PROTEINS: Structure, Function, and Genetics 24:413–426 (1996)

RESEARCH ARTICLES

Molecular Collapse: The Rate-Limiting Step in Two-State Cytochrome c Folding

Tobin R. Sosnick, Leland Mayne, and S. Walter Englander
The Johnson Research Foundation, Department of Biochemistry and Biophysics, University of Pennsylvania, Philadelphia, Pennsylvania 19104-6059

ABSTRACT **Experiments with cytochrome *c* (cyt c) show that an initial folding event, molecular collapse, is not an energetically downhill continuum as commonly presumed but represents a large-scale, time-consuming, cooperative barrier-crossing process. In the absence of later misfold-reorganization barriers, the early collapse barrier limits cyt c folding to a time scale of milliseconds. The collapse process itself appears to be limited by an uphill search for some coarsely determined transition state structure that can nucleate subsequent energetically downhill folding events. An earlier "burst phase" event at strongly native conditions appears to be a non-specific response of the unfolded chain to reduced denaturant concentration. The molecular collapse process may or may not require the co-formation of the amino- and carboxyl-terminal helices, which are present in an initial metastable intermediate directly following the rate-limiting collapse. After the collapse-nucleation event, folding can proceed rapidly in an apparent two-state manner, probably by way of a predetermined sequence of metastable intermediates that leads to the native protein structure (Bai et al., Science 269:192–197, 1995).** © 1996 Wiley-Liss, Inc.

Key words: protein folding, folding kinetics, folding barriers

INTRODUCTION

In earlier work we showed that equine cytochrome *c* (cyt c) can fold to the fully native state in a two-state manner with a 15 msec time constant (10°C, 0.7 M GdmCl)[1] (see also refs. 2,3). This established that the obligatory steps in folding—secondary and tertiary structure formation, side chain packing, water extrusion—are not intrinsically slow and cannot account for the barriers that limit folding to the usually observed time scale of seconds. The generality of this conclusion is supported by the similarly fast folding now seen in other proteins.[4–16] We also found that the slow step that limits cyt c folding under three-state folding conditions is the reorganization of misfolded structure trapped during the initial molecular collapse.[1] This too appears to occur widely and to limit the rate of multi-state folding in many proteins and also in some simplified model simulations of protein folding.[17–22] The present paper describes a search for the intrinsic step that limits the rate of cyt c folding when misfolding barriers are not encountered.

It is generally assumed that rate-limiting steps occur late in protein folding. This view has been promoted by several kinds of observations: 1) the widespread observation of metastable kinetic folding intermediates, which accumulate before a later rate-limiting step; 2) the demonstration of equilibrium molten globule forms, which have been thought to mimic kinetic folding intermediates that convert to the native state in some later rate-limiting step[23]; and 3) protein engineering results, which have been interpreted in terms of a late transition state.[24–26] The significance of these observations is open to question, especially in view of the prevalence of off-pathway misfold-reorganization barriers, for points one and two,[1] and some interpretational ambiguity for the third (see below).

Recent observations of fast, two-state folding in at least ten different proteins more directly challenge the view that the rate-limiting barrier is a late event. In two-state folding one cannot in general place the rate-limiting step or determine the order of events since all properties are acquired in a kinetically indistinguishable manner. Nevertheless, the observation of two-state kinetic folding does provide important information, namely, that *no intermedi-*

Received October 9, 1995; revision accepted December 30, 1995.

Address reprint requests to T.R. Sosnick, The Johnson Research Foundation, Department of Biochemistry and Biophysics, University of Pennsylvania, Philadelphia, PA 19104-6059.

This work was described at the Gordon Conference on Proteins in June, 1994 and at several international meetings since.

ate that is more stable than the unfolded state can occur before the rate-limiting barrier. Otherwise folding intermediates would accumulate on the unfolded side of the barrier and folding would not appear to be kinetically two state.

A conundrum then arises, which we will consider here for the specific case of cyt c. Although metastable intermediates of cyt c have been documented,[27,28] the present work demonstrates that cyt c can fold kinetically in a two-state manner.[1] The rate-limiting barrier for two-state folding cannot occur later than any of these intermediates. Experiments reported here resolve the conundrum by demonstrating that the rate-limiting step in cyt c folding is the initial chain collapse. Characterization of the collapse process shows it to be cooperative and much slower than expected for a diffusion-limited, energetically downhill process. These results have a number of implications for current views of protein folding.

MATERIALS AND METHODS

Stop-flow Spectroscopy

Starting from the fully unfolded state (3 M GdmCl at pH 2.4 or 4.2 M GdmCl at pH 6.2 in 20 mM sodium acetate), refolding was initiated by a dilution to native conditions (pH 4.9, 0.15–0.3 M sodium acetate, 10°C, variable GdmCl). For unfolding measurements, the starting native protein was in 1 M GdmCl, 0.2 M sodium acetate, pH 4.9. The intensity scale for each spectroscopic probe was normalized by measurements of the folded and the unfolded protein. Protein concentrations ranged from 20 μM, when folding was measured by fluorescence, to 0.3 mM when 695 nm absorbance was measured. The mixing experiments used a Biologic SFM-3 stop-flow apparatus equipped with a 200 W argon/mercury arc lamp and a sample cuvet of 10 mm optical pathlength and 1 mm width. Horse heart cyt c was type VI from Sigma Chemical Company.

HX Labeling With NMR Analysis

After 50 ms of folding, hydrogen exchange (HX) labeling was performed by a 1:1 dilution with 0.2 M glycine buffer at pH 9.6 for 40 ms, where the exchange halftime for an unstructured Ala–Ala peptide NH is 1 ms.[29] The labeling pulse was terminated by a 2:1 dilution into 0.3 M sodium citrate, 0.05 M sodium ascorbate to yield reduced cyt c at a final pH of 5.1–5.5 so that further H–D exchange is essentially halted. Samples were analyzed immediately by two-dimensional nuclear magnetic resonance (NMR) or temporarily stored at −85°C.

For NMR analysis, the protein samples were concentrated to ~4 mM (Amicon Centriprep-10 concentrators) and placed into 50 mM sodium acetate, 5 mM sodium ascorbate in D_2O buffer at pD_{read} 4.7 (Sephadex G25 spin columns). To determine ^1H-labeling, two-dimensional J-correlated spectra (COSY) were recorded in magnitude mode at 500 MHz on a Bruker AMX500 spectrometer at 20°C as described before.[1] Proton occupancies, using previously determined proton assignments,[30] were determined from NH-$C_\alpha H$ cross-peak volumes relative to a fully protonated sample adjusted for the 11% residual deuterium content in the labeling solution. The volume of a cross-peak for nonlabile hydrogens in the same spectral region (heme bridge-4) served as an internal intensity standard. Amide sites with proton labeling greater than 20% at pH 9.6 in the native protein were not included in the analysis. Occupancies at other sites were adjusted slightly for native state labeling.

Fragment Preparation

The cyt c heme-containing fragments 1-65 and 1-80 were prepared by adding 0.1 g of cyanogen bromide to 50 mg of cyt c in 5 ml at pH 1 and 10°C for 2 hr. Excess cyanogen bromide and other reaction products were removed by lyophilization and the peptides were purified by C18 reverse phase high-performance liquid chromatography. Purity was confirmed by mass spectrometry.

RESULTS

Early and Late Folding Probes

We want to learn about the process that ultimately limits folding when misfolding does not occur. The transition state that limits protein folding can be characterized by studying the effect of ambient conditions on folding and unfolding rates.

Cyt c provides a specific probe for molecular collapse, the fluorescence quenching of its single tryptophan (Trp59) due to resonance energy transfer as the tryptophan approaches the heme.[31] The sharp distance dependence[32] of this signal (r^6) and its zero baseline amplitude make it exceedingly sensitive to the Trp59 to heme separation and therefore to molecular condensation. A probe specific for final acquisition of the native state is provided by an absorbance band at 695 nm (A_{695}). This charge transfer signal depends on the weak ligation of Met80 to the heme iron, which occurs only in the fully native cyt c form.[33,34]

Figure 1 shows data for cyt c folding obtained by these early and late probes. When cyt c is initially fully unfolded at low pH (pH 2.4, 3 M GdmCl) and then diluted into the refolding condition used here (pH 4.9, 10°C), almost all of the molecules experience initial collapse (fluorescence) and 70% achieve final folding (A_{695}) at about the same fast rate (Fig. 1A–C). Thus, although intermediates play a role in cyt c folding,[27,28] large barriers are not encountered at these conditions, the intermediates do not become highly populated, and the kinetic folding of a major fraction of the cyt c population appears to be a two-state process.[1]

In the following, we consider the apparent two-state folding behavior exhibited by the major frac-

tion (~70%) of the protein population. Figure 1 shows that another fraction, ~30%, reaches the native state on a much slower time scale, reflecting molecules that encounter some non-obligatory barrier, perhaps a non-native histidine ligation or proline isomer.[35] We will ignore this fraction and focus attention on the faster folding molecules. As GdmCl in the refolding solution is reduced below 1 M concentration, an exceedingly fast decrease in fluorescence in the stopped-flow dead time becomes prominent. Experiments described below show that this so-called burst phase behavior represents a solvent-dependent contraction of the unfolded polypeptide chain that makes no positive contribution to the subsequent folding process.

