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

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

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.

*Corresponding author

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, 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.

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
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\(^1\) (RNase A) and cytochrome c\(^1\) (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\(^1\) and histidine to heme misligation in Cyt c\(^1\)\(^2\)–\(^5\) (see Figure 1(c)). Sosnick et al.\(^6\) 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\(^7\) 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 fraction of the misfolded intermediate and is transiently blocked. Intermediate formation itself is asynchronous; different population fractions form given structures at different rates. Baldwin\(^7\) 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 act to slow rather than promote folding. 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.\(^8\)–\(^21\),\(^26\)–\(^32\) 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 folding 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 \(\mathrm{H}_2\mathrm{O}\) at \(\mathrm{pD}_{\text{total}}\) 7.5. Folding was initiated by a large dilution into \(\mathrm{H}_2\mathrm{O}\) buffer at \(\mathrm{pH}\) 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
kinetics (695 nm absorbance, Figure 2(c); fit by $\tau$ 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 ($\tau$ = 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 $\approx 6$),$^{42}$ 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 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.

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...
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 (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 (Figure 4(b)).

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, CI2, bovine growth hormone, interleukin 1β, lysozyme, α-lactalbumin, phosphoglycerate kinase, myoglobin, barnase and ubiquitin.

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 2H2O was diluted into H2O 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 (tHX ~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–2H labeling profile imprinted during the sub-second labeling pulse. Samples were concentrated (centrifugal filtration) and the H–2H 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 (pKa ~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 (Hobs in equation (2)). 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 kop and kcl, and thus Kop and ΔGop. 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 (kch ! kcl; kex ~Kopkch). Accordingly, the horizontal displacement of the
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 amide hydrogen atoms in all of the helical and \( \Omega \) 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 \( \Omega \) loops show an average \( \Delta G_{op} \) of 1.0(±0.5) kcal/mol. For comparison, the blue and green helices in the native protein have a \( \Delta G_{HX} \) value of 12.6 kcal/mol and 9.8 kcal/mol, respectively, and the \( \Omega \) 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 \( \Omega \)-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 \( \Omega \)-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 ($H_{\text{obs}}$, 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, $k_{\text{op}}$ and $k_{\text{cl}}$). The black broken curves show the labeling that would be expected for each amide if it were wholly unprotected in the intermediate ($[1-f_{\text{cl}}]H_{\text{l, unprotected}}$, 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 ($[1-f_{\text{cl}}]H_{\text{l,pulse}}$, equation (5)) after the measured data ($H_{\text{obs}}$) are corrected for extraneous contributions ($H_{\text{bkgd}}$ and $H_{\text{N,pulse}}$). The fitting of the equations to the original data implicitly includes these corrections. For residues that show incomplete labeling, only the curve for $(1-f_{\text{cl}})H_{\text{l,pulse}}$ is shown.
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

---

**Figure 6.** Residue-resolved equilibrium and kinetic parameters in the trapped intermediate obtained by fitting the data in Figure 5. $\Delta G_{op}$ 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 $\Delta G_{op}$ value observed over the three $\Omega$-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 $\Omega$-loops.
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,66,69 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 equilibrium66,67,69 and kinetic24,63 modes:

$$\text{NRYGB (native)} \leftrightarrow \text{nRYGB} \leftrightarrow \text{nrYGB} \leftrightarrow \text{nryGB}$$

$$\leftrightarrow \text{nryGB} \leftrightarrow \text{nrygb (unfolded)}$$

(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.50,69,70 The subsequent identification of the nested-yellow foldon placed it even earlier in the unfolding sequence69 (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 nRYGB68). Stability labeling HX experiments55,61 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 $H_2$,82 apo Cyt $b$,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 misigate 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-workers91,92 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. 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 (ΔG‡, ΔH‡, ΔS‡) and structural (mT) parameters. 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. 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.

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 barrier (B1) that is intrinsic to the folding process and an error-dependent barrier (B2) that is optional rather than intrinsic.

**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.

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. The pathway is essentially linear because of the way that the different foldons are arranged in the native protein. 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. In three-state folding, the entire protein population traverses the same earlier barriers (this work and Sosnick et al.) 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.). 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. 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. 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.

