiment made it possible to define the structure, stability, and dynamics of an initial kinetic Cyt c intermediate at amino acid resolution. It represents a discrete partially formed replica of the native protein. The entire protein population moves through this same early intermediate on the way to the native state. When pulse labeling experiments are interpreted without attention to the character of high-pH HX behavior and appropriate data controls, the results will appear to show heterogeneous intermediates in alternative parallel pathways.

Results

Optimization

One wants to find conditions where refolding Cyt c enters some intermediate state rapidly and exits slowly so that a large fraction of the protein population occupies the intermediate and can be probed by HX pulse labeling. The same information helps to choose optimum pulse labeling times.

In exploratory experiments, Cyt c was initially unfolded in concentrated guanidinium chloride (GdmCl, p^{2}H 7.5). At this pH (and denaturant), the normal Met80-S to heme iron ligand is replaced by a misligating peripheral histidine residue. The misligation inserts a latent misfolding barrier that becomes obvious later in the folding process. Refolding was initiated by dilution into pH 6 buffer (stopped-flow). Spectroscopic probes were used to detect initial chain contraction (fluorescence; Figure 2(a)), helix formation (CD_{222}; Figure 2(b)), and final native state acquisition (A_{695}; Figure 2(c)).

An initial very fast burst phase represents, we believe, the relaxation of the chain to a more contracted but still unfolded form characteristic for the lower concentration of denaturant, although see Shastry & Roder[28] for a different view. Pertinent evidence has been analyzed in detail.[29] Most convincingly, the fast continuous flow results reported by Akiyama *et al*.[30] show that the entire CD spectrum of Cyt c, recorded long after the burst phase (at 400 μs), is identical with the thermally unfolded state and the acid low-salt unfolded state (compared in Figure 5 of Akiyama *et al*.[30] and Figure 6(F) of Krantz *et al*.[29]). This state has been referred to as U′ [29] or intermediate I.[30]

The protein starts to fold on a multi-millisecond time-scale (Figure 2(a) and (b)). The histidine to heme misligation barrier is encountered, folding pauses, and some intermediate accumulates.[31,32] The native state is reached much more slowly (~one second; Figure 2(c)), rate-limited by the misligation-dependent "error repair" barrier.[32,33]

The detailed kinetics observed depends on denaturant concentration. At 0.7 M GdmCl, multiple phases are seen, perhaps due to the multiple configurations possible for the misligation-dependent barrier (His26 and His33; under and over the heme, etc.). Lower denaturant concentration

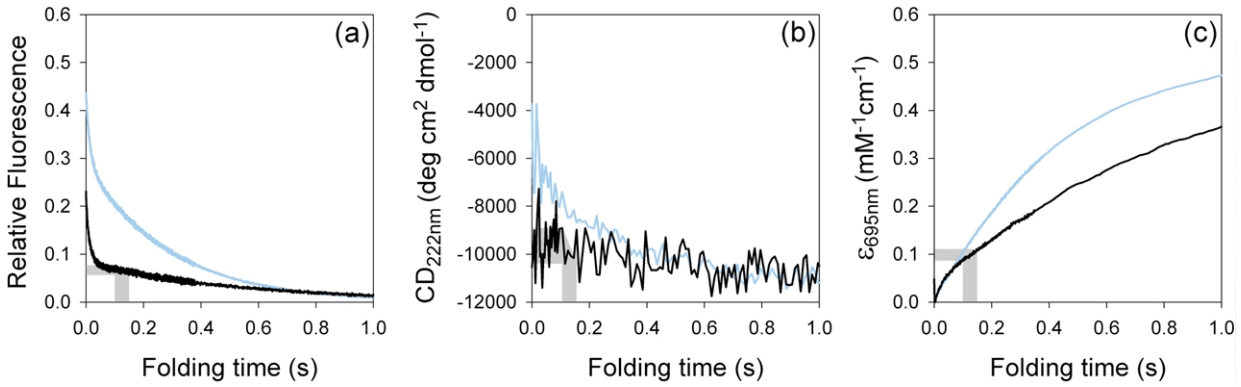

Figure 2. Stopped-flow folding kinetics of Cyt c. Unfolded Cyt c (4.2 M GdmCl, p^{2}H 7.5) was diluted into the folding buffer (pH 6.0, 10 °C) to final GdmCl concentrations of 0.23 M GdmCl (black) and 0.7 M GdmCl (cyan). Refolding was measured by (a) Trp59 fluorescence relative to the unfolded state (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; final value is 0.73; measured at 0.1 mM Cyt c as for the pulse labeling experiments). The gray shading indicates the time during which the HX pulse was applied in pulse labeling experiments.

(0.23 M GdmCl) speeds initial chain collapse (Figure 2(a)) and helix formation (Figure 2(b)) because the transition states and intermediate formation are stabilized. Inversely, the lower concentration of denaturant makes the native state formation slower (Figure 2(c); reverse denaturant effect) because the rate for overcoming the misfolding barrier is limited by an unfolding reaction.[24,32–37] Thus, lower denaturant concentration increases intermediate occupation.

The gray area in Figure 2 shows where the HX labeling pulse was applied in subsequent experiments. Figure 2(c) shows that 15% of the refolding protein reaches N before or during the HX labeling pulse, while the remaining 85% occupies the blocked intermediate state during the labeling pulse. These numbers must be known for the data analysis. It is also important to note that the fraction already native at the time of the pulse shows the same kinetic behavior (Figure 2(c)) and therefore has probably passed through the same condition as the intermediate fraction that is probed by the labeling pulse.

Final folding (Figure 2(c)) is not mono-exponential. A small fraction (~4%) folds more rapidly to N, probably because it loses the His-heme misligation before or during the initial chain collapse. In the 35% slower folding phase(s), 15% is due to proline mis-isomerization[38–40] (checked by us in double jump experiments). Much of the rest is protein concentration-dependent, due at least in part to intermolecular histidine to heme misligation.[29,31,32,41] This kinetic heterogeneity has no effect on the early N/C helix intermediate studied here, apparently because the histidine to heme misligation and the four proline residues are remote from the N and C helices and can only influence later events.

HX pulse labeling

For pulse labeling experiments, Cyt c was initially unfolded by denaturant in ^2H_2O (4.2 M GdmCl, p^{2}H 7.5). All of the amides become deuterated and almost all of the molecules have the histidine to heme iron misligation. Dilution into H_2O buffer (0.23 M GdmCl (pH 6.0), 10 °C) initiates refolding. At this condition, ^{2}H to H exchange is slow (seconds). A folding time of 100 ms was allowed in order to ensure uniform mixing and complete folding to the intermediate form ($\tau_f \sim 12$ ms; Figure 2(a)). ^{2}H to H exchange labeling was then catalyzed by a brief high-pH pulse (50 ms, pH 7.5 and above where $\tau_{ex} \sim$ ms). Amides that are still exposed in unstructured regions of the intermediate are labeled at their unprotected rate. Amides in already formed structure are protected and are labeled more slowly or not at all, as determined by the usual HX kinetic equations. A third mix into low pH (pH 5.3) terminates the labeling pulse. The protein folds to its native state within several seconds, trapping the H–^{2}H labeling profile imposed during the short pulse. Kept under slow HX conditions (~6 °C, reduced Cyt c, pH 5.3), the sample was then concentrated (centrifugal filtration) and the amount and position of H-label was read out by 2D NMR of the native protein at amino acid resolution.