Criteria for Two-State Folding

In addition to the quantitative agreement of different refolding probes, a second more rigorous test for two-state folding can be obtained by comparing equilibrium and kinetic parameters. The dependence of equilibrium unfolding on denaturant (den) is commonly described as in Eq. 1.

$$\Delta G^{\circ}(\text{den}) = -RT \ln K(\text{den}) = -RT \ln K(0) - m^{\circ}[\text{den}] \quad (1)$$

Analogous equations can be written for kinetic unfolding (u) and folding (f) reactions.

$$\Delta G^{\ddagger}_{u}(\text{den}) = -RT \ln k_u(0) - m^{\ddagger}_{u}[\text{den}] + \text{constant} \quad (2a)$$

$$\Delta G^{\ddagger}_{f}(\text{den}) = -RT \ln k_f(0) - m^{\ddagger}_{f}[\text{den}] + \text{constant} \quad (2b)$$

The m° value in Eq. 1 is proportional to the denaturant-sensitive surface newly exposed in the equilibrium unfolded state.[36,37] Similarly, $m^{\ddagger}_{u}$ and $-m^{\ddagger}_{f}$ in Eqs. 2a,b are proportional to the denaturant-sensitive surface exposed in moving from either initial state to the transition state.[38] When equilibrium and kinetic folding and unfolding are all adequately described as two-state reactions between the same native (N) and unfolded (U) states, limited by the same transition state barrier, then Eqs. 3a,b follow.

$$\Delta G^{\circ}(0) = \Delta G^{\ddagger}_{u}(0) - \Delta G^{\ddagger}_{f}(0) \quad (3a)$$

$$m^{\circ} = m^{\ddagger}_{u} - m^{\ddagger}_{f} \quad (3b)$$

Eqs. 3a,b show that when two-state conditions hold, ΔG° and m° can be obtained from kinetic experiments as well as from equilibrium unfolding data. Conversely, the demonstration of this equivalence provides a discriminating test for the validity of a two-state folding model.[5] The equivalence shows that the free energy and the surface exposure measured in the overall equilibrium unfolding transition is fully accounted for by the observed kinetic folding and unfolding reactions.

Demonstration of Two-State Folding

Figure 2A shows the effect of GdmCl concentration on cyt c folding under the same two-state conditions used before (Fig. 1A–C; initial unfolding at low pH). The results define a V-shaped (chevron[38]) curve. The relaxation rate observed in such kinetic experiments, plotted here as $RT \ln k_{obs}$, is the sum of folding and unfolding rates. Relaxation is dominated by the folding rate at low GdmCl (left limb of the chevron) and by the unfolding rate at high GdmCl (right limb). GdmCl increases the unfolding rate because the transition state exposes increased GdmCl binding surface relative to N. GdmCl slows the folding rate because the transition state exposes less surface than U.

Although the cyt c behavior in Figure 2A is not as simple or well behaved as for some other proteins, fairly accurate $m^{\ddagger}$ values can be obtained directly from the slopes of the V-shaped region of the kinetic experiments, and $\Delta G^{\ddagger}(0)$ values can be obtained by extrapolation to zero GdmCl. Eqs. 3a,b translate these kinetic determinations into equilibrium $\Delta G^{\circ}(0)$ and m° values. $\Delta G^{\circ}(0)$ and m° were also obtained from standard equilibrium melting experiments. The independent kinetic and equilibrium experiments yield equivalent free energy and surface burial parameters (Table I). This rules out the population of any folding intermediate that would account for significant surface burial or folding free energy in either folding or unfolding reactions. Thus kinetic folding and unfolding of cyt c through the chevron region act like two-state reactions rate-limited by a single, common barrier. The same conclusion comes from the agreement of the early fluorescence probe, the late A_{695} probe, and 34 H-bond formation probes.[1]

In the folding (left) limb of the chevron curve (Fig. 2A), the measured collapse rate (fluorescence) appears to be slightly faster than the rate for native state acquisition (A_{695}), suggesting that the fast folding seen here may not fully reach the two-state limit. The small rate discrepancy may reflect systematic error in the multiexponential analysis or may begin to resolve substeps in the folding pathway ($U \rightarrow I \rightarrow N$, with rate constants k_1 and k_2). In the latter case, the initial collapse-related step is still slower than the subsequent step (fitting gives $k_1/k_2 = 0.6$) so that the folding rate is largely limited by k_1, there is little accumulation of intermediates, Eqs. 3a,b are satisfied, and folding remains essentially two state. At GdmCl concentrations below about 1 M, the folding rate diverges from the chevron behavior. The agreement of early and late probes suggests that the two-state nature of folding is maintained, although it is clear that a different

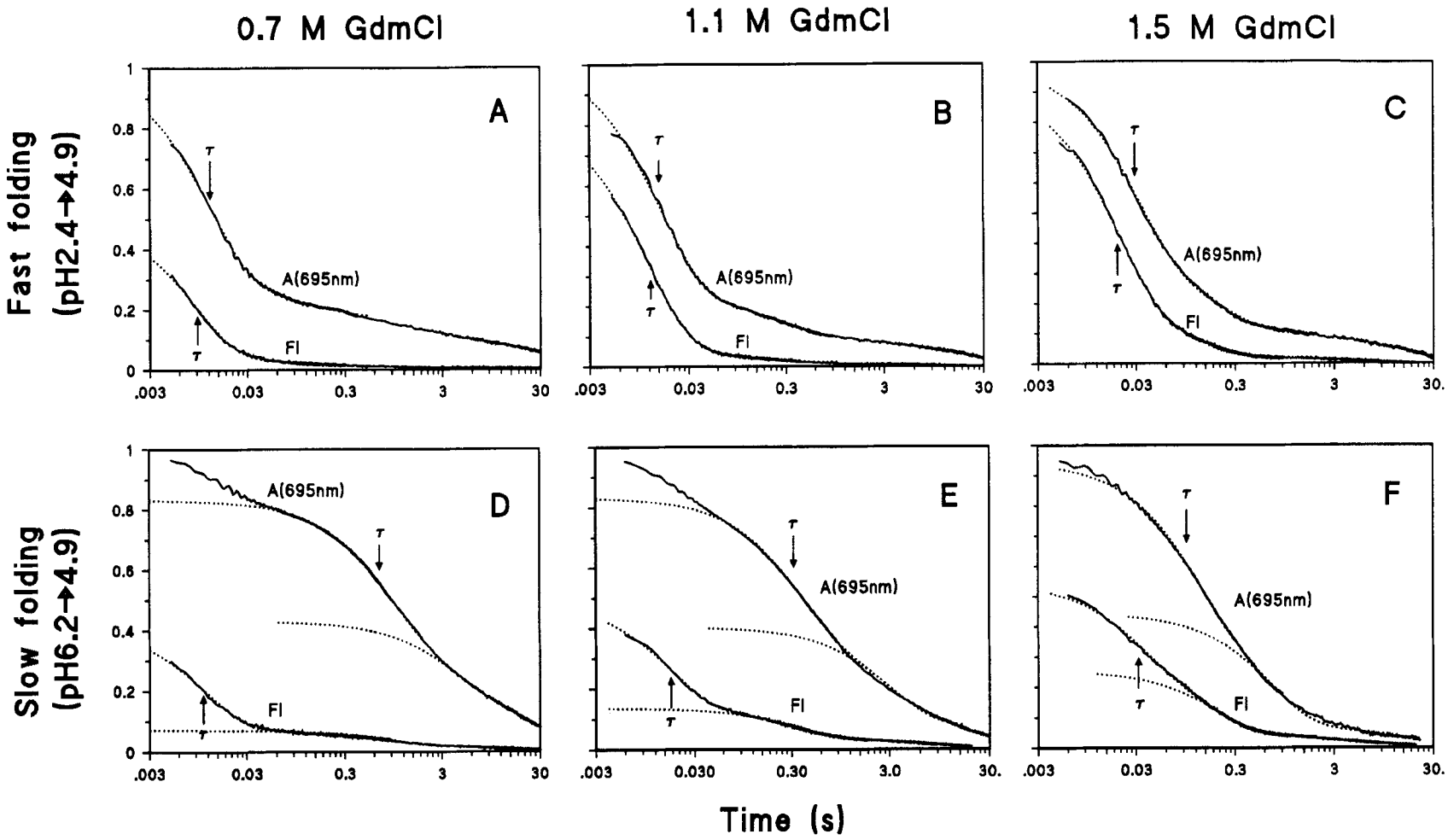

Fig. 1. Cyt c in fast folding (**A–C**) and slow folding (**D–F**) modes. Folding was followed in terms of molecular condensation (fluorescence quenching of Trp59, 20–40 μM protein) and final Met80 to heme ligation (absorbance at 695 nm, 200–300 μM protein). The ordinate gives the ratio of the time-dependent signal to the unfolded protein signal. Dotted lines indicate single exponential components, and arrows identify pertinent time constants. The folding reaction (at pH 4.9, 10°C in sodium acetate buffer and the GdmCl concentration shown) depends on the initial, unfolding condition, which was pH 2.4 (3 M GdmCl) for panels A–C (two-state behavior), and pH 6.2 (4.2 M GdmCl) for panels D–F (three-state behavior). In two-state folding the rate for collapse and final folding (A_{695} in A–C) is slowed by GdmCl, reflecting the large burial of GdmCl-binding surface in the rate-limiting transition state. In three-state folding the initial collapse rate mimics the two-state behavior. In contrast, the slower step in three-state folding (A_{695} in D–F) is accelerated by GdmCl because the rate-limiting reorganization barrier involves the unfolding of some structure.[1]

