Fast and Slow Folding in Cytochrome *c*

S. WALTER ENGLANDER,* TOBIN R. SOSNICK, LELAND C. MAYNE, MARK SHTILERMAN, PHOEBE X. QI, AND YAWEN BAI

Department of Biochemistry & Biophysics, The Johnson Research Foundation, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania 19104

FIGURE 1. Diagram of equine cyt *c* (using Molscript[37]) showing residues and structural elements noted in the text. The color coding indicates the measured free energy level of the cooperative unfolding units found by native state HX[5,31,38] (red to blue in order of ascending energy).

Introduction

A number of puzzling characteristics appear repeatedly in protein folding studies. Some proteins fold in a fast two-state manner with no apparent intermediates while others fold much more slowly in a multistate way. Many proteins fold heterogeneously; within the same controlled experiment some fraction of the population reaches the native state rapidly and others more slowly. These contradictory behaviors have made it difficult to discern general folding principles.

Cytochrome *c* (cyt *c*, equine), diagrammed in Figure 1, offers a number of advantages for folding studies. In addition to the usual folding probes (circular dichroism, hydrogen exchange protection, etc.), cyt *c* provides intrinsic optical probes that detect and characterize early and late folding events. The fluorescence of the single tryptophan, Trp 59, is sensitively quenched by Förster transfer to the covalently held heme in early chain contraction. The appearance of an optical absorbance band at 695 nm signals the final step in native state acquisition, the Met 80-S to heme iron ligation. Another major advanta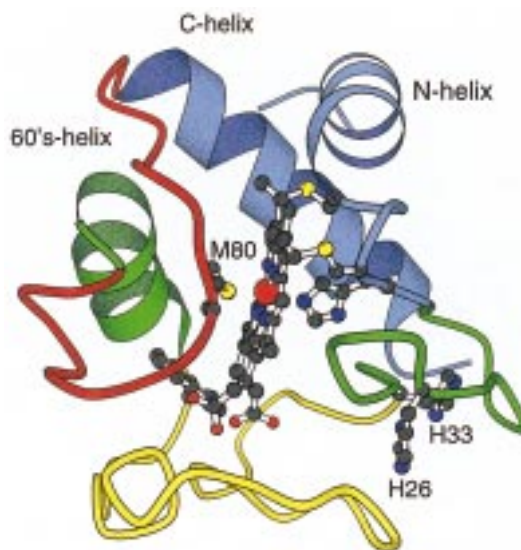ge is the ability to produce fast two-state folding or slower three-state folding at will by removing or inserting a defined kinetic barrier (Figure 2).

Figure 2 illustrates the important kinetic events that can be measured in folding experiments initiated by diluting cyt *c* from denaturing guanidinium chloride (GmCl). A submillisecond burst phase contraction (arrows) is seen in both two-state and three-state folding, followed by a further chain condensation on a millisecond time scale (fluorescence quenching probe). Under two-state folding conditions, the native state is reached by the majority of the protein population (695 nm absorbance probe) on the same time scale as the millisecond chain condensation and intermediates are not seen. Under three-state conditions, the protein pauses after the millisecond condensation and an intermediate accumulates before folding can continue more slowly to the native state. Figure 2 also shows that cyt *c* folding is heterogeneous in both conditions. These events can be explained in terms of four important folding processes, as follows.

(1) Burst phase chain contraction.[1–3] The submillisecond burst phase represents an energetically downhill polymer-like chain contraction that readjusts the unfolded ensemble from its more extended distribution in a good solvent (high denaturant) to the more contracted ensemble characteristic for low denaturant.

(2) Search–nucleation–collapse.[1] The first step inherent to the folding program is a free-energy uphill conformational search to assemble a transition state configuration that can act to nucleate smaller forward folding steps in an energetically downhill manner. Successful nucleation is accompanied by a massive chain collapse that becomes more apparent at higher denaturant where the

S. Walter Englander received M.S. and Ph.D. degrees in biophysics at the University of Pittsburgh. He is a Professor of Biochemistry & Biophysics and the Jacob Gershon-Cohen Professor of Medical Science at the University of Pennsylvania and is a member of the National Academy of Sciences.

Tobin R. Sosnick received a Ph.D. degree in applied physics at Harvard and did postdoctoral work at Los Alamos and the University of Pennsylvania. He is an Assistant Professor in Biochemistry and Molecular Biology at the University of Chicago.

Leland C. Mayne received a Ph.D. degree in chemistry at the University of Oregon. He is a Research Assistant Professor of Biochemistry & Biophysics at the University of Pennsylvania and Assistant Director of the Johnson Research Foundation.

Mark Shtilerman received a B.S. degree in physics from St. Petersburg University, Russia. He is a graduate student at the University of Pennsylvania.

Phoebe X. Qi received a Ph.D. degree in biochemistry at the University of Illinois in Urbana. She is a postdoctoral researcher at the University of Pennsylvania.

Yawen Bai received M.S. and Ph.D. degrees in chemistry and biophysics at the University of Pennsylvania and did postdoctoral work at the Scripps Research Institute. He is a staff scientist at the NIH.