This analysis, which takes hours, reveals the H–^{2}H labeling profile imprinted on a millisecond time-scale, identifying the structure that was formed and unformed at the time of the pulse. Repetition of the experiment, varying the pre-folding time or the pulse intensity (pH, time), can provide site-resolved information on the time-course of structure formation[8,9] and its stability[10,11] and even on dynamics (below). We measured the labeling obtained in multiple experiments with an increasingly strong labeling pulse (pH 7.5–10), but with constant pre-folding time (100 ms) and pulse time (50 ms).

Figure 3(a) shows a typical H-labeling pattern (pulse pH 9). Figure 3(b) shows the residues found to have significant protection in the

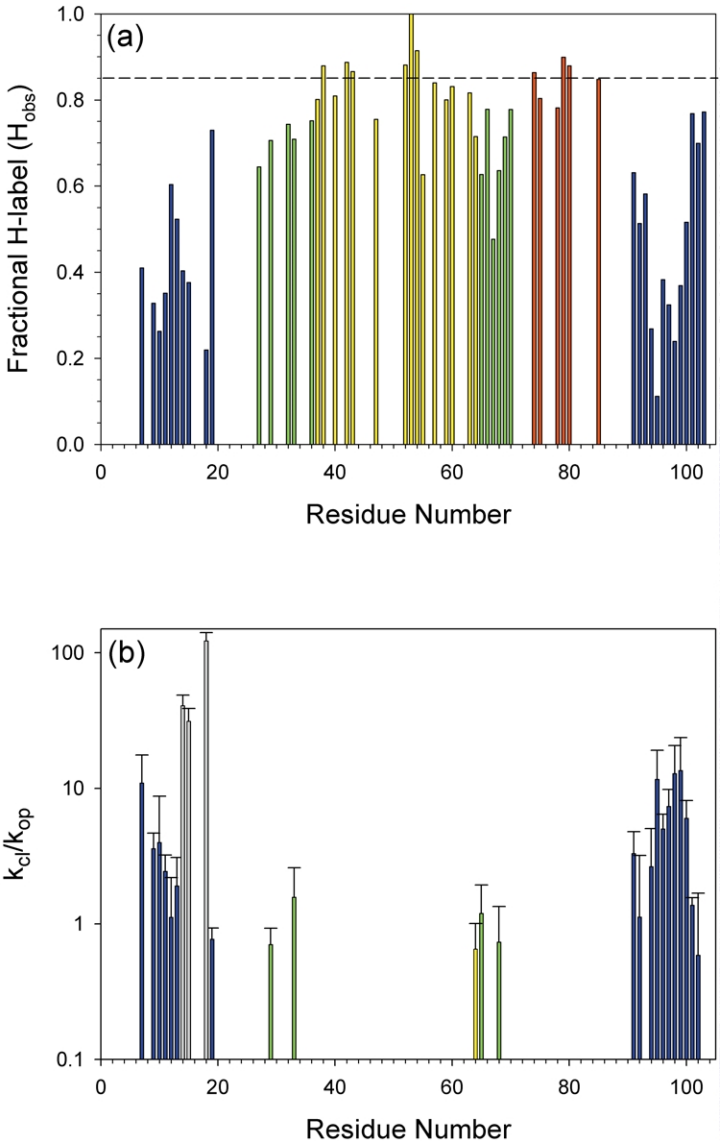

Figure 3. Labeling results. (a) The directly observed fractional H-label accumulated by the various residue amides in HX pulse labeling experiments (H_{obs} in equations (1) and (2)), in this case at pH 9.0 before correcting for the other factors described in Materials and Methods. The broken line indicates the fractional population of intermediate (15% is native). (b) The folding equilibrium constant (k_{cl}/k_{op}) calculated for each amide in the trapped intermediate, as detailed in Materials and Methods. Only the N and C helices are stably formed. Color coding relates to Figure 1. Residues 14, 15, and 18 (gray bars) are protected even in the unfolded state.

intermediate, evaluated as described below. The more complete data are in Figure 4, which shows the H-labeling at 45 amides that could be measured with good accuracy across the pulse pH range. The color coding (compare with Figure 1) identifies amides in the N and C helices (blue unit), the green loop and helix, the nested-yellow loop and outer yellow neck, and the red loop.

Data fitting

In Figure 4, the colored continuous curves fit the pH-dependent labeling data for each amide with the complete HX equations given by Hvidt[42] (equations (2)–(7)), with no assumption about HX mechanism (EX1 or EX2). The fitting parameters are k_{op} and k_{cl}, which also specify site-resolved protection (structure) and stability (K_{op}, ΔG_{op}).

The broken curves in Figure 4 relate to labeling in the intermediate alone. The black broken curves show the labeling that is expected for each amide when it is wholly unprotected (scaled to 0.85), calculated from k_{ch} as described.[43,44]† The colored broken curves show the labeling obtained at each amide ($H_{I,pulse}$), after correcting for labeling of the native protein during the pulse ($H_{N,pulse}$) and during the sample workup (H_{bkgd}) (see Materials and Methods). The corrections are nearly zero for amides in the blue and green units but are larger for some faster exchanging amides in the yellow and red segments (compare the colored curves in Figure 4).

It can be noted that the brief high-pH pulse used here, at pH 10 and lower, is unlikely to seriously distort the parameters measured. Only the two histidine residues are titratable below pH 10, both are in the still unstructured green loop, and one is already titrated and misliganded to the heme, which in addition protects against the usual