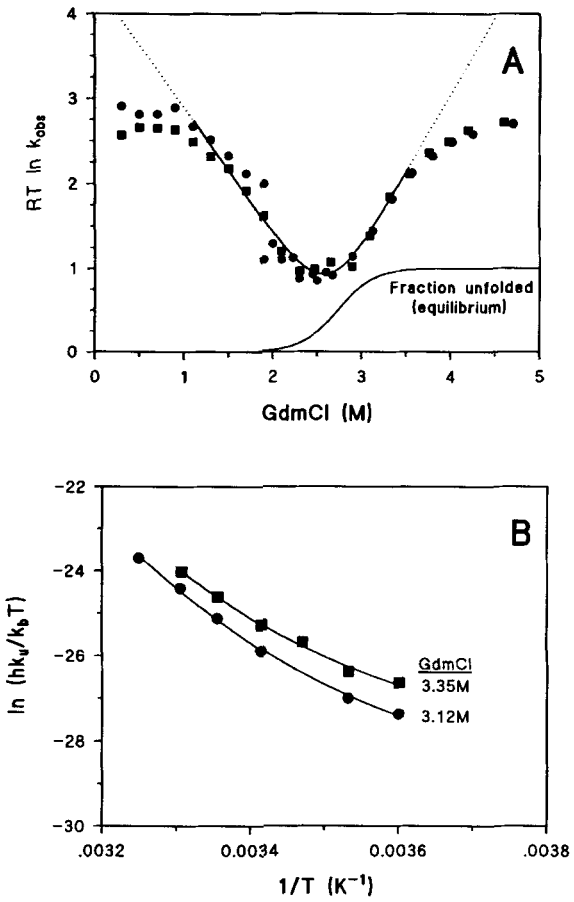

Fig. 2. Dependence of rates on denaturant and temperature. **A:** GdmCl dependence of folding and unfolding rates monitored by fluorescence quenching (circles) and A_{695} (squares) under two-state conditions. Initial conditions were as in Figure 1A for the folding experiments and were at 1 M GdmCl, pH 4.9 for the unfolding experiments. Kinetics were measured at the GdmCl concentration shown on the abscissa (pH4.9, 10°C). The parameter plotted on the ordinate, RT ln k_{obs}, puts the slope (m in Eq. 2) in units of kcal/mol/M. A linear fit to the data (Eqs. 1, 2) as shown yields the parameters in Table I. The fluorescence signal near 1.9 M GdmCl requires two exponentials for adequate fitting (both rates shown), perhaps due to the changing heme absorbance spectrum, which may be affected by local[62] as well as global behavior. **B:** Eyring plot for unfolding at 3.12 and 3.35 M GdmCl (pH 4.9). The lines are fits to the fluorescence data using classical transition state analysis.[75] Computed $\Delta C_p^{\ddagger}{}_u$ values are 0.67 ± 0.1 and 0.62 ± 0.2 kcal/mol K.

barrier or process becomes limiting (see section on the burst phase).

Parameters of the Two-State Barrier

The $m^{\ddagger}{}_f$ value obtained from the GdmCl dependence of folding rate (Fig. 2A) reflects primarily H-bonding polar groups[37] that interact favorably with the denaturant in U and are buried in the transition state. The ratio $m^{\ddagger}{}_f/m^o$ (Table I) shows that the transition state for folding buries ~45% of the GdmCl binding surface that is buried in the native state.

An independent measure, the difference in heat capacity between the unfolded state and the folding transition state ($\Delta C_p^{\ddagger}{}_f$), is largely sensitive to the burial of apolar surface.[39,40] This parameter can be obtained from the dependence of folding rate on temperature. In the present case folding becomes too fast to measure accurately over a sufficiently wide temperature range. $\Delta C_p^{\ddagger}{}_f$ was obtained indirectly from the measurement of unfolding rate as a function of temperature (Fig. 2B). The unfolding measurements yield a $\Delta C_p^{\ddagger}{}_u$ value of 0.65 kcal/mol K. The heat capacity increment for the folding transition can then be obtained from the relation $\Delta C_p^{\circ} = \Delta C_p^{\ddagger}{}_u - \Delta C_p^{\ddagger}{}_f$ = 1.6–1.7 kcal/mol K.[41,42] The value found for $\Delta C_p^{\ddagger}{}_f$, −1.0 kcal/mol K, is 60% of the equilibrium ΔC_p° value.

In summary, the dependence of folding rate on denaturant and temperature shows that the rate-limiting step in two-state folding includes a large-scale molecular condensation that buries about half of the molecular surface that is buried in native cyt c. The condensation is biased toward hydrophobic surface burial, but a great deal of polar surface is buried also. Clearly, a major collapse of the polypeptide chain occurs in the transition state or in the uphill process that leads to it.

The Burst Phase

The considerations just discussed refer to cyt c folding above 1 M GdmCl, where the chevron analysis and the correspondence of early and late probes document two-state folding and define some parameters of the rate-limiting process. When GdmCl in the refolding solution falls below 1 M, two new elements enter. A "burst phase" becomes prominent in which a fraction of the fluorescence signal is quenched within the stopped-flow dead time (3–5 msec) (Figs. 1, 2), and folding deviates from the chevron ("roll over" in Fig. 2A).

Since tryptophan fluorescence is independent of GdmCl,[31] the fluorescence burst phase seen in both two-state and three-state folding conditions (Fig. 1) must reflect some structural event faster than the early collapse-related step just discussed. Burst phase signals for other proteins have often been interpreted as indicating the fast formation of a folding intermediate. The increasing amplitude of the burst phase with GdmCl dilution can be translated into a melting curve for the putative intermediate. Figure 3 compares this with the melting curve for native cyt c.

A particularly revealing test for the structural significance of the cyt c burst phase is possible. An initial intermediate in cyt c folding is known to include the formation of the amino and carboxyl helices,[27] which dock against each other in the native protein. We prepared a fragment of cyt c extending from residue 1 to 80 that lacks the C-terminal helix region. This helix is the most stable region in native

TABLE I. Equilibrium and Transition State Parameters*

	m°	$-m^{\ddagger}_{f}$ (kcal/mol/M)	$m^{\ddagger}_{u}$	$\Delta G^{\circ}(0)$	$\Delta G^{\ddagger}_{u}(0) - \Delta G^{\ddagger}_{f}(0)$ (kcal/mol)
Cyt c	3.0 ± 0.2	1.4 ± 0.1	1.9 ± 0.2	8.2 ± 0.6	8.8 ± 0.4
Ubiquitin[6]	2.1	1.4	0.8	7.3	7.4
CI2[5]	1.8	1.1	0.7	7.0	7.2
AcoAbp[13]	2.7	2.1	1.4	6	8
GCN4-p1[64]	1.7	1.0	0.8	10.5	10.3
RNase A[4]	3	1.3			
RNase A[16]	3	1.4			

*Cyt c activation parameters obtained from kinetic experiments under two-state conditions are compared with equilibrium parameters from GdmCl melting experiments (pH 4.9, 10°C). Other proteins were studied between pH 5 and 6.3 and between 8° and 35°C. Parameters are defined in Eqs. 1 and 2. The success of Eqs. 3a,b documents two-state folding. The similarity in $m^{\ddagger}_{f}$ values for different proteins suggests that a minimum size is required to achieve a nucleus that can support polypeptide chain collapse.