* To whom correspondence should be addressed. Phone: (215) 898-8042. Fax: (215) 898-4217. E-mail: Walter@HX2.Med.UPenn.Edu.
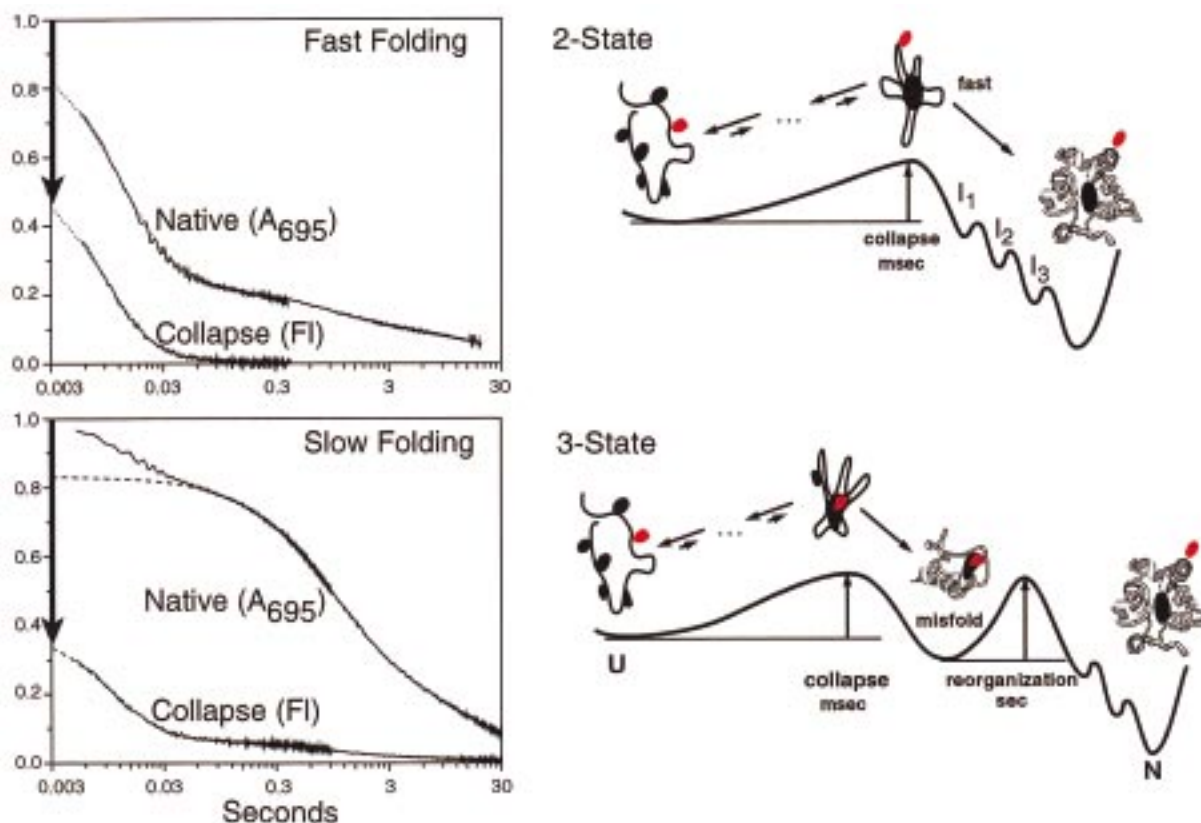
FIGURE 2. Kinetic folding behavior seen when cyt *c* is diluted from denaturing GmCl in stopped-flow experiments. Fluorescence quenching measures chain contraction and collapse; absorbance at 695 nm measures final native state acquisition. Two-state vs three-state folding was manipulated by exposing the initially unfolded cyt *c* to pH 2.9 (upper) or pH 6.3 (lower). The folding condition was the same in both cases (pH 4.9, 10 °C, 0.7 M GmCl). The arrows indicate an unresolved submillisecond burst phase which signals a polymer-like chain contraction on dilution from a good to a poorer solvent and not the production of a distinct folding intermediate. The energy diagrams (right) interpret the kinetic behavior seen (see the text). In both cases productive folding is initiated by a time-requiring (millisecond) long-range search to find a coarsely nativelike nucleation transition state, formed by a set of side chain interactions that can support chain collapse and the energetically downhill formation of intermediates. Cyt *c* then proceeds through a small sequence of intermediates (I_i) determined by the stepwise cooperative folding of the units color coded in Figure 1. In two-state folding these latter steps are fast, folding is limited by the initial search−nucleation, and intermediates do not accumulate. However, an adventitious misfolding error can insert a large reorganization barrier that slows folding and causes an intermediate to accumulate. The optional nature of misfolding produces folding heterogeneity; different molecules in the protein population fold faster or slower. An ambiguity in this reaction scheme is clarified in reactions 4 and 5 (text). The folding molecule diagrammed is meant to suggest the large-scale search in which a number of sequence-remote side chains find each other to produce the initial nucleating collapse. The native state topology of the transition state nucleus promotes secondary structure (Figure 1) formation and stabilization via fast small-scale searches. The incipient misfold that leads later to a reorganization barrier is suggested by the misligated histidine (red).

burst phase contraction is suppressed. The same search-dependent nucleation−collapse process on a time scale of milliseconds initiates both two-state and three-state folding.

(3) Intermediates.[4,5] After the initial nucleation, folding proceeds through a short sequence of programmed meta-stable nativelike intermediates to the native state (I_1 to I_3 in Figure 2; cooperative folding units color coded in Figure 1). When all goes well, the step from each intermediate to the next depends only on a more local search that is faster than the large-scale nucleation process. In this case the rate for reaching the final native state is limited by the initial search−nucleation, intermediates do not accumulate, and folding appears to be a two-state process.