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

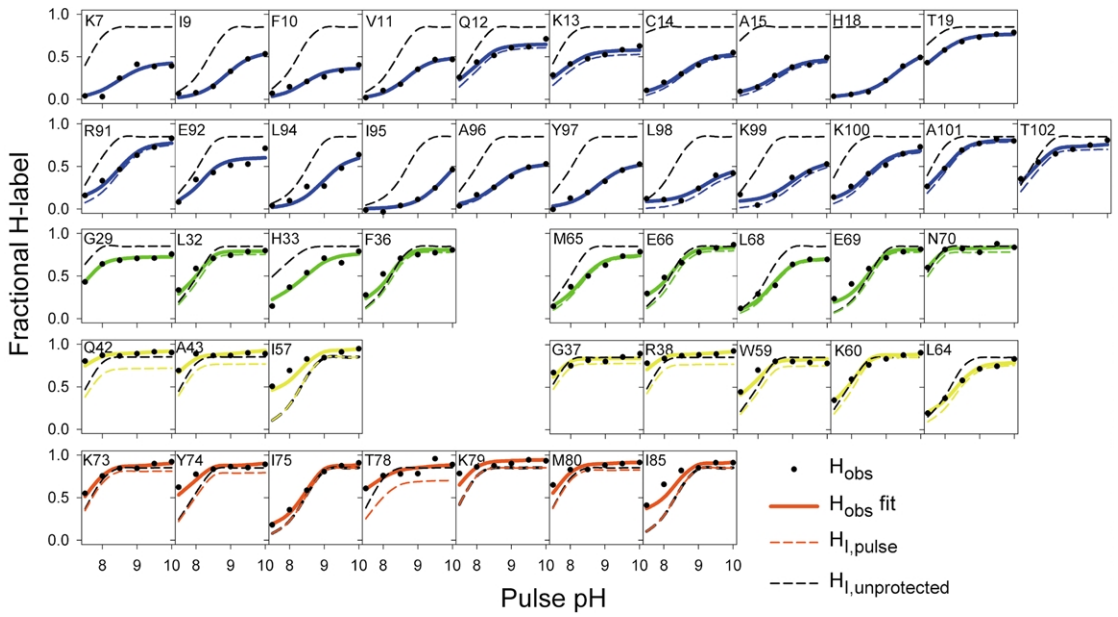

Figure 4. Residue-resolved H-labeling results for the kinetic intermediate, which specify the position of protected structure and its equilibrium and kinetic parameters (color coded to match the foldons in Figure 1) The data points show the experimentally observed pH-dependent labeling (H_{obs}; Materials and Methods). The continuous colored curves fit the equations given in Materials and Methods to the data, which yields k_{op} and k_{cl}, and therefore also K_{op}. The broken colored curves show the labeling in the trapped intermediate alone ($H_{I,pulse}$) after correcting for the extraneous contributions (H_{bkgd} and $H_{N,pulse}$). For comparison, the black broken curves show the labeling for each amide in the intermediate that would be expected in the absence of protection (scaled to 0.85, the fraction of the population in the intermediate at the time of the HX pulse).

alkaline transition.[45] The pH 10 data point is slightly elevated for some residues, perhaps due to the onset of some destabilization at elevated pH,[26] which can increase the HX labeling rate even for EX1 exchange.[46]

Structure

The results shown in Figures 3(b) and 4 identify the structured and unstructured regions of the trapped intermediate and quantify their stability and kinetic parameters. Major protection in the intermediate is found for all of the amides that are H-bonded in the N and C helices of the native protein, indicating that the entire length of both helices is formed. A similar conclusion was inferred before on the basis of many fewer protons.[8,11]

In native Cyt c, the amide hydrogen atoms of residues 64–70 are H-bonded in the 60s helix. In the intermediate, the core residues Leu64 and Leu68 and the more exposed Met65 have small but clear protection ($K_{op} \sim 1$; $k_{op} < 40\ s^{-1}$). The other residues in the 60s segment are at most marginally protected. A possible interpretation is marginal helix formation. A more likely option is that the helix is not yet formed but exposed hydrophobic residues tend to become buried. Energy minimization that forms non-native hydrophobic interactions in a folding intermediate has been demonstrated before.[47,48]

A number of amides that are H-bonded in β-turn configurations in the native protein have HX protection in the intermediate.[49,50] His33 and the heme-related residues Cys14, Ala15, and His18, all known to be protected in the unfolded state,[19,23,31] are protected in the intermediate. Arg38, Gln42 and Thr78, show small HX protection in the intermediate in the sense of a low EX1 plateau, but their equilibrium protection is within the noise level ($K_{cl} < 0.2$). Four other β-turns could not be monitored due to weak cross-peaks and fast exchange during sample workup and analysis.

Most of the amides in the red, yellow, and green loops are labeled at very close to their expected unprotected rates. This agreement provides a test of the experiment and the data analysis, and also of the previously calibrated rate constants for unprotected amides.[43,44]

Stability

Figure 5 shows equilibrium and kinetic parameters obtained for the N-terminal and C-terminal segments. For the C helix, maximum protection occurs where the hydrophobic side-chains of the N and C helices interact in the core of the native protein (Figure 5(a)). Native-like N to C interactions were inferred before in a mutational study.[51] A fraying pattern appears to emanate from that region. The N helix may also fray but this is less clear. (Note: this kind of fraying behavior may contribute to the dispersion of ΔG_{HX}