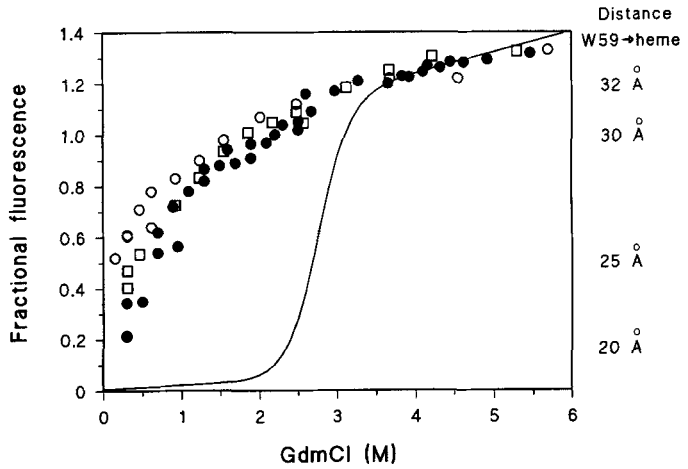

Fig. 3. The "burst phase" fluorescence signal for holo cyt c under fast folding conditions (•) and for cyanogen bromide fragments (□, 1–80; ○, 1–65) containing the heme and Trp59. The amplitude plotted is the ratio of fluorescence immediately after dilution into various GdmCl concentrations to fluorescence in the unfolded chain (3 M GdmCl, pH 2.4). The fluorescence signal for equilibrium melting of native cyt c (pH 4.9, 10°C) is normalized to the same reference conditions. The Förster-averaged Trp59 to heme distance[76] is shown (right ordinate).

cyt c[28] and exhibits some helical CD signal even when isolated from the protein.[43,44] Another experimental fragment, 1-65, lacks regions that compose the C-terminal helix, the short 70s helix, and almost all of the major 60s helix. Both fragments retain the fluorescence quenching probe provided by Trp59 and the covalently bound heme (at Cys14 and Cys17). When these fragments are diluted from denaturing GdmCl, as in the cyt c folding experiments, they exhibit an instantaneous (dead time) increase in fluorescence quenching that is remarkably similar to the holoprotein burst phase (Fig. 3). This result is against the development of an authentic folding intermediate in the burst phase.

Additional observations support this conclusion. The hypothetical intermediate formed in the burst phase produces no significant hydrogen exchange protection, indicating the absence of stable H-bonded structure.[1,27] The fragments and the holoprotein duplicate the upper fluorescence baseline for the equilibrium denaturation curve of cyt c (Fig. 3); the burst behavior appears simply to extend the unfolded baseline to native conditions, as has been clearly shown for CspB, which folds in a rigorously two-state manner.[14] It has long been known that denatured proteins reach expected random coil dimensions only in concentrated GdmCl and tend to adopt what Tanford[45,46] referred to as an "incompletely disordered state" in low denaturant. Such a contraction in diluted GdmCl can directly explain the burst phase fluorescence energy transfer results for cyt c and similar results for barstar.[47] The degree of molecular contraction in the cyt c burst phase, quantified by the Förster relationship,[32] is small; a fluorescence quenching of 50% represents a decrease in the Trp59 to heme distance (6th power av-

erage) from 30 to 25 Å (Fig. 3, right axis; native distance is 5–8 Å). Finally, cyt c develops considerable far ultraviolet circular dichroism (CD) in the burst phase upon GdmCl dilution.[48] The 1-80 fragment develops much less (mean residue ellipticity), and the 1-65 fragment even less, even though their fluorescence burst phase signals are nearly identical (manuscript in preparation). It is relevant that temperature-unfolded cyt c retains a sizeable CD signature that can be removed by increasing GdmCl and that other workers have found clear indications of excess, non-native CD in other refolding systems.[49–51] All these results indicate that the burst phase signals result from a trivial readjustment of the unfolded polymer chain to the new poorer solvent (low GdmCl; see also Kuszewski et al.[8]).

This view also provides an interpretation of the roll-over behavior that follows the burst phase; the spontaneous (downhill) chain contraction increases the activation free energy gap (trough to peak) for the rate-limiting collapse step. The degree of roll-over observed requires a $\Delta\Delta G^{\ddagger}$ (below the chevron) of only 1.5 kcal/mol or less. In the roll-over region cyt c folding continues to occur on a millisecond time scale and to appear essentially two state, as indicated by the nearly identical kinetics for early and late spectral signals (Fig. 2A) and hydrogen bond formation.[1]

A roll-over in the chevron at low denaturant has been seen for other proteins and, like the burst phase, has often been interpreted as signifying the accumulation of a new folding intermediate, either an off-pathway intermediate in the sense $I \leftrightarrow U \leftrightarrow N$ or an on-pathway intermediate, as in $U \leftrightarrow I \leftrightarrow N$. The present results indicate that the cyt c burst phase represents a solvent-induced contraction of the polymer chain rather than the formation of a productive intermediate.

Induction of Three-State Folding by Barrier Insertion

We want to investigate further the large-scale, collapse-related process that accounts for the rate-limiting step in the fast, two-state folding of cyt c. Is the barrier in two-state folding late, as has been generally conceived, so that most folding events occur on the upslope of the barrier before the transition state is reached, or is the barrier early so that structural elements form preponderantly afterwards, on the downslope?

These questions are difficult to deal with under two-state conditions where all properties appear to be acquired in a single step. Nevertheless the folding process in any single molecule may well proceed through a relatively defined sequence of events, such as might be investigated by computer simulations. Alternatively, the ability to manipulate cyt c folding behavior provides an experimental approach. It is possible to distinguish early and late folding events in cyt c by deliberately modifying the initial unfolding condition while the folding condition itself is kept constant.

When cyt c is initially unfolded at a pH somewhat higher than before (pH 6.2, 4.2 M GdmCl), a non-native histidine to heme-iron ligation forms.[1–3] Upon dilution into the standard refolding condition (pH 4.9), the fluorescence-detected molecular collapse occurs as before, but subsequent acquisition of the native state is greatly slowed because the molecular collapse traps the false heme ligation together with the accompanying misfolded chain. This introduces a sizeable reorganization barrier,[1] causes a collapsed intermediate to be transiently populated and produces three-state folding (Fig. 1D–F). Folding events that occur earlier than the populated intermediate can then be clearly distinguished from events that occur later.

Manipulation of the cyt c unfolding condition to insert a new reorganization barrier in this way can be especially revealing because the new barrier only becomes effective at a somewhat advanced stage in folding and appears not to perturb the earlier events. This is suggested by the results in Figure 1. The arrows in Figure 1 show that the collapse process measured specifically by fluorescence quenching exhibits the same rate in two-state and three-state folding. This equivalence suggests that the same barrier crossing process that limits two-state folding may be isolated within the early phase of three-state folding where it serves to limit formation of the early collapsed intermediate. Such a conclusion would be consistent with the principle that the rate-limiting barrier in two-state folding must precede all stable intermediates, and further that the ultimate rate limitation is connected with the collapse process itself.

Comparison of the Two-State and Three-State Collapse Barriers

If the early events are in fact preserved in the three-state case, then the barrier insertion experiment provides a means for overcoming the indistinguishability of events in two-state folding. The ultimate rate-limiting step could then be more cleanly studied under three-state conditions. In the barrier insertion experiment that produces three-state folding, the initial misligation of a histidine (His26 or His33) to the heme in the unfolded state imposes an incipient constraint that is expressed later in refolding as a misfold-reorganization barrier. One wants to know whether the presence of the non-native heme ligation significantly alters the early collapse-related barrier. To compare the two-state barrier with the early three-state barrier more critically, experiments were done to measure their thermodynamic and structural parameters.

Figure 4 compares folding rates due to the kinetic barrier that limits native state formation in two-

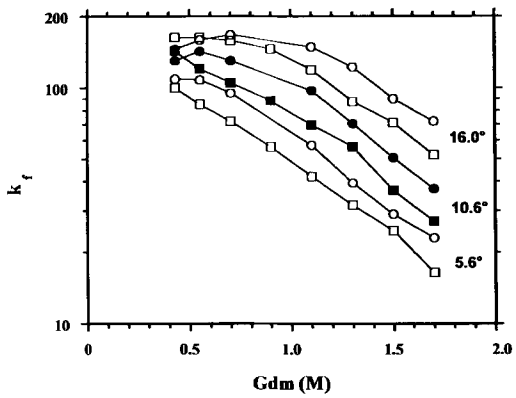

Fig. 4 The collapse rate measured by fluorescence quenching under two-state (circles) and three-state (squares) folding conditions as a function of temperature and GdmCl concentration.

state folding and the barrier that limits the collapse step in three-state folding. The dependence of these rates on temperature and GdmCl concentration was measured by fluorescence quenching, in all cases at the same final condition (pH 4.9). We find essentially the same activation free energy (within 0.2 kcal/mol), the same activation enthalpy (16.6 ± 0.4 and 16.0 ± 0.9 kcal/mol, above 1 M GdmCl), and therefore also the same activation entropy. ($\Delta G^{\ddagger}$ is roughly 10 kcal/mol, depending on the prefactor assumed.) The complex dependence of rates on GdmCl concentration is also very similar in the two cases. At higher GdmCl concentrations, the rate vs. GdmCl dependence is steep (see also Fig. 2A), indicating that the transition states in the two cases exhibit similarly large surface burial, equal to about half the surface that is buried in the native state. [In contrast, the later misfold-reorganization barrier in three-state folding is considerably larger and has the opposite dependence on GdmCl (Fig. 1).] At low GdmCl, measured rates "roll over" and fall below the extrapolated values, limiting at the same rate in both two- and early three-state folding at each temperature tested. All these identities, in both thermodynamic and structural parameters, indicate that the same process that limits the early molecular collapse step in three-state folding also acts to limit the entire folding process when the subsequent misfold-reorganization barrier is absent and folding is two state.