(4) Misfold−reorganization barriers.[6] When all does not go well, folding errors can occur, resulting in slowed three-state folding and intermediate accumulation (Figure 2). With cyt *c* a misfolding error can be manipulated by a small change in the pH of the unfolding solution. At pH above about 6, a peripheral histidine residue can deprotonate and bind avidly to the unmasked heme iron,[7] replacing the weak Met 80-S ligand and forcing a segment of chain to the "wrong side" of the heme (see Figure 1). Upon GmCl dilution, the burst phase contraction, the millisecond chain condensation (nucleation step), and the initiation of intermediate formation occur as before, but then a new barrier is encountered (Figure 2, bottom). The misplaced segment, trapped by the chain condensation, must be freed before the programmed folding sequence can continue. The time-consuming chain reorganization acts as an inserted barrier that slows folding and causes the accumulation of the preceding intermediate. The optional nature of this barrier and other analogous ones produces folding heterogeneity.
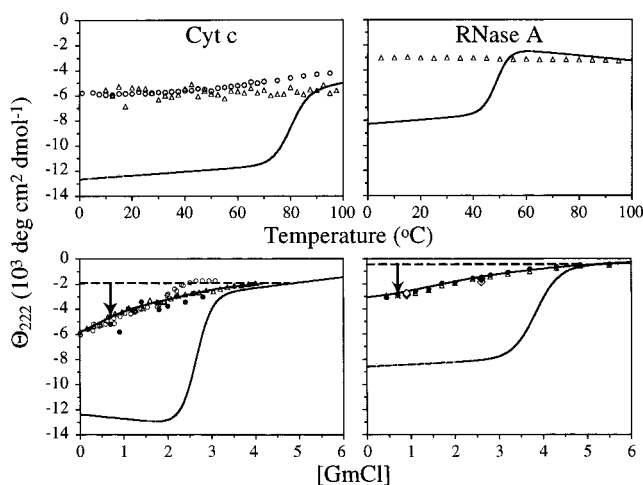
FIGURE 3. Equilibrium and kinetic behavior of cyt *c* (left panels[2]) and RNase A (right panels[3]) and their nonfolding polypeptides as a function of temperature and GmCl. The nonfolding polypeptides maintain the same CD (and fluorescence[1,2]) as the unfolded proteins at all temperatures and at high GmCl concentration, and they define the equilibrium parameters of the unfolded state ensemble under native conditions (low temperature and GmCl concentration). On dilution from high to low denaturant concentration, the burst phase signal starts from the CD value at high GmCl concentration (dashed line) and jumps to the end points shown by the data, as suggested by the arrows. In short, the burst phase simply readjusts the measured parameters to the unfolded equilibrium values characteristic for the lower GmCl concentration. Closed and open symbols are for data for the intact proteins and their nonfolding polypeptide analogues. The diamonds (RNase A) are for data from Houry et al.[39] (The preparation of fragment 1–80 from cyt *c* (circles) was contaminated with some nativelike species.)

This Account summarizes the experiments that have led to these hypotheses and places these events in the context of a coherent, relatively determinate folding pathway.

The Ultrafast Burst Phase

Submillisecond burst phase behavior like that in Figure 2 (arrows) is seen for many proteins and has generally been interpreted in terms of the fast formation of a productive folding intermediate. This interpretation appears to be incorrect. When nonfolding polypeptide chains are exposed to the same change in solvent, they produce the very same burst phase signals.

The circular dichroism (CD) results in Figure 3 show the melting behavior of intact cyt *c* (left panels) and intact ribonuclease A (RNase A; right panels). Two truncated fragments of cyt *c* (residues 1–65 and 1–80) were obtained by selective cyanogen bromide cleavage and subsequent purification. These fragments lack the carboxy-terminal segment, the most stable[8] and earliest folding[4] segment of cyt *c*. An unfolded analogue of RNase A was obtained by breaking the four disulfide bonds that are necessary to stabilize the native state. Figure 3 shows that these polypeptides exhibit the CD characteristic of the unfolded chain above the high-temperature melting transition and maintain this same CD under fully native conditions (low temperature, no denaturant). The lack of any structural

protection for the exchangeable amide hydrogens in these peptides under strongly native conditions confirms the absence of H-bonded structure.[2,9] Thus, they provide a good model for the unfolded state of cyt *c* and RNase A even at low temperature and denaturant concentration. We use them to define the unfolded CD equilibrium curves under native conditions, shown in Figure 3.

The residual CD of the unfolded peptides is removed noncooperatively by denaturant. Upon dilution from concentrated GmCl, the nonfolding polypeptides show a burst phase relaxation to new CD values (arrows in Figure 3) and then fold no further. The kinetic burst phase readjusts the nonfolding chains precisely to their previously determined equilibrium condition at the lower GmCl concentration. When the GmCl-unfolded proteins are similarly diluted, they also exhibit a burst phase change and then fold much more slowly to the native state. The interesting result is that the burst phase amplitude for the intact proteins quantitatively duplicates the burst phase seen for their demonstrably unfolded polypeptide analogues over the whole range of GmCl concentration. The same result was found for cyt *c* using its intrinsic fluorescence quenching probe, demonstrating a significant chain contraction in the burst phase.[1,2]

These results show that the condition reached at the end of the burst phase is the unfolded state ensemble characteristic for the low denaturant concentration used. The burst phase change occurs because the stopped-flow experiment starts at high denaturant concentration, with a differently biased unfolded state ensemble. Analogous good solvent to poor solvent contraction occurs for all polymers. Evidently polypeptides are no exception.