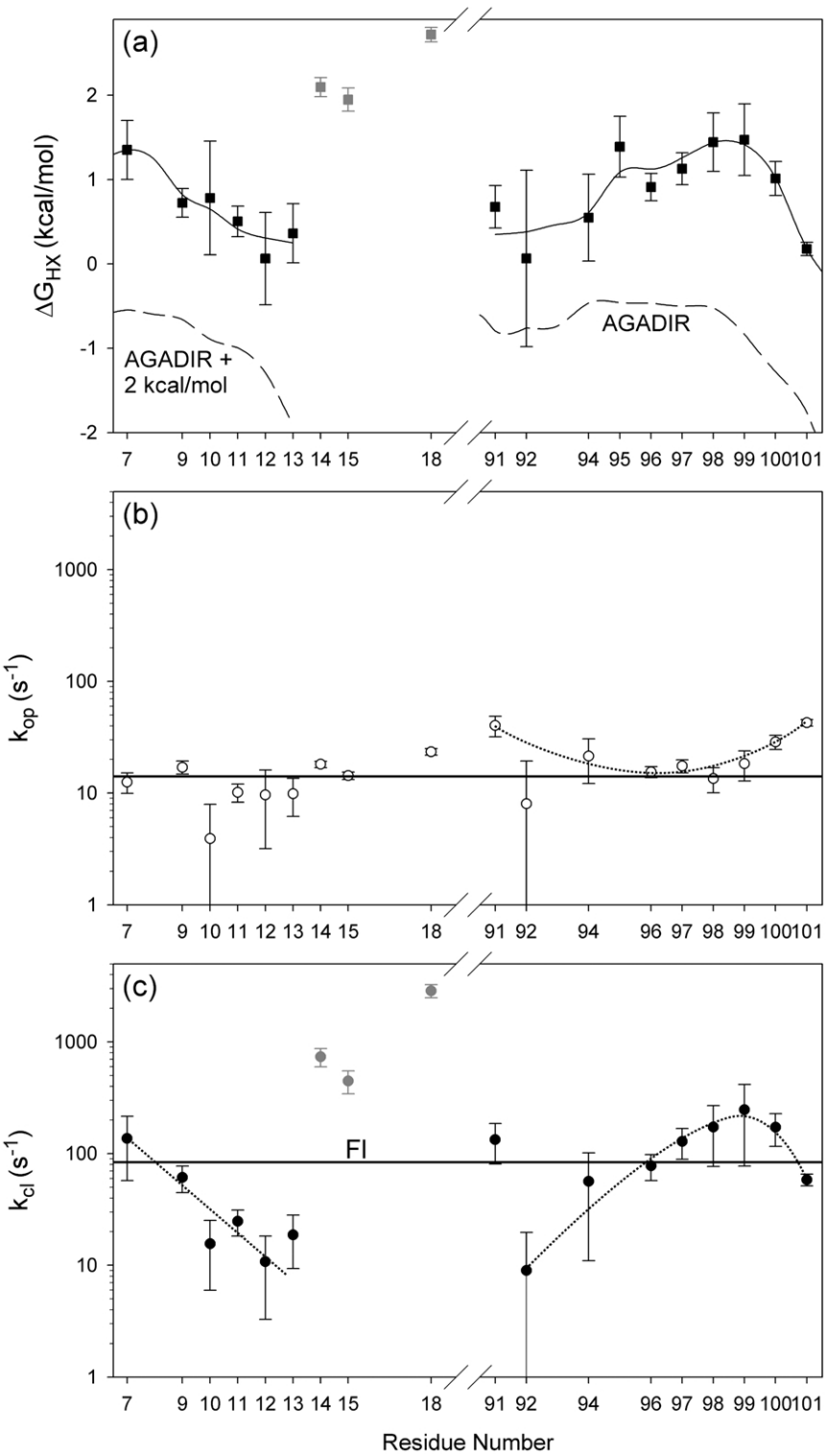

Figure 5. Equilibrium and kinetic parameters through the N and C segments in the intermediate. Error flags are at $\pm 1\sigma$. ΔG_{HX} in (a) is the residue-resolved stabilization free energy level of the unfolding reaction that allows exchange. The data are fit to Lifson–Roig helix-coil theory (continuous line). Also shown is the residue-resolved AGADIR helical stability prediction for the isolated segments[56,57] (broken line). (b) The approximately constant opening rate (continuous line at 14 s^{-1}). (c) Closing rates. The horizontal continuous line in (c) shows the folding rate measured by tryptophan fluorescence (from Figure 2(a)). The fraying behavior in (a) appears to be due mainly to a progression in refolding rates (c) but also in part to unfolding rates (b), as suggested by the dotted lines. Residues 14, 15, and 18, shown in gray, are protected even in the unfolded state.

values often seen within cooperative elements in native state HX experiments.)

The presence of helix fraying seems very reasonable. Its reality is supported by the pattern observed and the maximum apparent stability where the two helices normally interact, even though the error level for individual protons is sizeable. If the fraying pattern is real, then a fit to the Lifson–Roig equation[52,53] leads to propensity values (w) of 0.6–0.8 for the peripheral residues, in the range normally seen,[54,55]† and much higher, 4–50, for the middle helical residues, denoting external stabilization. The broken curves show the AGADIR[56,57] prediction of helical stability for these segments when unsupported‡. The N and C helices in the intermediate are more stable than calculated for the separately isolated segments by 4 kcal/mol and 2 kcal/mol, respectively (they have negative stability when unsupported).

The maximum stability for the bi-helix in the kinetic intermediate, 1.4 kcal/mol, is less than found before by native state HX for the same intermediate at the same GdmCl concentration (2.6 kcal/mol at 0.23 M GdmCl). Given the differences in conditions of measurement (10 °C at high pH for pulse labeling *versus* 30 °C at p^{2}H 7 for NHX), the actual discrepancy is probably significantly less, since Cyt c goes through maximum stability at 20–25 °C and is more stable in ^2H_2O than in H_2O.

Kinetic parameters

The labeling pattern observed in Figure 4 indicates EX2 exchange at lower pH and EX1 exchange at elevated pH. The EX2 to EX1 transition is determined by the competition between each amide's unprotected HX rate (k_{ch}) and the reprotection rate (k_{cl}).[26] In the EX2 region (lower pH), HX labeling of the protected sites increases sigmoidally with pH, as expected. At pH 8.5 and 10 °C an average unprotected amide HX rate is 100 s^{-1}, about equal to the measured rate for refolding to the intermediate. This is the condition ($k_{ch} \sim k_{cl}$) where EX1 exchange is expected to emerge, as is found. Every opening results in labeling, and the degree of labeling reaches a pH-independent plateau. Different amides reach their plateau values at higher or lower pH, consistent with the known intrinsic HX rate of each amide (k_{ch}). EX1 exchange at high pH has now been observed in a number of studies.[26,27,46,58–62]

The plateau level for each amide (after corrections) shows the fraction that fails to label because it does not open even once during the labeling pulse. The fractional labeling obtained taken together with the labeling time (50 ms) specifies k_{op}. Figure 5(b) shows that almost all of the hydrogen bonds in the two helices open at about the same rate, indicating that the two helices unfold concertedly ($k_{op} = 14(\pm 5)\ s^{-1}$). Residues near the helix ends unfold faster, as can be expected, as do the protected β-turn amides. Figure 5(c) shows amide-resolved values for k_{cl}. The calculated reclosing rates match the rate for folding to the intermediate, independently measured by fluorescence (from Figure 2(a)) (horizontal line), which helps to validate the HX result (see also Sivaraman *et al.*[63]).

The helix fraying (Figure 5(a)) appears to be due to a progression of both unfolding and refolding rates, moving from the bi-helix contact region and not from the heme attachment region. Most surprising, the helices appear to propagate (closing rate) on a millisecond time-scale, a million-fold slower than the nanosecond time-scale seen for independently stable helices in free solution. This result suggests slow phi–psi rotation in the condensed milieu, presumably due to side-chain hindrance and interactions.[64–67] The observation of fraying (Figure 5(a)) and its dependence on the helix propagation (reclosing) rate (Figure 5(c)) depends on data with significant error levels, but is supported by the persistent pattern of the data and the likelihood of the fraying result.

Homogeneity

Is the intermediate population homogeneous? The results in Figure 4 show that all intermediate forms present must have all amides in the N and C segments well protected, and many other defined sites essentially unprotected. In agreement, the CD_{222} measured at this stage in kinetic folding (Figure 2(b)) indicates that all of the molecules in the intermediate population must have essentially all of the N/C bi-helix formed (CD parameters from Chin *et al.*[68]). The intermediate population therefore appears to be homogeneous in a structural sense.

Can the intermediate population be heterogeneous in a stability sense, with differing N/C protection? The labeling curve for any given amide in Figure 4 would then consist of different components with different stabilities and opening rates. Heterogeneous intermediates would produce a superposition of curves like those in Figure 4, with the component curves variably displaced in the rising EX2 portion of the labeling curve (note that a one-component labeling curve is identical in form to the standard one proton pH titration curve). In fact, the data in Figure 4 fit well to the equations used, which assume a single homogeneous intermediate and use only two free fitting parameters. The goodness-of-fit obtained for many protons does not allow for the presence of significant heterogeneity in the measured labeling.

These considerations still do not rule out the presence of another population with only the N/C residues protected but with protection so large, or

† ftp://cmgm.stanford.edu/pub/helix

‡ http://www.embl-heidelberg.de/Services/serrano/agadir/agadir-start.html

k_{op} so small, that it does not label measurably under the conditions used. What fraction could this population represent? The data in Figure 4 show that at least 60% of each protected site is well accounted for by a single fitting curve. This fraction is limited at high pH by an EX1 plateau, indicating that the major intermediate characterized here represents a proportion of the population that is larger than 60%. Any more stable population must be less than 40%. In the regions nearer the helix ends and elsewhere, a single fitting curve accounts for a homogeneous fraction close to 100%. Also, the matching of the k_{cl} value with the folding rate measured from fluorescence (Figure 5(c)) is against the possibility that some very different fraction exists.

These results leave no room for significant heterogeneity.

Equilibrium behavior of the kinetic intermediate

The kinetic parameters found lead to an interesting perspective. The trapped intermediate unfolds and refolds repeatedly, ~ ten times ($1/k_{op} \sim 70$ ms; $1/k_{fold} \sim 800$ ms), before the barrier that blocks its forward progress is overcome and folding can proceed to N. Implications are noted in Discussion.