This conclusion requires that a significant difference in the unfolded state, the histidine misligation, has no significant effect on early folding events. In the light of present knowledge, this conclusion is not surprising. Most generally, one can note that the unfolded state of any protein includes an exceedingly broad range of conformational forms, yet proteins often exhibit apparent two-state, mono-exponential folding behavior.[5–9,13–15] Thus major differences among unfolded forms need not affect the barrier that separates the unfolded form from subsequent stages in folding.[25] Information on the structure and folding pathway of cyt c allows a more specific statement. It is obvious that the cross-heme misligation of His26 or His33 should block folding when the histidine-containing segment attempts to organize, but the histidine-containing segment and the heme iron site are spatially separated from the amino and carboxyl helices in the native protein. There is no evident structural reason why the His-heme interaction should interfere with formation of the amino and carboxyl-terminal helices (see below and Roder et al.[27] and Bai et al.[28]). Finally, independent information depicts a folding pathway in which the amino and carboxyl helices form first, followed by a step that structures the histidine-containing segment.[28] It seems unlikely that early events would be affected by chain regions that are still highly flexible as well as structurally remote. These considerations point to an independence of the early three-state barrier from the His-heme ligation that produces the later misfold barrier and thus support the experimental demonstration that the limiting two-state and early three-state barriers are essentially identical.

We conclude that the rate-limiting barrier in two-state folding is crossed relatively early in the sequence of folding events and that it, like the early three-state barrier, includes a massive, energetically uphill molecular collapse.

Structure in the Early Transition State

We want to learn what structure in addition to the condensed polypeptide chain might contribute to the transition state, and in this sense just how early the rate-limiting step is. To gain more detailed information on hydrogen-bonded structure in the transition state, we studied the intermediate that is populated directly following it in the three-state folding sequence.

Folding was allowed to proceed for 50 ms under the three-state conditions used in Figure 1D. At the 50 ms time point, the collapse phase has ended and an intermediate is maximally populated. Figure 1D shows that about 20% of the protein population has already reached the native state (A_{695}), while essentially all the remainder occupies some collapsed intermediate form. An H-D exchange labeling pulse (40 ms, pH 9.6, 10°C) was then imposed to assess the presence of stably H-bonded structure.[52] The results (Fig. 5) show the degree of H-bonding present at each measurable peptide group in the protein. The H-bonding signal expected for the 20% native population is indicated by the hatched region in Figure 5. Almost all the molecules have formed structure in their two terminal helices (see also Roder et al.[27]). About 15% of the molecules may have formed additional structure throughout, perhaps reflecting a subfraction that has progressed to a later, near-native state but still lacks the Met80 to heme ligation (A_{695} in Fig. 1). These results essentially repeat pre-

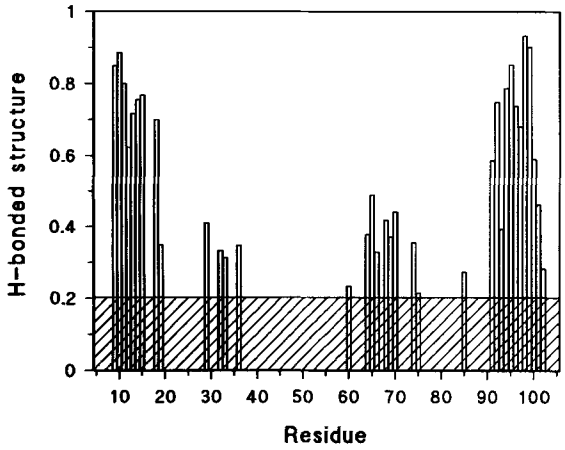

Fig. 5. Pulsed HX labeling results for three-state folding showing structure formation at 31 local and 4 non-local H-bonding sites. Conditions were as in Figure 1D, with the H-D exchange labeling pulse applied after 50 msec of folding where 20% of the molecules are fully native (A_{695}). The expected labeling for this fully native subpopulation is indicated by the hatched region. This pattern of HX protection develops on the same time scale as the collapsed intermediate (fluorescence, Fig. 1D). In the representation shown here, a fully native population would show a uniform value of 1.0; a fully unfolded population would show 0.0; a homogeneous partially folded species would show 1.0 in structured regions and 0.0 elsewhere. Heterogeneous mixtures exhibit linear combinations of these alternatives.

vious observations[27] with minor quantitative differences, which may be due to the difference in the final folding conditions (pH 4.9 compared with pH 6.3) and other experimental uncertainties.

Most interesting for present considerations is the large amount of still unstructured chain (white space in Fig. 5). The transition state that precedes the collapsed intermediate in three-state folding can contain no more structure than the intermediate itself (the two terminal helices), and may contain less. The helices found to be present in the intermediate are unlikely to form before the transition state is reached because the bihelical complex itself is known to have some independent stability.[43,44] The helices may form in the transition state itself or may form afterwards, on the downhill traverse to the stable intermediate. Therefore the collapse transition state must represent a very early step in the folding process.

These results also reflect on the large surface burial that occurs in the transition state, involving 50% or more of the surface that is buried in final folding. This surface cannot be accounted for by formation of the two terminal helices alone but must include the burial of many additional polar and apolar groups.

All these considerations portray the transition state as an initial, extensively condensed form that contains no major quantity of defined secondary structure but does include many coarsely defined interactions brought together by some rare although not unique conformations. Thus, the earliest significant folding event in cyt c is a massive uphill time-consuming (ms) chain collapse.

DISCUSSION

The results obtained here can be summarized as follows. When misfolding is avoided the large fraction of cyt c molecules fold rapidly in a two-state manner. The two-state character is rigorously demonstrable in GdmCl concentrations above about 1 M (same rate for early and late probes; ΔG and m values match the two-state chevron criteria). The process that limits two-state folding occurs within the early collapse step in three-state folding (same rate, temperature dependence, and denaturant dependence) where it can be more directly studied. The limiting process is a very early step in the folding sequence (no stable intermediate occurs earlier; much H-bonded structure forms later). The even earlier burst phase and the chevron roll-over at low GdmCl are due to a shrinkage of the chain dimension in the poorer solvent (fragment experiments and additional observations). The rate-limiting process is a large-scale molecular collapse (buries about half of the polar and apolar surface that is buried in the native state) and perhaps some secondary structure formation (HX pulse labeling).

Molecular Collapse as a Cooperative, Search-Nucleation Process

It has been widely assumed that molecular collapse is a barrier-free, energetically downhill process that occurs very early in folding,[53] before other structure-forming events, by a random, diffusion-limited association of hydrophobic side chains. This view is inconsistent with the present results and more generally with the fact that folding often appears to be a two-state reaction, since no downhill process can precede an authentic two-state barrier. Furthermore, we find that molecular collapse is slower by orders of magnitude than one expects for a random association process, since the effective interaction concentration between groups in an unfolded polypeptide, in the 10 mM range,[54,55] would produce random intra-residue association on the submicrosecond time scale. Thus the great majority of intrachain encounters must fail to achieve productive association and tend to separate. Factors that militate against random pairwise association and presumably account for the energetically uphill nature of the collapse process include unfavorable loop[56,57] and side chain[58] entropy and the burial of unsatisfied H-bonding groups.[59]

We conclude that the limiting on-pathway barrier in cyt c folding is an initial, large-scale, energetically uphill, molecular collapse, perhaps together with some helix formation. The molecular collapse process involves many side chains in a search for some relatively rare, large-scale transition state