The Millisecond Barrier: Search–Nucleation–Collapse

Work with cyt *c* has identified a second and more fundamental kind of chain collapse, a massive chain condensation that forms the transition state for the millisecond kinetic process in Figure 2. This is a time-requiring nucleation process that serves as an obligatory *initial step* in folding and sets an ultimate limit for the folding rate.[1] Cyt *c* can be made to fold in either a two-state or a three-state manner. Under two-state conditions the transition state can be directly characterized. The insertion of a new barrier to produce three-state folding makes it possible to distinguish and characterize earlier and later events.

Fast Two-State Folding to the Native State. The upper left curve in Figure 2 exhibits fast cyt *c* folding. We found the same kinetic rate for a number of probes (fluorescence quenching, 695 nm absorbance, hydrogen exchange (HX) labeling and degree of HX protection, near and far UV CD, unfolding rate of the product).[6] The results show that cyt *c* can fold in milliseconds in a 2-state manner all the way to the native state.

Another discriminating test for two-state folding and further important information is provided by the kinetic folding chevron. The chevron refers to the V-shaped plot
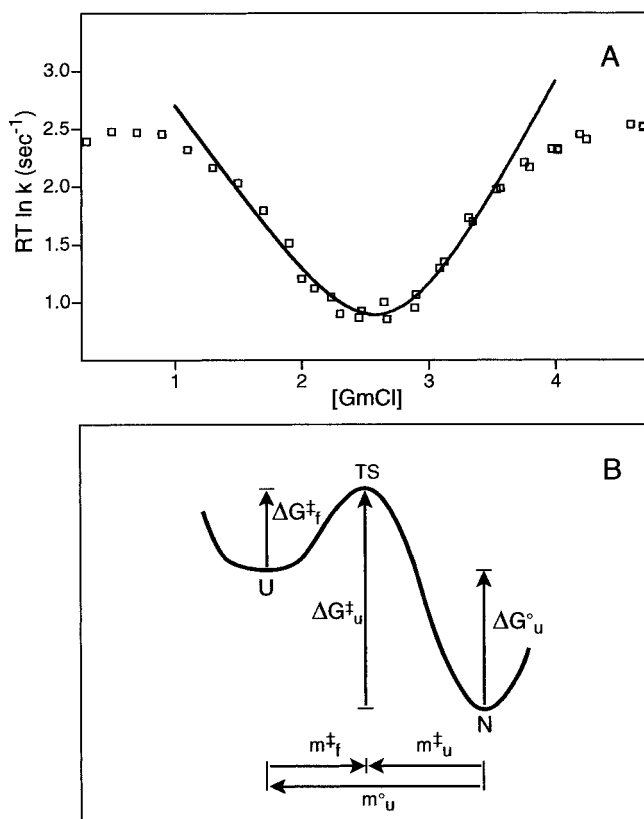
FIGURE 4. The folding chevron. Panel A shows the measured relaxation rate of cyt *c* against final GmCl concentration through the V-shaped chevron region (pH 4.9, 10 °C). ("Rollover" deviations occur at higher and lower GmCl concentrations.) Panel B illustrates the relationships that connect equilibrium and kinetic parameters when both are two-state, formalized in eqs 1–3.

of log(observed rate) vs denaturant concentration[10−12] shown in Figure 4A. The chevron criteria for two-state behavior are expressed in eqs 1 and 2 and illustrated in Figure 4B. Equation 1 requires that the same free energy

$$\Delta G^{\circ}_{u}(\text{den}) = \Delta G^{\ddagger}_{u}(\text{den}) - \Delta G^{\ddagger}_{f}(\text{den}) \qquad (1)$$

$$m^{\circ}_{u} = m^{\ddagger}_{u} - m^{\ddagger}_{f} \qquad (2)$$

$$\Delta G(\text{den}) = \Delta G(0) - m[\text{den}] \qquad (3)$$

measured for equilibrium unfolding (ΔG°_{u}) must be shown by the difference between the unfolding and refolding activation free energies at the same denaturant (den) concentration. Equation 2 is an analogous criterion in terms of the *m* parameter which relates to the denaturant-sensitive surface newly exposed or buried in each reaction ($m^{\circ}_{u} = \Delta(RT \ln K_{u})/\Delta[\text{den}]$; $m^{\ddagger} = \Delta(RT \ln k)/\Delta[\text{den}]$). ΔG and *m* are related as in eq 3, for either equilibrium or kinetic parameters.

Equilibrium and kinetic folding data obtained for cyt *c* at the low pH condition that produces two-state folding (Figure 4A) match the chevron criteria. This confirms the two-state nature of cyt *c* folding within the range of GmCl concentrations where folding data fall on the chevron.

Nucleation: Energetically Uphill and Topologically Nativelike. As for cyt *c*, many proteins have been found

to match the chevron criteria, with equilibrium and activation parameters related as in eqs 1 and 2. These results dictate a picture like that in Figure 4B in which the transition state for two-state folding and unfolding is reached in an energetically uphill barrier-crossing process. The alternative is that two-state folding might represent an energetically downhill slide limited in rate by the equivalent of a frictional resistance due to the roughness of the reaction surface.[13,14] This would require a fortuitous match in many proteins between the denaturant dependence of equilibrium melting and the denaturant dependence of the kinetic "friction" in downhill and uphill folding and unfolding (eqs 1 and 2).