Discussion

The N/C intermediate: distinct and native-like

The present experiments reconfigured the HX pulse labeling experiment and analysis. The intermediate population was maximized, control experiments corrected for labeling that occurs before and after the pulse and for protection due to a population of already native molecules, and the data analysis allowed a role for high pH EX1 behavior. These advances made it possible to portray the structure, energetics, and dynamics of the populated intermediate in some detail.

The residue-resolved pattern of HX protection shows that the entire N and C-helices are formed and they appear to interact as in the native protein. A small number of protected amides elsewhere in the protein suggest several native-like β-turns, which are known to be marginally stable even in the unfolded state. Some small protection is seen for other amides, perhaps due to the burial of hydrophobic residues in some non-native but energy minimized[48] format. The unstructured regions of the partially folded intermediate undoubtedly represent a broad dynamic ensemble, as emphasized by theoretical studies, but it is clear that fairly extensive native-like structural elements are also present.

In summary, the kinetic folding intermediate strongly resembles a partially constructed native protein.

On or off pathway: the initial nucleation barrier

Is the trapped N/C intermediate on or off of the Cyt c folding pathway? A demonstration comes from studies of the folding barriers.

A previous study showed that the rate-limiting barrier for two-state Cyt c folding (U to N) is identical with the rate-limiting barrier for formation of the N/C intermediate in three-state folding (U to I). They exhibit the same rate ($\Delta G^{\ddagger}$ and pre-exponential), the same temperature dependence ($\Delta H^{\ddagger}$ and $\Delta S^{\ddagger}$), and the same denaturant dependence ($m^{\ddagger}$).[35] It is obvious that the barrier that limits the rate of two-state folding is on the folding pathway. Since the very same barrier leads directly to the N/C intermediate in three-state folding, this intermediate must also be on the main pathway (U to I to N) (the misfolding barrier that traps the N/C intermediate, producing three-state folding, is placed afterwards). In support, Krantz *et al.*[69,70] found evidence that the N and C helices are formed in the initial rate-limiting transition state in two-state folding.

Thus, it appears that Cyt c folds through the same initial on-pathway N/C intermediate whether it becomes apparent in three-state folding or remains kinetically hidden in two-state folding. The same conclusion comes from a quantity of native state HX data,[19,20,22–27] which places the N/C intermediate at the beginning of the folding pathway, even when no kinetic trapping occurs, and shows that all of the molecules move through it.

On or off pathway: the I ↔ U → N issue

The on-pathway question is raised also by the observation that the trapped N/C intermediate unfolds and refolds repeatedly (perhaps all the way to U), before the misligation-dependent barrier is overcome and folding to N can proceed.

When refolding is initiated by denaturant dilution, an unfolded protein may initially relax to some form that is not on the productive folding pathway (chain contraction, aggregation, etc.), in a sequence usually written as I ↔ U → N. This description is often taken as the definition of an off-pathway species. An off-pathway form must travel back through the U state in order to go forward. However, it should be noted that the reverse logic does not hold. A trapped intermediate that is on-pathway may still re-equilibrate with U before going forward (U ↔ I (blocked) → N). This will occur if the barrier that blocks forward folding is higher than the I to U barrier. The trapped N/C intermediate is one example. Other trapped on-pathway intermediates may do the same.

In the trapped condition, the repeated unfolding enters the rate for final folding to N as a pre-equilibrium step, as does the heme deligation step, which may occur only when the collapsed intermediate transiently unfolds. The need to unfold in order to repair the misfolding and fold to N

explains the reverse dependence of folding rate on denaturant concentration[24,32,34–37] and it duplicates a suggested mechanism for GroEl action.[71]

In summary, the N/C intermediate is clearly on-pathway, best described as a U ↔ I ↔ N situation, with a large inserted error-repair barrier slowing the I → N rate.

Homogeneity of the intermediate population

The pH-dependent pattern of labeling seen in Figure 4 has been observed in previous HX pulse labeling studies with Cyt c and other proteins. Some fraction of any given amide is labeled while another fraction appears altogether resistant to labeling, even at high pulse pH. The common interpretation, based on assuming EX2 HX behavior, has been that different fractions of the refolding population have different protection, indicating alternative folding intermediates with different structure and/or stability, supposedly in alternative parallel pathways.

The present work identifies a different reason. At pH above 9 or so, the chemical HX rate of transiently exposed amides becomes faster (<1 ms) than reclosing. This is the condition for EX1 exchange ($k_{ch} > k_{cl}$). A plateau level of labeling is reached because amides have reached their maximum labeling rate, equal to k_{op}. Amides that open at least once during the brief pulse are labeled. Those that do not open even once are not labeled, no matter what their stability may be. When EX2 exchange is assumed, the resistant fraction appears misleadingly to represent a more stable form. Another contribution to the protected fraction can come from molecules that reach the native state before the labeling pulse. These factors have generally not been considered in previous pulse labeling work.

Analysis shows that the N/C bi-helical intermediate observed here is a distinctly structured form with defined stability that is homogeneously shared by essentially all of the refolding population. One appreciates that conformation in unstructured regions must be grossly heterogeneous from one chain to another, but the structured regions are essentially the same in conformation and stability. Some folding models picture that the heterogeneous nature of multiple trapped intermediates is most pronounced early on the funnel-like folding landscape where conformational freedom is large, and much more restricted near the bottom of the funnel where conformations merge to the native form. The present intermediate is distinctly structured, homogeneous, and native-like even though it is an early one.

Pathway heterogeneity and optional misfolding barriers

Other results in the literature seemed to show heterogeneity in folding pathways. Different population fractions are often seen to refold at different rates, taken to imply that proteins fold through multiple alternative pathways.

A different possibility[32] is that all of the molecules fold by way of the same sequence of structure formation (intermediates), but that some fraction of the population encounters an optionally inserted barrier (misfold). When optional misfolding barriers are inserted into a pathway with probability between 0 and 1, different fractions of a refolding population will encounter a limiting barrier at different points along the pathway, populate different intermediates, and reach N at different rates.

This occurs in Cyt c where a misligation-dependent barrier can be inserted into the folding sequence of some variable fraction of the population simply by adjusting the pH of the initially unfolded protein, even though folding is always done at the same low pH value.[32] Molecules with the misligation encounter the misfolding barrier, populate the N/C intermediate, and fold in a slow three-state manner. The rest of the population folds in a fast two-state manner.[4,32,72,73] Folding appears to be heterogeneous. However, a great deal of data now support a single dominant pathway that can be variably interrupted by an optionally inserted error repair barrier.[7,25,33]

Analogous examples in other proteins depend on the fractional mis-isomerization of proline residues,[4,73] alternative disulphide bond formation,[4] alternative docking modes for different folding domains,[4] and formation of non-native hydrophobic clusters.[66] These kinds of probabilistically inserted barriers have often been discussed in terms of parallel pathways. The same results might be better described as single pathways that are variably interrupted by optionally inserted barriers.

Implications

The present experiments trap and characterize a folding intermediate during kinetic folding. The intermediate resembles a partially constructed native state. The entire refolding population passes through it on the way to N. In agreement, a quantity of native state HX results indicate that Cyt c and some other proteins fold in a stepwise manner through a small series of native-like intermediates, sequentially adding cooperative native-like elements to progressively assemble the final native structure. These results point to a classical, stepwise, folding pathway.