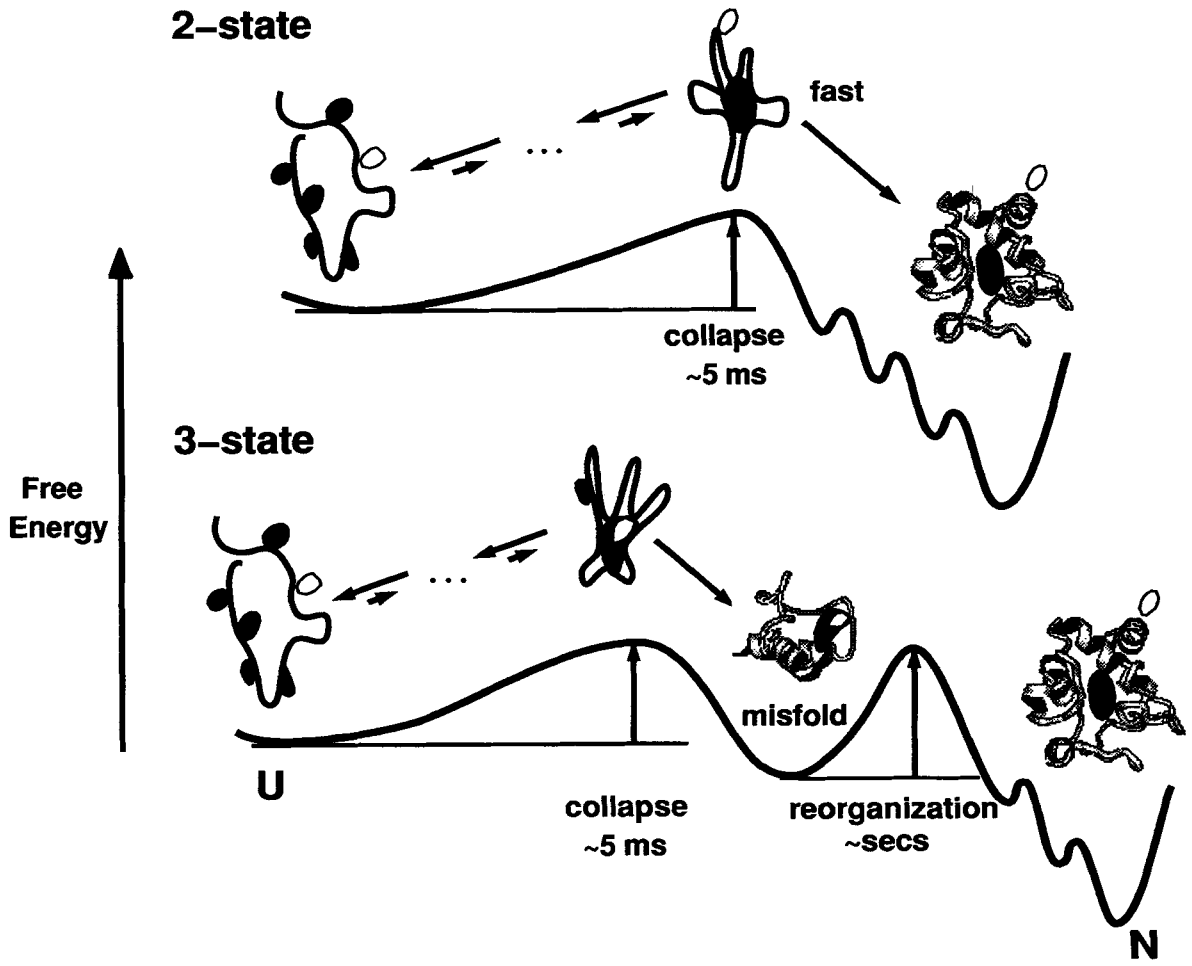

Fig. 6. Search-nucleation model for collapse and folding under two-state (above) and three-state (below) conditions. In the collapse process, an energetically uphill search consisting of many unfavorable association reactions ultimately finds a combination of interacting groups that can nucleate chain collapse and subsequent energetically downhill growth reactions. In the collapse transition state, preferentially hydrophobic groups (black in the diagram) form a large collapsed core while some main chain loops remain exposed to solvent. Under two-state conditions, formation of the collapsed core is the slow step (ms), followed by the rapid energetically favorable condensation of the exposed loops onto the core to form a series of intermediates leading to the native state.[28] Under three-state folding conditions some incipient "error" in either the core (e.g., the white residue) or an external loop (e.g., a non-native proline isomer) leads, during or after collapse, to a misfolding that must be corrected before folding can continue. The reorganization barrier slows folding and causes a metastable intermediate to accumulate. For example, the folding sequence inferred in Bai et al.[28] predicts that the adventitious His-heme barrier exploited in the present experiments permits formation of the early bihelical intermediate but blocks the formation of the subsequent intermediate, which involves the histidine-containing segment.

configuration, which can then support subsequent favorable interactions to carry the protein downhill toward the native state. This kind of uphill multistep core formation is by definition a nucleation process. Figure 6 illustrates the process and provides some further detail.

The Issue of Generality

The behavior found here for cyt c appears to be characteristic of a number of proteins known to exhibit fast, two-state folding. For many proteins two-state folding occurs on a similar time scale, in the millisecond range[5–9,13–15] or slightly faster at higher temperature.[12] Also similar values for $m^{\ddagger}_{f}$ are found, almost all in the range between 1 and 1.4 kcal/mol/M (Table I). An extensive mutational Φ-analysis by the Fersht laboratory characterized the transition state in the two-state folding in chymotrypsin inhibitor 2. The hydrophobic residues located in the protein core are loosely structured, achieving at most 55% of their native-like association energies in the folding transition state, while most core-related residues interact much less, and little if any secondary structure is formed.[24,60] Similarly, the substitution of many sites by alanine (one at a time) in the P22 Arc repressor caused little change in folding rates, indicating that side-chain

interaction free energies in the transition state are less than 30% of those in the native state.[26]

Although these results have been interpreted in terms of a late rate-limiting barrier, all these results for other proteins are consistent with the formation of a large weakly interacting core of significant size in the early rate-determining collapse transition state, as found here for cyt c. Insofar as subsequent barriers represent off-pathway misfold-reorganization steps,[1] the early collapse-nucleation barrier will account for the intrinsic rate-limiting step in protein folding.

Implications

It is widely assumed that the intrinsic rate-limiting step in protein folding occurs late, after collapse and extensive secondary structure formation. This conclusion rests mainly on the observation that intermediates often accumulate in folding experiments,[61] pointing to some later barrier. While distinct intermediates do evidently participate in cyt c folding,[27,28] the late barrier that causes kinetic intermediates to accumulate is known to represent an adventitious misfold-reorganization step.[1] In the absence of misfolding, we now see that the intrinsic rate-limiting barrier is the early molecular collapse step. Whether the nature of cooperative polypeptide collapse is imposed by inherent physical constraints or has been selected for in biological evolution remains to be seen.

An early folding/late unfolding barrier might be thought to be incompatible with protein rigidity and stability. Rather, rigidity is determined by the shape of the native well and not its absolute depth, and an early folding barrier ensures that most stabilizing interactions contribute to the height of the unfolding barrier. Furthermore, the kinetic folding process may well benefit by placing the major conformational search process early, before collapse and stable structure formation, since the large-scale search problem is then faced in the mobile unfolded state and not in the condensed state where the reorganization of misfolded structure within a collapsed state would be slow, as we have found.[1]

The collapse-nucleation barrier studied here limits cyt c folding to a time scale of about 5 msec (measured at 16.5°C in zero GdmCl). The literature considers earlier folding events. Photoinitiated folding experiments with cyt c[62] have detected events that occur on a submillisecond time scale. These observations were made by Soret absorbance changes, which reflect heme ligation reactions and not necessarily structure formation. Furthermore, this work was done in high GdmCl where folding appears to be two state. Cyt c and nine other proteins[5,6,9,12–15,63,64] have now been shown to satisfy the strict two-state criteria in Equation 3, with folding on a millisecond time scale. This rigorously rules out the faster formation of stable structure that would account for any significant fraction of a protein's equilibrium folding free energy or surface burial.

The energetically uphill, multi-step search that is required to find a satisfactory collapse transition state (Fig. 6) is by definition a nucleation process. Computer simulations by Shakhnovich and co-workers[20] and by Guo and Thirumalai[65] point to the efficacy of nucleation by a subset of native-like interactions. Nucleation models for protein folding have been considered before[66,67] but were discarded when evidence was found for well-populated intermediates,[61] pointing to later barriers. We have found that, although folding intermediates do occur, their accumulation signals off-pathway misfold-reorganization barriers[1] that are not inherent to the folding process and so do not refute the reality of an early nucleation mechanism.

Available data suggest that the transition state for successful collapse involves a collapsed core of significant size ($m^{\ddagger}_f$ in Table I). This requirement will make the transition state for the two-state folding of small proteins appear to be late, close to the native state, and early for larger proteins. That is, if a sizeable $m^{\ddagger}_f$ must be maintained to support successful collapse ($\geq$1 kcal/mol/M; Table I), then small proteins with small equilibrium m^o values must preferentially decrease their $m^{\ddagger}_u$ values (Eq. 3b).

Although the requirements for correct collapse appear to be substantial, the formation of an incorrect core or the presence of an incipient misfold (nonnative proline isomer, heme ligation, etc.) can occur, and this can give rise to a misfold-reorganization barrier as folding progresses[1] (Fig. 6). Several consequences follow. Small proteins with few candidate core residues are unlikely to err and should generally fold in a fast, two-state manner. For larger proteins, the probability of misfolding is high. Therefore different fractions of a folding population are likely to encounter different barriers and to exhibit a heterogeneity of faster and slower folding fractions.[68,69] This kind of heterogeneous folding behavior has been generally described in terms of parallel pathways. More precisely, this behavior probably reflects a heterogeneity of off-pathway reorganizational barriers and not necessarily alternative sets of on-pathway intermediates. That is, misfolding barriers are ubiquitous; alternative sets of stable intermediates are probably not.

The view that folding is ultimately limited by an initial conformational search makes folding analogous to metabolic pathways that are limited by an initial enzymatic step. Although intermediates do not accumulate in such pathways, one does not doubt that intermediates exist; this distinction has often not been clearly appreciated in the case of two-state protein folding. For protein folding as for enzyme pathways, biological evolution is evidently able to ensure that the later on-pathway steps do not

become limiting. The common scale of two-state folding rates seen for many proteins (~1–10 ms) suggests that selective evolution has little ability, or little need, to further enhance the initial search.

CONCLUSIONS

Results now available for cyt c point to some of the principles that govern its folding behavior and perhaps the folding of other proteins.