Once the transition state is reached, the uphill course of folding is reversed (see also Figure 2). The transition state allows new structure formation in an energetically downhill manner.

Here a consideration of the subsequent events in folding becomes pertinent. Increasing evidence indicates that folding intermediates exist, contain nativelike structure, and can be more stable than the unfolded state. (For a brief discussion of the current situation, see an exchange among A. R. Fersht, C. K. Woodward, and S.W.E. in *TIBS*, October 1998.) Figure 1 shows the cooperative units of structure that produce folding intermediates in cyt *c*, found by native state HX in an equilibrium mode[5] and in part by HX pulse labeling in a kinetic mode.[4] The transition state must function to direct the formation of and stabilize these nativelike intermediates. The ability to do this seems to require that the transition state itself adopts a configuration that is coarsely nativelike in a topological sense, as suggested diagrammatically in Figure 2.

An energetically uphill conformational search necessary to assemble some nativelike multicontact configuration that can serve the transition state function is by definition a nucleation process.[1] Nucleation processes have also been inferred from model simulations[15,16] and experimental studies.[17,18]

Nucleation: Massive Chain Collapse. The two-state nature of cyt *c* folding in the chevron region makes it possible to characterize the rate-limiting transition state by measuring the response of the folding rate to various conditions.

The dependence of folding rate on GmCl concentration[1] (Figure 4A) shows that a great deal of denaturant sensitive surface (primarily polar groups[19]) is buried in the folding transition state, about half of that buried in native cyt. The dependence of folding rate on temperature[1] points to a specific heat of activation (largely apolar burial[20]) that is ~60% of the equilibrium value. Thus, formation of the folding transition state is associated with a massive chain condensation that buries much polar and apolar molecular surface. A similar *amount* of surface burial occurs in other proteins.[1] It seems reasonable that the condensate acts to help stabilize the subsequent formation of elements of structure that would not be stable all by themselves, perhaps by providing exposed hydrophobic docking surface.

Major chain condensation requires the concerted association of sequence-remote residues. Apparently it is necessary to assemble multiple interactions in some coarsely nativelike configuration in order to offset the unfavorable loop closure entropy and uncompensated polar burial in the collapse. This requires a long-range time-consuming search. What is nucleated is the chain collapse itself.

The nucleated collapse limits the rate for reaching the native state when subsequent larger barriers do not intervene (see reorganization barriers below). The same kind of nucleation process presumably occurs at lower denaturant also, below the chevron region, but here the prior burst phase contraction (Figures 2 & 3) tends to obscure the organizing chain condensation.

Nucleation: The Initial Step in Folding. It has been widely believed that the rate-limiting barrier occurs late in folding.[21,22] This belief stems from two sources: (1) the molten globule view which suggests that the folding rate of already well-formed intermediates is limited by final tight side chain packing;[23] (2) the fact that the ratio $m^{\ddagger}_{\mathrm{f}}/m^{\circ}_{\mathrm{u}}$ is generally large.[24]

The molten globule to native state transition can be directly observed in cyt *c* by virtue of the special 695 nm absorbance band. When the low pH molten globule of cyt *c* is mixed into neutral pH, the transition to the native state is complete in less than 3 ms.[6] Thus, final side chain packing and interdigitation are not intrinsically slow processes. The rate limitation must be earlier.

When folding is diagrammed using *m* as the reaction coordinate as in Figure 4B, the large $m^{\ddagger}_{\mathrm{f}}/m^{\circ}_{\mathrm{u}}$ positions the transition state well to the right, so that the rate-limiting step appears to be late in folding. This view is misleading. The large ratio is due to the large $m^{\ddagger}_{\mathrm{f}}$, reflecting the massively collapsed nature of the early transition state, and also to the relatively small m°_{u} characteristic of the small proteins usually studied. It is more pertinent to evaluate the placement of the limiting transition state relative to the sequence of specific folding events.

The sequence of events in folding is obscured when folding appears to be two-state. In cyt *c* earlier and later folding events can be separated by inserting an optional barrier into the folding sequence. When unfolded cyt *c* is exposed to pH above 6, peripheral histidines in the neutral form can misligate to the heme iron, creating an incipient misfolding. Early events in folding occur as before, but the newly inserted misfolding barrier delays later steps. Events that occur before and after the inserted barrier can then be distinguished by direct experiment.

We compared the collapse-related two-state barrier with the initial barrier in three-state folding[6] (see Figure 2). The two barriers have the same activation parameters, including free energy, enthalpy, entropy, and specific heat. They also exhibit the same structural $m^{\ddagger}_{\mathrm{f}}$ parameter, indicating the amount of surface burial. Clearly the limiting two-state barrier represents the same kinetic process as the early barrier in three-state folding. There-

fore, the limiting two-state barrier must be placed before the histidine-dependent inserted barrier (diagrammed in Figure 2).

The inserted barrier produces three-state folding and causes a preceding intermediate to become populated. The intermediate contains the N- and C-terminal helices and no other structure.[4] The limiting two-state barrier is therefore placed during or earlier than the formation of the N/C helix intermediate. It is a very early step in the sequence of folding events.