From all of these experiments some principles have emerged. The unit folding steps are determined by the intrinsically cooperative nature of the secondary structural elements of the native structure or pairs thereof, termed foldons. The pathway sequence is determined by the arrangement of the foldons in the native protein; prior native-like structure guides and stabilizes the subsequent formation of adjacent units (sequential

stabilization[25,74]). In this process, easily reversible non-native interactions may commonly serve to minimize the energy of transient forms.[48] Less rapidly reversible interactions may insert kinetic barriers that slow folding and cause intermediate accumulation. These principles can be seen as a natural consequence of thoroughly established protein structural behavior that seem unlikely to differ fundamentally in other proteins.

However, many studies have been interpreted differently, in terms of multiple heterogeneous intermediates in multiple parallel pathways. In the case of HX pulse labeling experiments, the present work shows that these conclusions have depended on incorrect assumptions about HX behavior and some other experimental problems. We note that other studies often interpreted in terms of heterogeneous pathways might be reconsidered in terms of optionally inserted error-dependent barriers.

Materials and Methods

The purity of WT equine Cyt c (type VI; Sigma Chemical Co.) was checked and further purified by reverse phase HPLC when necessary.[75] The pH buffers used were of highest quality 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 (for 695 nm absorbance and tryptophan fluorescence) or an Aviv stopped-flow module attached to an AVIV CD spectrometer (model 202). Buffers and pH values were the same as those in HX pulse labeling experiments.

The following HX pulse labeling procedures were used. Cyt c was unfolded in 4.2 M d-GdmCl in 10 mM phosphate, (p^2H_{read} 7.5), ^2H_2O buffer. Refolding was initiated by diluting 18 times into the folding buffer (0.23 M GdmCl (pH 6), 10 mM Mes, H_2O). Protein concentration during folding was about 100 μM. Folding was continued for 100 ms. A high-pH pulse (pH 7.5–10) was 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)). Ascorbate was used to reduce Cyt c and pH 5.3 minimizes background labeling during the sample workup.

Final samples were concentrated (15 ml–0.5 ml; 4 °C; Amicon centriprep YM-10), then buffer-exchanged (Sephadex spin columns pre-equilibrated with NMR buffer; 100 mM deuterated acetate, 12 mM ascorbate, ^2H_2O, p^2H_{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 value and the data were averaged.

Data analysis

The 2D 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). NH–C^αH cross-peak volumes for each amide were calculated to obtain the fractional H labeling, H_{obs}, as in equation (1):

$$H_{obs} = \frac{V/V^{ref}}{V_C/V_C^{ref}} \quad (1)$$

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 native protein control sample equilibrated (pH 7.5, 55 °C, three hours) to the same H/^{2}H ratio as during the pulse, representing 100% labeling. H_{obs} and the analogous terms considered below vary 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). We want to remove the extraneous contributions and extract the information, structural, equilibrium, and kinetic, that is contained in the way that the intermediate alone becomes labeled during the pulse. The two extraneous contributions, H_{bkgd} and $H_{N,pulse}$, were evaluated in control experiments.

A no-pulse control was used to determine the summed background labeling at each amide during the entire experiment except during the pulse. In this control experiment, the stopped-flow mixing and sample workup were done exactly as before but with the labeling pulse omitted (100 ms folding at pH 6, no pulse, quench to pH 5.3, workup at pH 5.3). The fractional labeling obtained as in equation (1) is H_{bkgd}, which is dominated by the labeling that occurs during the sample workup at sites that exchange significantly in the native protein. H_{obs} for the total experiment can then be expressed as in equation (2):

$$H_{obs} = H_{pulse} + (1 - H_{pulse})H_{bkgd} \quad (2)$$

H_{pulse} includes mainly the fractional labeling of the intermediate during the pulse ($H_{I,pulse}$), which is the experimental parameter of major interest. It also includes the labeling in the 15% of native protein (f_N) present during the pulse ($H_{N,pulse}$), as in equation (3):

$$H_{pulse} = (1 - f_N)H_{I,pulse} + f_N H_{N,pulse} \quad (3)$$

To evaluate $H_{N,pulse}$, pre-deuterated native protein was carried through the entire pulse labeling procedure (in 0.23 M d-GdmCl at each pH) to obtain $H_{N,obs}$ (like equation (1)). A no-pulse control experiment was run with the native protein to obtain $H_{N,bkgd}$. These parameters determine $H_{N,pulse}$ (like equation (2)). $H_{N,pulse}$ 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 omega loops but this correction is multiplied by f_N (equation (3)). (Note: this control experiment is identical with the kinetic native state HX experiment described elsewhere,[26] which measured the stability and kinetic parameters for unfolding of the loops in the native protein).

Equations (2) and (3) together with the independently evaluated parameters, H_{bkgd}, $H_{N,pulse}$, and f_N, connect the measured H_{obs} to the desired $(1 - f_N)H_{I,pulse}$. 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 amide-resolved opening and closing rates in the intermediate. Accuracy improves as the term $H_{I,pulse}$ becomes the dominant contributor to H_{obs} and the other contributions decrease. This condition was promoted by the experimental manipulations described above.

HX equations

The Linderstrøm-Lang scheme for opening dependent HX was assumed:[76]

$$\mathrm{NH(closed)} \underset{k_{cl}}{\overset{k_{op}}{\rightleftarrows}} \mathrm{NH(open)} \xrightarrow{k_{ch}} \mathrm{exchanged} \qquad (4)$$

As usual, k_{op} and k_{cl} are the opening (unfolding) and closing (refolding) rates of the protecting structure, and k_{ch} is the unprotected chemical exchange rate calculated for the ambient conditions from model peptide calibrations.[43,44]

Analysis of the present HX pulse labeling data requires the complete equations given by Hvidt[42] (equations (5)–(7)), rather than the usual steady-state rate equation.[76,77] The parameters t_f and t_p are the folding (0.1 second) and pulse (0.05 second) times, and $[\mathrm{NH(open)}]_{t_f}$ is the population fraction in the open form during the pulse, after folding for time t_f. Equation (7) expresses $[\mathrm{NH(open)}]_{t_f}$ in these terms (assuming no protection in the unfolded state):

$$H_{I,pulse} = 1 - \left(\frac{k_{ch}[\mathrm{NH(open)}]_{t_f} - \lambda_2}{\lambda_1 - \lambda_2}\right) e^{-\lambda_1 t_p} - \left(\frac{\lambda_1 - k_{ch}[\mathrm{NH(open)}]_{t_f}}{\lambda_1 - \lambda_2}\right) e^{-\lambda_2 t_p} \qquad (5)$$

where:

$$\lambda_{1,2} = \frac{k_{op} + k_{cl} + k_{ch} \pm \sqrt{(k_{op} + k_{cl} + k_{ch})^2 - 4k_{op}k_{ch}}}{2} \qquad (6)$$

$$[\mathrm{NH(open)}]_{t_f} = \frac{k_{op}}{k_{op} + k_{cl}} + \frac{k_{cl}}{k_{op} + k_{cl}} e^{-(k_{op}+k_{cl})t_f} \qquad (7)$$

When $k_{cl} \gg k_{op}$, k_{ch}, and t_f is sufficiently long, equation (5) reduces to the usual steady-state rate equation. When these conditions are not met, the additional first bracketed term in equation (5) is necessary in order to account for the fraction of the population, NH(open), that is exposed when HX labeling is initiated. As in equation (7), this term is negligible when structure is very stable ($k_{cl} \gg k_{op}$), in which case [NH(open)] and $\lambda_2 \sim 0$, and/or when the initially exposed sites are rapidly protected before labeling can occur ($k_{cl} \gg k_{ch}$). This term is significant in the present work because the stability of the intermediate is low. Therefore, a significant fraction of the intermediate is unfolded when labeling is initiated and it can become labeled without having to wait for opening (neither EX1 nor EX2).