1. The unfolded polypeptide engages in dynamic, generally unfavorable, internal association-dissociation reactions that ultimately find a rare although not unique combination (nucleus) of cooperatively stabilizing interactions that can together support a large-scale chain condensation. Successful conformations are likely to resemble the gross native topology, since the collapse transition state is defined by its ability to nucleate subsequent energetically downhill steps (see also refs. 20,65). The early association-dissociation behavior can be viewed as an energetically uphill conformational search for the transition state.
2. Main chain loops that remain unstructured in the collapse transition state can then condense onto the stabilizing core nucleus to form elements of secondary structure in locally cooperative reactions that depend on small-scale searches[28] Here local interactions determine the detailed structure, as in the LINUS algorithm.[70] These second-level reactions appear to step cyt c through a sequence of three stable intermediates before reaching the native state.[28]
3. Incorrect collapse or incipient defects within correctly condensed molecules (non-native isomers; improper ligations) can produce misfold-reorganization barriers that greatly slow folding and populated intermediates.[1] The destabilization of a populated intermediate can make folding faster,[71] but in principle only when further folding is limited by a reorganization barrier that requires the loss of the interactions (native or non-native) that are experimentally destabilized in the intermediate. Proteins that facilitate folding, including chaperonins, disulfide isomerases, and prolyl isomerases, may similarly be considered to function by promoting the loss of misfolding interactions.[72,73]

These conclusions suggest that proteins solve the conformational search problem[74] in two stages, involving an initial coarse sorting of structural segments followed by the more local development of secondary folding and tertiary packing. The initial search for a collapse-competent conformation need only find a coarse approximation to the native fold and so can be faster by far than would be necessary to find the unique native state.[74] The large-scale search for the coarsely organized nucleus represents the inherent slow step in folding and brings about a major reduction in the dimensionality of the search problem. The fine scale structure can then be achieved in smaller scale local searches, which appear to carry cyt c through several intermediates to the native state.[28] The fact that such intermediates are not well populated in two-state folding has often been interpreted to mean that they do not exist or are not important. We suggest that it is precisely the existence of these intermediates and the pathway they define that makes fast folding possible.

ACKNOWLEDGMENTS

We thank J. J. Englander, Y. Bai, J. Lear, K. A. Sharp, N. R. Kallenbach, R. L. Baldwin, G. P. Connelly, K. Dill, H.S. Chan, E. Shakhnovich, D. Thirumalai, and G. D. Rose for useful discussions and comments on the manuscript, and E. Margoliash for supplying cyt c mutants. We thank mark Shtilerman for help in preparation and characterization of the cyt c peptide fragments. This work was supported by NIH grant GM31847. Mass spectrometry was provided by Dr. William T. Moore of the Protein Chemistry Laboratory, Medical School, University of Pennsylvania and supported by core grants of the Diabetes and Cancer Centers (NIH DK-19525 and CA-16520 Dr. J. D. Lambris).

NOTE ADDED IN PROOF

A. Fersht[77] has reinterpreted protein engineering results for C12 in terms of a "nucleation-condensation" mechanism, similar to the early step in the search-nucleation model described here. Also, computer simulations[78] show that strongly stabilizing conditions tend to promote the rapid formation of a "structureless globule" which must then reorganize to reach the nucleation transition state[20,65] like the cyt c burst phase behavior seen at low GdmCl, while nucleation can proceed without a burst phase under less stable conditions, as for cyt c at higher GdmCl.

REFERENCES

1. Sosnick, T.R., Mayne, L., Hiller, R., Englander, S.W. The barriers in protein folding. Nature Struct. Biol. 1:149–156, 1994.
2. Brems, D.N., Stellwagen, E. Manipulation of the observed kinetic phases in the refolding of denatured ferricytochromes c. J. Biol. Chem. 258:3655–3660, 1983.
3. Elöve, G.A., Bhuyan, A.K., Roder, H. Kinetic mechanism of cytochrome c folding: Involvement of the heme and its ligands. Biochemistry 33:6925–6935, 1994.
4. Garel, J.R., Nall, B.T., Baldwin, R.L. Guanidine-unfolded state of ribonuclease A contains both fast- and slow-refolding species. Proc. Natl. Acad. Sci. USA 93:1853–1857, 1976.
5. Jackson, S.E., Fersht, A.R. Folding of chymotrypsin inhibitor 2. 1. Evidence for a two-state transition. Biochemistry 30:10428–10435, 1991.
6. Khorasanizadeh, S., Peters, I.D., Butt, T.R., Roder, H. Folding and stability of a tryptophan-containing mutant of ubiquitin. Biochemistry 32:7054–7063, 1993.
7. Alexander, P., Orban, J., Bryan, P. Kinetic analysis of folding and unfolding of the 56 amino acid IgG-binding domain of streptococcal protein G. Biochemistry 31:7243–7248, 1992.
8. Kuszewski, J., Clore, G.M., Gronenborn, A.M. Fast folding