Other observations suggest the absence of any folding event preceding the limiting two-state barrier. The adherence to the chevron criteria shows that no event that buries significant surface or accounts for any significant fraction of the folding free energy occurs before the major nucleation–collapse step (in the chevron region). The intrinsic Trp59 fluorescence quenching probe, which is exceedingly sensitive to chain extent, shows that no chain contraction occurs earlier than the major uphill chain collapse that forms the transition state (except for the nonspecific burst phase contraction).

These considerations indicate that the limiting two-state barrier, i.e., the nucleated collapse transition state, occurs at the very beginning of the folding sequence in both two-state and three-state folding.

Summary. Following early suggestions,[25,26] a role for nucleation has more recently been proposed on the basis of simulated model folding[15,16] and experiment.[1,17,18,27] Cyt *c* results[1,27] demonstrate the uphill nature of the nucleated transition state, its massively collapsed condition, and the significant time requirement for its formation and imply a long-range search for a nativelike transition state topology. The nucleation barrier occurs at the very beginning of the cyt *c* folding process. When subsequent larger barriers (misfold-reorganization barriers) are not encountered, the initial nucleation becomes the rate-limiting step for native state acquisition, subsequent intermediates are kinetically invisible, and folding *appears to be* a two-state process.

Misfold–Reorganization Barriers

Many proteins including cyt *c* at neutral pH fold on a ~1 s time scale in a multistate way. Until recently this rate was considered to be extraordinarily fast and was thought to be limited by a late intrinsic on-pathway barrier, perhaps the final molten globule to native state transition. A study of cyt *c*[6] led us to a different conclusion.

As noted above, the folding of cyt *c* can be manipulated by small changes in pH that select between apparent two-state and three-state folding behavior, conditioned by the histidine-dependent ligation of the heme iron in the unfolded state.[7,28] Under low-pH two-state conditions, cyt *c* can be made to fold all the way to the native state in milliseconds.[6] Other proteins can do the same. This shows that none of the obligatory steps in folding are intrinsically "slow" (seconds).

The parallels found between fast two-state folding and slow three-state folding are especially indicative. When
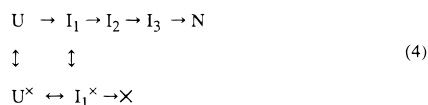
three-state folding is initiated, the initial millisecond barrier is traversed just as in fast two-state folding and these events appear thermodynamically and structurally identical (noted above). The difference appears later. In the three-state case, cyt *c* folding proceeds at the nucleation–collapse rate through formation of the N/C helix intermediate[4] (blue in Figure 1, I_1 in Figure 2) and then is blocked. Apparently the misligated histidine segment is trapped out of place by the fast molecular collapse, and this inserts a new barrier that makes folding slow. The resumption of forward folding requires some unfolding to free the trapped segment as indicated by an increase in the slow folding rate when GmCl concentration is increased (reverse GmCl dependence). These observations suggested[6] that the late barriers that make the folding of cyt *c* and other proteins much slower than milliseconds reflect nonobligatory off-pathway events.

Two influential reviews based on this work[22,29] favored an explanation derived from theoretical studies of non-protein models, termed the "new view" of protein folding. In this view, the coappearance of intermediates and slow folding and the fact of folding heterogeneity are all caused by obstructive off-pathway intermediates that form when the folding chain happens to become trapped in some deep energy well. On the contrary, we interpreted our results in terms of an optionally inserted reorganization barrier. In this view slow folding and intermediate accumulation occur together because they are both caused by the time-requiring reorganization needed to repair an optional misfolding error. The misfolding, in our view, is contained within an otherwise productive on-pathway intermediate.

Ambiguity arises concerning the nature of the misfolded intermediate and the barrier it produces. Does the observed intermediate reflect an off-pathway obstructive form that must be unfolded to allow folding to proceed, or is it a normally occurring on-pathway intermediate that is somehow corrupted? The view formalized below addresses this ambiguity.

A Coherent Folding Model

All of the results reviewed here can be understood in terms of a modified classical pathway (eq 4; Figure 2) in which intermediates are required and productive. The large-

$$U \rightarrow I_1 \rightarrow I_2 \rightarrow I_3 \rightarrow N$$
$$\updownarrow \quad \updownarrow \qquad\qquad (4)$$
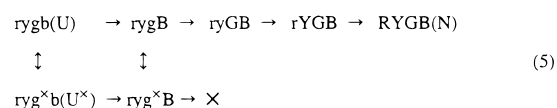$$U^\times \leftrightarrow I_1^\times \rightarrow \times$$

scale conformational search that finds the initial nucleating transition state is represented by the first reaction arrow in eq 4. (The U state pertinent for the folding condition is used, and not the high denaturant U form that leads to the burst phase readjustment in stopped-flow experimentation.) The chain can then move energetically downhill through a small series of obligatory and productive intermediate forms to the native state (upper path). When all goes well, each of these steps requires only a small-scale search that is faster than the initial

large-scale search–nucleation process, the intermediates are not seen, and the native state is reached in milliseconds in an apparent two-state process. However, misfolding errors that can mislocate and trap particular chain segments often occur (lower path).