These equations are written without any assumptions about the relative magnitudes of k_{op}, k_{cl} and k_{ch} and therefore about the role of EX1 HX (k_{op}, $k_{cl} \ll k_{ch}$; $k_{ex} = k_{op}$) or EX2 HX ($k_{cl} \gg k_{ch}$, k_{op}; $k_{ex} = K_{op}k_{ch}$; $K_{op} = k_{op}/k_{cl}$). For more discussion about the application of these equations, see Krishna *et al*.[78]

To implement these equations, the basic experimental data for each amide (H_{obs} *versus* pulse pH; as in Figure 4) were fit to equation (2), substituting equation (3) for H_{pulse} and equations (5)–(7) for $H_{I,pulse}$. The k_{ch} values were calculated from peptide models[43,44] using the Excel spreadsheets†. The numerical parameters, H_{bkgd} and $H_{N,pulse}$, for each amide were fixed at the values determined in the control experiments. There are only two independent fitting parameters, k_{op} and k_{cl}. Figure 4 also shows these fitted curves after correction for the background contributions (colored broken lines). For each amide, the corrected plateau level (EX1 region) determines the opening rate (k_{op}); the offset from the expected unprotected curve (black) at lower pH (EX2 region) largely gives the equilibrium stability (K_{op}). Data fitting used Sigma Plot 2001.

The residue-resolved unfolding free energy of protecting structure was then obtained using (8):

$$\Delta G_{HX} = -RT \ln\left(\frac{k_{op}}{k_{cl}}\right) \qquad (8)$$

The errors in ΔG_{HX} (Figure 5(a)) and k_{cl}/k_{op} (Figure 3(b)) were calculated from the individual errors in k_{op} and k_{cl} obtained from data fitting using standard error propagation formulas.[79]

Acknowledgements

This work was supported by research grants from the NIH and the Mathers Foundation. We thank Yawen Bai for sharing his manuscript[48] before publication.

References

1. Wolynes, P. G., Onuchic, J. N. & Thirumalai, D. (1995). Navigating the folding routes. *Science*, **267**, 1619–1620.
2. Shakhnovich, E. I. (1997). Theoretical studies of protein-folding thermodynamics and kinetics. *Curr. Opin. Struct. Biol.* **7**, 29–40.
3. Dill, K. A. & Chan, H. S. (1997). From Levinthal to pathways to funnels. *Nature Struct. Biol.* **4**, 10–19.
4. Wallace, L. A. & Matthews, C. R. (2002). Sequential *vs*. parallel protein-folding mechanisms: experimental tests for complex folding reactions. *Biophys. Chem.* **101–102**, 113–131.
5. Levinthal, C. (1968). Are there pathways for protein folding. *J. Chim. Phys.* **65**, 44–45.
6. Kim, P. S. & Baldwin, R. L. (1990). Intermediates in the folding reactions of small proteins. *Annu. Rev. Biochem.* **59**, 631–660.
7. Englander, S. W. (2000). Protein folding intermediates and pathways studied by hydrogen exchange. *Annu. Rev. Biophys. Biomol. Struct.* **29**, 213–238.
8. 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.
9. 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.
10. Udgaonkar, J. B. & Baldwin, R. L. (1990). Early folding

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

intermediate of ribonuclease A. *Proc. Natl Acad. Sci. USA*, **87**, 8197–8201.
11. Elöve, G. A. & Roder, H. (1991). Structure and stability of cytochrome c folding intermediates. In *Protein Refolding* (Georgiou, G. & Bernardez-Clark, E. D., eds), pp. 50–63, American Chemical Society, Washington, DC.
12. Englander, S. W. & Mayne, L. (1992). Protein folding studied using hydrogen-exchange labeling and two-dimensional NMR. *Annu. Rev. Biophys. Biomol. Struct.* **21**, 243–265.
13. Baldwin, R. L. (1993). Pulsed H/D-exchange studies of folding intermediates. *Curr. Opin. Struct. Biol.* **3**, 84–91.
14. Dobson, C. M., Evans, P. A. & Radford, S. E. (1994). Understanding how proteins fold: the lysozyme story so far. *Trends Biochem. Sci.* **19**, 31–37.
15. Evans, P. A. & Radford, S. E. (1994). Probing the structure of folding intermediates. *Curr. Opin. Struct. Biol.* **4**, 100–106.
16. Woodward, C. K. (1994). Hydrogen exchange rates and protein folding. *Curr. Opin. Struct. Biol.* **4**, 112–116.
17. Dyson, H. J. & Wright, P. E. (1996). Insights into protein folding from NMR. *Annu. Rev. Phys. Chem.* **47**, 369–395.
18. Li, R. & Woodward, C. (1999). The hydrogen exchange core and protein folding. *Protein Sci.* **8**, 1571–1591.
19. Bai, Y., Sosnick, T. R., Mayne, L. & Englander, S. W. (1995). Protein folding intermediates: native-state hydrogen exchange. *Science*, **269**, 192–197.
20. Bai, Y. & Englander, S. W. (1996). Future directions in folding: the multi-state nature of protein structure. *Proteins: Struct. Funct. Genet.* **24**, 145–151.
21. Chamberlain, A. K. & Marqusee, S. (2000). Comparison of equilibrium and kinetic approaches for determining protein folding mechanisms. *Advan. Protein Chem.* **53**, 283–328.
22. Xu, Y., Mayne, L. & Englander, S. W. (1998). Evidence for an unfolding and refolding pathway in cytochrome *c*. *Nature Struct. Biol.* **5**, 774–778.
23. 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.
24. Bai, Y. (1999). Kinetic evidence for an on-pathway intermediate in the folding of cytochrome *c*. *Proc. Natl Acad. Sci. USA*, **96**, 477–480.
25. 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.
26. 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.
27. 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.
28. Shastry, M. C. R. & Roder, H. (1998). Evidence for barrier-limited protein folding kinetics on the microsecond time scale. *Nature Struct. Biol.* **5**, 385–392.
29. 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.
30. Akiyama, S., Takahashi, S., Ishimori, K. & Morishima, I. (2000). Stepwise formation of α-helices during cytochrome *c* folding. *Nature Struct. Biol.* **7**, 514–520.
31. 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.
32. Sosnick, T. R., Mayne, L., Hiller, R. & Englander, S. W. (1994). The barriers in protein folding. *Nature Struct. Biol.* **1**, 149–156.
33. Englander, S. W., Sosnick, T. R., Mayne, L. C., Shtilerman, M., Qi, P. X. & Bai, Y. (1998). Fast and slow folding in cytochrome *c*. *Accts Chem. Res.* **31**, 737–744.
34. Weissman, J. S. & Kim, P. S. (1995). A kinetic explanation for the rearrangement pathway of BPTI folding. *Nature Struct. Biol.* **2**, 1123–1130.
35. 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.
36. 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.
37. Bhuyan, A. K. & Udgaonkar, J. B. (2001). Folding of horse cytochrome *c* in the reduced state. *J. Mol. Biol.* **312**, 1135–1160.
38. 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.
39. Osterhout, J. J. & Nall, B. T. (1985). Slow refolding kinetics in yeast iso-2 cytochrome *c*. *Biochemistry*, **24**, 7999–8005.
40. 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.
41. 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.
42. Hvidt, A. (1964). A discussion of the pH dependence of the hydrogen–deuterium exchange of proteins. *C.R. Trav. Lab. Carlsberg*, **34**, 299–317.
43. 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.
44. 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.
45. 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.
46. Arrington, C. B. & Robertson, A. D. (2000). Kinetics and thermodynamics of conformational equilibria in native proteins by hydrogen exchange. *Methods Enzymol.* **323**, 104–124.
47. Capaldi, A. P., Kleanthous, C. & Radford, S. E. (2002). Im7 folding mechanism: misfolding on a path to the native state. *Nature Struct. Biol.* **9**, 209–216.
48. 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*, in the press.
49. Dyson, H. J., Rance, M., Houghten, R. A., Lerner, R. A. & Wright, P. E. (1988). Folding of immunogenic peptide fragments of proteins in water solution. I. Sequence requirements for the formation of a reverse turn. *J. Mol. Biol.* **201**, 161–200.
50. Dyson, H. J., Bolinger, L., Feher, V. A., Osterhout, J. T.,