of a prototypic polypeptide: The immunoglobulin binding domain of streptococcal protein G. Protein Sci. 3:1945–1952, 1994.
9. Viguera, A.R., Martinez, J.C., Filimonov, V.V., Mateo, P.L., Serrano, L. Thermodynamic and kinetic analysis of the SH3 domain of spectrin shows a two-state folding transition. Biochemistry 33:2142–2150, 1994.
10. Houry, W.A., Rothwarf, D.M., Scheraga, H.A. A very fast phase in the refolding of disulfide-intact ribonuclease A: Implications for the refolding and unfolding pathways. Biochemistry 33:2516–2530, 1994.
11. Huang, G.S., Oas, T.G. Structure and stability of monomeric lambda repressor: NMR evidence for two-state folding. Biochemistry 34:3884–3892, 1995.
12. Huang, G.S., Oas, T.G. Submillisecond folding of monomeric lambda repressor. Proc. Natl. Acad. Sci. USA 92: 6878–6882, 1995.
13. Kragelund, B.B., Robinson, C.V., Knudsen, J., Dobson, C.M., Poulsen, F.M. Folding of a four-helix bundle: Studies of acyl-coenzyme A binding protein. Biochemistry 34: 7217–7224, 1995.
14. Schindler, T., Herrler, M., Marahiel, M.A., Schmid, F.X. Extremely rapid protein folding in the absence of intermediates. Nature Struct. Biol. 2:663–673, 1995.
15. Villegas, V., Azuaga, A., Catasus, L., Reverter, D., Mateo, P.L., Aviles, F.X., Serrano, L. Evidence for a two-state transition in the folding process of the activation domain of human procarboxypeptidase A2. Biochemistry 34:15105–15110, 1995.
16. Houry, W.A., Rothwarf, D.M., Scheraga, H.A. The nature of the initial step in the conformational folding of disulfide-intact ribonuclease A. Nature Struct. Biol. 2:495–503, 1995.
17. Dill, K.A., Bromberg, S., Yue, K., Fiebig, K.M., Yee, D.P., Thomas, P.D., Chan, H.S. Principles of protein folding—a perspective from simple exact models. Protein Sci. 4:561–602, 1995.
18. Camacho, C.J., Thirumalai, D. Kinetics and thermodynamics of folding in model proteins. Proc. Natl. Acad. Sci. USA 90:6369–6372, 1993.
19. Sali, A., Shakhnovich, E., Karplus, M. How does a protein fold? Nature 369:248–251, 1994.
20. Abkevich, V.I., Gutin, A.M., Shakhnovich, E.I. Specific nucleus as the transition state for protein folding: Evidence from the lattice model. Biochemistry 33:10026–10036, 1994.
21. Bryngelson, J.D., Wolynes, P.G. Intermediates and barrier crossing in a random energy model with application to protein folding. J. Phys. Chem. 93:6902–6915, 1989.
22. Socci, N.D., Onuchic, J.N. Folding kinetics of proteinlike heteropolymers. J. Chem. Phys. 101:1519–1528, 1994.
23. Ptitsyn, O.B. Structures of folding intermediates. Curr. Opin. Struct. Biol. 5:74–78, 1995.
24. Jackson, S.E., elMasry, N., Fersht, A.R. Structure of the hydrophobic core in the transition state for folding of chymotrypsin inhibitor 2: A critical test of the protein engineering method of analysis. Biochemistry 32:11270–11278, 1993.
25. Creighton, T.E. The energetics ups and downs of protein folding. Nature Struct. Biol. 1:135–138, 1994.
26. Milla, M.E., Brown, B.M., Waldburger, C.D., Sauer, R.T. P22 Arc repressor: Transition state properties inferred from mutational effects on the rates of protein unfolding and refolding. Biochemistry 34:13914–13919, 1995.
27. Roder, H., Elöve, G.A., Englander, S.W. Structural characterization of folding intermediates in cytochrome c by H-exchange labelling and proton NMR. Nature 335:700–704, 1988.
28. Bai, Y., Sosnick, T., Mayne, L., Englander, S.W. Protein folding intermediates studied by native-state hydrogen exchange. Science 269:192–197, 1995.
29. Bai, Y., Milne, J.S., Mayne, L., Englander, S.W. Primary structure effects on peptide group hydrogen exchange. Proteins 17:75–86, 1993.
30. Wand, A.J., DiStefano D.L., Feng, Y., Roder, H., Englander, S.W. Proton resonance assignments of horse ferrocytochrome c. Biochemistry 28:186–194, 1989.
31. Tsong, T.Y. Ferricytochrome c chain folding measured by the energy transfer of tryptophan to the heme group. Biochemistry 15:5467–5473, 1976.
32. Förster, T.H. Delocalized excitation and excitation transfer. In: "Modern Quantum Chemistry." Sinanoglu, O. (ed.). New York: Academic Press, 1965:93–137.
33. Schechter, E., Saludjian, P. Conformation of ferricytochrome c. IV. Relationship between optical absorption and protein conformation. Biopolymers 5:788–790, 1967.
34. Ridge, J.A., Baldwin, R.L., Labhardt, A.M. Nature of the fast and slow refolding reactions of iron(III) cytochrome c. Biochemistry 20:1622–1630, 1981.
35. Wood, L.C., Muthukrishnan, K., White, T.B., Ramdas, L., Nall, B.T. Replacement of a conserved proline eliminates the absorbance-detected slow folding phase of iso-2-cytochrome c. Biochemistry 27:8562–8568, 1988.
36. Pace, C.N. The stability of globular proteins. CRC Crit. Rev. Biochem. 3:1–43, 1975.
37. Makhatadze, G.I., Privalov, P.L. Protein interactions with urea and guanidinium chloride. A calorimetric study. J. Mol. Biol 226:491–505, 1992.
38. Matthews, C.R. Effects of point mutations on the folding of globular proteins. Methods Enzymol. 154:498–511, 1987.
39. Makhatadze, G.I., Privalov, P.L. Heat capacity of proteins I. Partial molar heat capacity of individual amino acid residues in aqueous solution: Hydration effect. J. Mol. Biol. 213:375–384, 1990.
40. Xie, D., Freire, E. Molecular basis of cooperativity in protein folding. V. Thermodynamic and structural conditions for the stabilization of compact denatured states. Proteins 19:291–301, 1994.
41. Privalov, P.L., Khechinashvili, N.N. A thermodynamic approach to the problem of stabilization of globular protein structure: A calorometric study. J. Mol. Biol. 86:665–684, 1974.
42. Privalov, P.L., Gill, S.J. Stability of protein structure and hydrophobic interactions. Adv. Protein Chem. 39:191–234, 1988.
43. Kuroda, Y. Residual helical structure in the C-terminal fragment of cytochrome c. Biochemistry 32:1219–1224, 1993.
44. Wu, L.C., Laub, P.B., Elöve, G.A., Carey, J., Roder, H. A noncovalent peptide complex as a model for an early folding intermediate of cytochrome c. Biochemistry 32:10271–10276, 1993.
45. Tanford, C. Protein denaturation. Adv. Protein Chem. 23: 121–282, 1968.
46. Tanford, C. Protein denaturation. C. Theoretical models for the mechanism of denaturation. Adv. Protein Chem. 24:1–95, 1970.
47. Agashe, V.R., Shastry, M.C.R., Udgaonkar, J.B. Initial hydrophobic collapse in the folding of barstar. Nature 377: 754–757, 1995.
48. Elove, G.A., Chaffotte, A.F., Roder, H., Goldberg, M.E. Early steps in cytochrome c folding probed by time-resolved circular dichroism and fluorescence spectroscopy. Biochemistry 31:6876–6883, 1992.
49. Goldberg, M.E., Guillou, Y. Native disulfide bonds greatly accelerate secondary structure formation in the folding of lysozyme. Protein Sci. 3:883–887, 1994.
50. Chaffotte, A.F., Cadieux, C., Guillou, Y., Goldberg, M.E. A possible initial folding intermediate: The C-terminal proteolytic domain of tryptophan synthase beta chains folds in less than 4 milliseconds into a condensed state with nonnative-like secondary structure. Biochemistry 31:4303–4308, 1992.
51. Radford, S.E., Dobson, C.M., Evans, P.A. The folding of hen lysozyme involves partially structured intermediates and multiple pathways. Nature 358:302–307, 1992.
52. Englander, S.W., Mayne, L. Protein folding studied using hydrogen-exchange labeling and two-dimensional NMR. Annu. Rev. Biophys. Biomol. Struct. 21:243–265, 1992.
53. Matthews, C.R. Pathways of protein folding. Annu. Rev. Biochem. 62:653–683, 1993.
54. Muthukrishnan, K., Nall, B.T. Effective concentrations of amino acid side chains in an unfolded protein. Biochemistry 30:4706–4710, 1991.
55. Peng, Z.Y., Wu, L.C., Kim, P.S. Local structural preferences in the alpha-lactalbumin molten globule. Biochemistry 34:3248–3252, 1995.
56. Jacobson, H., Stockmayer, W.H. Intramolecular reactions and polycondensation: I. The theory of linear systems. J. Chem. Phys. 18:1600–1606, 1950.

57. Thomas, D.J. The entropic tension of protein loops. J. Mol. Biol. 216:459–65, 1990.
58. Bromberg, S., Dill, K. Side-chain entropy and packing in proteins. Protein Sci. 3:997–1009, 1994.
59. Honig, B., Sharp, K., Yang, A.S. Macroscopic models of aqueous-solutions—biological and chemical applications. J. Phys. Chem. 97:1101–1109, 1993.
60. Otzen, D.E., Itzhaki, L.S., elMasry, N.F., Jackson, S.E., Fersht, A.R. Structure of the transition state for the folding/unfolding of the barley chymotrypsin inhibitor 2 and its implications for mechanisms of protein folding. Proc. Natl. Acad. Sci. USA 91:10422–10425, 1994.
61. Kim, P.S., Baldwin, R.L. Specific intermediates in the folding reactions of small proteins and the mechanism of protein folding. Annu. Rev. Biochem. 51:459–489, 1982.
62. Jones, C.M., Henry, E.R., Hu, Y., Chan, C.K., Luck, S.D., Bhuyan, A., Roder, H., Hofrichter, J., Eaton, W.A. Fast events in protein folding initiated by nanosecond laser photolysis. Proc. Natl. Acad. Sci. USA 90:11860–11864, 1993.
63. Milla, M.E., Sauer, R.T. P22 Arc repressor: Folding kinetics of a single-domain, dimeric protein. Biochemistry 33: 1125–1133, 1994.
64. Zitzewitz, J.A., Bilsel, O., Luo, J., Jones, B.E., Matthews, C.R. Probing the folding mechanism of a leucine zipper peptide by stopped-flow circular dichroism spectroscopy. Biochemistry 34:12812–12819, 1995.
65. Guo, Z.Y., Thirumalai, D. Kinetics of protein-folding: Nucleation mechanism, time scales, and pathways. Biopolymers 36:83–102, 1995.
66. Anfinsen, C.B. Principles that govern the folding of protein chains. Science 181:223–230, 1973.
67. Wetlaufer, D.B. Nucleation, rapid folding, and globular intrachain regions in proteins. Proc. Natl. Acad. Sci. USA 70:697–701, 1973.
68. Thirumalai, D., Guo, Z. Nucleation mechanism for protein folding and theoretical predictions for hydrogen-exchange experiments. Biopolymers 35:137–140, 1995.
69. Abkevich, V.I., Gutin, A.M., Shakhnovich, E.I. Free energy landscape for protein folding kinetics: Intermediates, traps, and multiple pathways in theory and lattice model simulations. J. Chem. Phys. 101:6052–6062, 1994.
70. Srinivasan, R., Rose, G.D. LINUS: A hierarchical procedure to predict the fold of a protein. Proteins 224:81–99, 1995.
71. Zhang, J.X., Goldenberg, D.P. Amino acid replacement that eliminates kinetic traps in the folding pathway of pancreatic trypsin inhibitor. Biochemistry 32:14075–14081, 1993.
72. Todd, M.J., Viitanen, P.V., Lorimer, G.H. Dynamics of the chaperonin ATPase cycle: Implications for facilitated protein folding. Science 265:659–666, 1994.
73. Weissman, J.S., Kashi, Y., Fenton, W.A., Horwich, A.L. GroEl- mediated protein folding proceeds by multiple rounds of binding and release of nonnative forms. Cell 78: 693–702, 1994.
74. Levinthal, C. Are there pathways for protein folding? J. Chim. Phys. 65:44–45, 1968.
75. Chen, B., Baase, W.A., Schellman, J.A. Low-temperature unfolding of a mutant of phage T4 lysozyme. 2. Kinetic investigations. Biochemistry 29:691–699, 1989.
76. Jeng, M.F., Englander, S.W. Stable submolecular folding units in a non-compact form of cytochrome c. J. Mol. Biol. 221:1045–1061, 1991.
77. Fersht, A. Proc. Natl. Acad. Sci. USA 92:10869–10873, 1995.
78. Gutin, A.M., Abkevich, V.I., Shakhnovich, E.I. Biochemistry 34:3066–3076, 1995.