The ambiguity noted above can be resolved by the view that each blocked $I_i^\times$ form, although "off *the fast* pathway", contains the same predetermined nativelike elements as the cognate on-pathway I_i form and *in addition* some misfolding that can stop or slow the forward progression. The misfolding error may require a time-consuming reorganization to return the chain to the fast track. The reorganization may or may not carry the chain back through U, and in any case is not a nucleation process. If the error-correcting reorganization is slower than the initial nucleation step, it will become rate-limiting, causing the accumulation of the corrupted intermediate and slow three-state folding. The optional nature of misfolding produces folding heterogeneity.

More concretely, native state HX results for cyt c^5 portray the cooperative folding units shown in Figure 1, referred to as red (R), yellow (Y), green (G), and blue (B). Accumulating evidence indicates that kinetic folding intermediates are produced by the sequential (hierarchical[30]) folding of these units as in eq 5, where lower case

$$rygb(U) \rightarrow rygB \rightarrow ryGB \rightarrow rYGB \rightarrow RYGB(N)$$
$$\updownarrow \qquad\qquad \updownarrow \qquad\qquad\qquad (5)$$
$$ryg^\times b(U^\times) \rightarrow ryg^\times B \rightarrow \times$$

letters represent the initially unstructured segments and upper case letters the formed units. Equation 5 will produce fast two-state folding when g is freely available and slow three-state folding when the unfolded green unit is trapped ($g^\times$), as in the histidine misligation experiment. The $ryg^\times B$ intermediate, corresponding to the trapped N/C helix kinetic intermediate described by HX pulse labeling,[4] may be off-pathway in a strict sense (eq 5), but it reveals the structure of its on-pathway cognate, rygB.

Evidence that distinct nativelike intermediates exist comes most directly from HX and protein engineering experiments. HX pulse labeling and ϕ analysis experiments are performed in a kinetic mode and so will only detect intermediates that are blocked and populated, like $ryg^\times B$. The native state HX experiment makes it possible to detect and characterize intermediates as they cycle through the free energy region between the native and unfolded states under native conditions. The experiment has now found nativelike partially folded forms in cyt $c^{5,31}$ (Figure 1), RNase H,[32] a thermophilic rubredoxin,[33] and apocytochrome b_{562}.[34,35] Accumulating evidence suggests that these forms represent distinct kinetic intermediates as in eqs 4 and 5. In no case has experiment detected a continuum of partially unfolded forms as expected for some theoretical models.

Much recent work has stressed the indefinite nature of intermediates and transition states in coordinate space. Nevertheless, experimental work points to the reality of relatively distinct partially folded intermediates. These

forms are of course not as definitively structured as the unique native state, but it now appears that they can be usefully described in terms of distinct elements that form in a programmed manner as above. The H-bonded structure of intermediates may be defined by HX labeling experiments. Side chain interactions in intermediates and transition states may be defined by the ϕ analysis approach.[36] The stability and distinctness of these elements may be expressed in terms of ΔG from HX experiments and the magnitude of ϕ in mutational analyses. At this level of definition, folding intermediates can be usefully considered to be distinct as well as productive.

In summary, the folding pathway proposed here is simply a modified classical pathway with relatively defined and productive nativelike intermediates. The intrinsic rate-limiting process occurs at the initial step and has certain remarkable properties. A long range conformational search ultimately assembles a topologically nativelike constellation of interactions, nucleating a massively condensed transition state form that can support forward folding steps in an energetically downhill manner. Slow folding, heterogeneity, and intermediate accumulation are all due to optional misfolding errors within the otherwise productive intermediates.

References

(1) Sosnick, T. R.; Mayne, L.; Englander, S. W. Molecular collapse: The rate-limiting step in two-state cytochrome *c* folding. *Proteins* **1996**, *24*, 413−426.

(2) Sosnick, T. R.; Shtilerman, M. D.; Mayne, L.; Englander, S. W. Ultrafast signals in protein folding and the polypeptide contracted state. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 8545−8550.

(3) Qi, P. X.; Sosnick, T. R.; Englander, S. W. The burst phase in ribonuclease A folding: Solvent dependence of the unfolded state. *Nature Struct. Biol.* **1998**, *5*, 882−884.

(4) Roder, H.; Elöve, G. A.; Englander, S. W. Structural characterization of folding intermediates in cytochrome *c* by H-exchange labeling and proton NMR. *Nature* **1988**, *335*, 700−704.

(5) Bai, Y.; Sosnick, T. R.; Mayne, L.; Englander, S. W. Protein folding intermediates studied by native state hydrogen exchange. *Science* **1995**, *269*, 192−197.

(6) Sosnick, T. R.; Mayne, L.; Hiller, R.; Englander, S. W. The barriers in protein folding. *Nature Struct. Biol.* **1994**, *1*, 149−156.

(7) Brems, D. N.; Stellwagen, E. Manipulation of the observed kinetic phases in the refolding of denatured ferricytochromes c. *J. Biol. Chem.* **1983**, *258*, 3655−3660.

(8) Kuroda, Y. Residual helical structure in the C-terminal fragment of cytochrome *c*. *Biochemistry* **1993**, *32*, 1219−1224.

(9) Robertson, A. D.; Baldwin, R. L. Hydrogen exchange in thermally denatured ribonuclease A. *Biochemistry* **1991**, *30*, 9907−14.

(10) Ikai, A.; Fish, W. W.; Tanford, C. Kinetics of unfolding and refolding of proteins. II Results for cytochrome *c*. *J. Mol. Biol.* **1973**, *73*, 165−184.