The pH buffers used were the highest grade available from Sigma. H2O was from Isotec or Aldrich. Deuterated GdmCl (d-GdmCl) was prepared by dissolving in 2H2O 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 SFM-400 apparatus (695 nm absorbance and fluorescence). Cyt c were measured at 10 °C and vacuum evaporating three times. Folding kinetics of the initial barrier (B1) is rate-limiting. In three-state kinetic folding, the Cyt c intermediates are invisible because the initial barrier (B1) is rate-limiting. In three-state folding, the entire protein population traverses the same earlier barriers (this work and Sosnick et al.) 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.). 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. 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. 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.

The pH buffers used were the highest grade available from Sigma. H2O was from Isotec or Aldrich. Deuterated GdmCl (d-GdmCl) was prepared by dissolving in 2H2O 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.
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

$H_{\text{pulse}}$ includes mainly the fractional labeling of the intermediate during the pulse ($H_{\text{pulse}}$), 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 ($H_{\text{N, pulse}}$), as expressed in equation (4) (42% = 39% fast phase + 3% of the slower phase). $H_{\text{N, pulse}}$ at each pulse pH was calculated from the two native controls $H_{\text{N, obs}}$ and $H_{\text{N, bkgd}}$ (like equation (3)). $H_{\text{N, pulse}}$ 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)).

$$H_{\text{pulse}} = f_N H_{\text{pulse}} + f_H H_{\text{pulse}} + f_N H_{\text{N, pulse}}$$

In the present case, $f_N \approx 0$ (Figure 2(a)). Hence $f_H$ in equation (4) becomes $(1-f_N)$. Equations (3) and (4) together with the independently evaluated parameters, $H_{\text{bkgd}}$, $H_{\text{N, pulse}}$ and $f_N$, connect the measured $H_{\text{obs}}$ to the desired $(1-f_N)H_{\text{pulse}}$:

$$(1-f_N)H_{\text{pulse}} = H_{\text{obs}} - H_{\text{bkgd}} - f_N H_{\text{N, pulse}}$$

($H_{\text{bkgd}}$ 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 $H_{\text{pulse}}$ can be obtained. The equations described below relate $H_{\text{pulse}}$ to opening and closing (deprediction and repropagation) rates. These relationships allow experimental pulse labeling data for $H_{\text{obs}}$ 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 $(1-f_N)H_{\text{pulse}}$ with that expected when there is no protection in the unfolded state, calculated as:

$$(1-f_N)H_{\text{unprotected}} = (1-f_N)(1 - \exp(-k_{ch} \tau_p))$$

where $\tau_p$ (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:25

$$\begin{align*}
&k_{op} \quad \text{NH(open)} \\
&k_{cl} \quad \text{NH(closed)} \quad \text{exchanged}
\end{align*}$$

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 $\tau_i$ 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 Hvidt24,60,96 (equations (8)–(10)):

$$H_{\text{pulse}} = 1 - \left(\frac{k_{op}[\text{NH(open)}]_{\text{bkgd}} - \lambda_2}{\lambda_1 - \lambda_2}\right) e^{-\lambda_1 \tau_p}$$

$$- \left[\frac{1}{\lambda_1 - \lambda_2}\right] \tau_p$$

where:

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

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

$$[\text{NH(open)}]_{\text{bkgd}} = \frac{k_{op} + k_{cl}}{k_{op} + k_{cl} + k_{ch}} e^{-k_{ch} \tau_i}$$

To implement these equations, the basic experimental data for each amide ($H_{\text{obs}}$ versus pulse pH; Figure 5) were fit to equation (3) using equation (4) for $H_{\text{pulse}}$ and equations (7)–(10) for $H_{\text{pulse}}$, with $H_{\text{bkgd}}$ and $H_{\text{N, pulse}}$ fixed at the values determined in the control experiments described. Alternatively, one can calculate $H_{\text{pulse}}$ 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):

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

The errors in $\Delta G_{\text{HX}}$ were calculated from the individual errors in $k_{op}$ and $k_{cl}$ obtained from data fitting using standard error propagation formula.92

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


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