Jr, Yao, J. & Wright, P. E. (1998). Sequence requirements for stabilization of a peptide reverse turn in water solution. Proline is not essential for stability. *Eur. J. Biochem.* **255**, 462–471.
51. Colón, W., Elöve, G. A., Wakem, P., Sherman, F. & Roder, H. (1996). Side chain packing of the N- and C-terminal helices play a critical role in the kinetics of cytochrome *c* folding. *Biochemistry*, **35**, 5538–5549.
52. Lifson, S. & Roig, A. (1961). On the theory of the helix-coil transition in polypeptides. *J. Chem. Phys.* **34**, 1963–1974.
53. Qian, H. & Schellman, J. A. (1992). Helix-coil theories: a comparative study for finite length polypeptides. *J. Phys. Chem.* **96**, 3987–3994.
54. Chakrabartty, A., Kortemme, T. & Baldwin, R. L. (1994). Helix propensities of the amino acids measured in alanine-based peptides without helix-stabilizing side-chain interactions. *Protein Sci.* **3**, 843–852.
55. Padmanabhan, S., York, E. J., Stewart, J. M. & Baldwin, R. L. (1996). Helix propensities of basic amino acids increase with the length of the side-chain. *J. Mol. Biol.* **257**, 726–734.
56. Muñoz, V. & Serrano, L. (1994). Elucidating the folding problem of helical peptides using empirical parameters. *Nature Struct. Biol.* **1**, 399–409.
57. Lacroix, E., Viguera, A. R. & Serrano, L. (1998). Elucidating the folding problem of α-helices: local motifs, long-range electrostatics, ionic strength dependence and prediction of NMR parameters. *J. Mol. Biol.* **284**, 173–191.
58. Roder, H., Wagner, G. & Wuthrich, K. (1985). Amide proton exchange in proteins by EX_1 kinetics: studies of basic pancreatic trypsin inhibitor at variable p^2H and temperature. *Biochemistry*, **24**, 7396–7407.
59. Pedersen, T. G., Thomsen, N. K., Andersen, K. V., Madsen, J. C. & Poulsen, F. M. (1993). Determination of the rate constants k_1 and k_2 of the Linderstrøm-Lang model for protein amide hydrogen exchange. A study of individual amides in hen egg-white lysozyme. *J. Mol. Biol.* **230**, 651–660.
60. Arrington, C. B. & Robertson, A. D. (2000). Microsecond to minute dynamics revealed by EX1-type hydrogen exchange at nearly every backbone hydrogen bond in a native protein. *J. Mol. Biol.* **296**, 1307–1317.
61. Canet, D., Last, A. M., Tito, P., Sunde, M., Spencer, A., Archer, D. B. *et al.* (2002). Local cooperativity in the unfolding of an amyloidogenic variant of human lysozyme. *Nature Struct. Biol.* **9**, 308–315.
62. 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.
63. Sivaraman, T., Arrington, C. B. & Robertson, A. D. (2001). Kinetics of unfolding and folding from amide hydrogen exchange in native ubiquitin. *Nature Struct. Biol.* **8**, 331–333.
64. Duan, Y. & Kollman, P. A. (1998). Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution. *Science*, **282**, 740–744.
65. Penel, S. & Doig, A. J. (2001). Rotamer strain energy in protein helices—quantification of a major force opposing protein folding. *J. Mol. Biol.* **305**, 961–968.
66. Chowdary, S., Zhang, W., Wu, C., Xiong, G. & Duan, Y. (2003). Breaking non-native hydrophobic clusters is the rate-limiting step in the folding of an alanine-based peptide. *Biopolymers*, **68**, 63–75.
67. Kussell, E., Shimada, J. & Shakhnovich, E. I. (2003). Side-chain dynamics and protein folding. *Proteins: Struct. Funct. Genet.* **52**, 303–321.
68. Chin, D.-H., Woody, R. W., Rohl, C. A. & Baldwin, R. L. (2002). Circular dichroism spectra of short, fixed-nucleus alanine helices. *Proc. Natl Acad. Sci. USA*, **99**, 15416–15421.
69. 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.
70. 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.
71. Shtilerman, M., Lorimer, G. H. & Englander, S. W. (1999). Chaperonin function: folding by forced unfolding. *Science*, **284**, 822–825.
72. Wildegger, G. & Kiefhaber, T. (1997). Three-state model for lysozyme folding: triangular folding mechanism with an energetically trapped intermediate. *J. Mol. Biol.* **270**, 294–304.
73. 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. *J. Mol. Biol.* **323**, 309–325.
74. Englander, S. W., Mayne, L. & Rumbley, J. N. (2002). Submolecular cooperativity produces multi-state protein unfolding and refolding. *Biophys. Chem.* **101–102**, 57–65.
75. 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.
76. Linderstrøm-Lang, K. (1958). Deuterium exchange and protein structure. In *Symposium on Protein Structure* (Neuberger, A., ed.), pp. 23–24, Methuen, London.
77. Hvidt, A. & Nielsen, S. O. (1966). Hydrogen exchange in proteins. *Advan. Protein Chem.* **21**, 287–386.
78. Krishna, M. M. G., Hoang, L., Lin, Y. & Englander, S. W. (2003). Hydrogen exchange methods to study protein folding. *Methods*, in the press.
79. Bevington, P. R. & Robinson, D. K. (1994). *Data Reduction and Error Analysis for the Physical Sciences*, 2nd edit., McGraw-Hill, New York.
80. 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.
81. Kraulis, P. J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallog.* **24**, 945–949.

Edited by P. Wright

(Received 19 June 2003; received in revised form 26 September 2003; accepted 26 September 2003)