(11) Jackson, S. E.; Fersht, A. R. Folding of chymotrypsin inhibitor 2. 1. Evidence for a two-state transition. *Biochemistry* **1991**, *30*, 10428−10435.

(12) Chen, X.; Matthews, C. R. Thermodynamic properties of the transition state for the rate-limiting step in the folding of the alpha subunit of tryptophan synthase. *Biochemistry* **1994**, *33*, 6356−62.

(13) Bryngelson, J. D.; Onuchic, J. N.; Socci, N. D.; Wolynes, P. G. Funnels, pathways, and the energy landscape of protein folding: a synthesis. *Proteins* **1995**, *21*, 167−95.

(14) Dill, K. A.; Chan, H. S. From Levinthal to pathways to funnels. *Nature Struct. Biol.* **1997**, *4*, 10−19.

(15) Abkevich, V. I.; Gutin, A. M.; Shakhnovich, E. I. Specific nucleus as the transition state for protein folding: evidence from the lattice model. *Biochemistry* **1994**, *33*, 10026−36.

(16) Thirumalai, D.; Guo, Z. Nucleation mechanism for protein folding and theoretical predictions for hydrogen-exchange experiments. *Biopolymers* **1995**, *35*, 137−140.

(17) Fersht, A. R. Optimization of rates of protein folding: the nucleation-condensation mechanism and its implications. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 10869−73.

(18) Fersht, A. R. Nucleation mechanisms in protein folding. *Curr. Opin. Struct. Biol.* **1997**, *7*, 3−9.

(19) Makhatadze, G. I.; Privalov, P. L. Protein interactions with urea and guanidinium chloride. A calorimetric study. *J. Mol. Biol.* **1992**, *226*, 491−505.

(20) Xie, D.; Fox, R.; Freire, E. Thermodynamic characterization of an equilibrium folding intermediate of staphylococcal nuclease. *Protein Sci.* **1994**, *3*, 2175−84.

(21) Ptitsyn, O. B. Kinetic and equilibrium intermediates in protein folding. *Protein Eng.* **1994**, *7*, 593−596.

(22) Creighton, T. E. The energetic ups and downs of protein folding. *Nature Struct. Biol.* **1994**, *1*, 135−138.

(23) Ptitsyn, O. B. Molten globule and protein folding. *Adv. Protein Chem.* **1995**, *47*, 83−229.

(24) Matthews, C. R. Pathways of protein folding. *Annu. Rev. Biochem.* **1993**, *62*, 653−83.

(25) Tsong, T. Y.; Baldwin, R. L.; McPhie, P. A sequential model of nucleation dependent protein folding. Kinetic studies of ribonuclease A. *J. Mol. Biol.* **1972**, *63*, 453−475.

(26) Wetlaufer, D. B. Nucleation, rapid folding, and globular intrachain regions in proteins. *Proc. Natl. Acad. Sci. U.S.A.* **1973**, *70*, 697−701.

(27) Sosnick, T. R.; Englander, S. W. What limits protein folding? In *Dynamics and the problem of recognition in biological macromolecules*; Jardetzky, O., Lefevre, J.-F., Eds.; Plenum Press: New York, 1996; Vol. 288; pp 65−71.

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

(29) Baldwin, R. L. The nature of protein folding pathways: The classical versus the new view. *J. Biomol. NMR* **1995**, *5*, 103−109.

(30) Rose, G. D. Hierarchic organization of domains in globular proteins. *J. Mol. Biol.* **1979**, *134*, 447−70.

(31) Bai, Y.; Englander, S. W. Future directions in folding: the multi-state nature of protein structure. *Proteins* **1996**, *24*, 145−51.

(32) Chamberlain, A. K.; Handel, T. M.; Marqusee, S. Detection of rare partially folded molecules in equilibrium with the native conformation of RNase H. *Nature Struct. Biol.* **1996**, *3*, 782−787.

(33) Hiller, R.; Zhou, Z. H.; Adams, M. W. W.; Englander, S. W. Stability and dynamics in a hyperthermophilic protein with melting temperature close to 200 °C. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 11329−11332.

(34) Fuentes, E. J.; Wand, A. J. Local dynamics and stability of apocytochrome b_{562} examined by hydrogen exchange. *Biochemistry* **1998**, *37*, 3687−98.

(35) Fuentes, E. J.; Wand, A. J. The local stability and dynamics of apocytochrome b_{562} examined by the dependence of hydrogen exchange on hydrostatic pressure. *Biochemistry* **1998**, *37*, 3687−3698.

(36) Fersht, A. R.; Matouschek, A.; Serrano, L. Folding of an enzyme I. Theory of protein engineering analysis of stability and pathway of protein folding. *J. Mol. Biol.* **1992**, *224*, 771−782.

(37) Kraulis, P. J. Molscript: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **1991**, *24*, 946−950.

(38) Englander, S. W.; Mayne, L. C.; Bai, Y.; Sosnick, T. R. Hydrogen exchange: the modern legacy of Linderstrom-Lang. *Protein Sci.* **1997**, *6*, 1101−1109.

(39) Houry, W. A.; Rothwarf, D. M.; Scheraga, H. A. Circular dichroism evidence for the presence of burst phase intermediates on the conformational folding pathway of ribonuclease A. *Biochemistry* **1996**, *35*, 10125−33.

AR970085H