Acc. Chem. Res. **1998,** *31,* 727−736

Cytochrome *c* Folding and Unfolding: A Biphasic Mechanism

SYUN-RU YEH, SANGHWA HAN, AND DENIS L. ROUSSEAU*

Department of Physiology and Biophysics, Albert Einstein College of Medicine, Bronx, New York 10461

Received December 12, 1997

Introduction

How an unstructured polypeptide chain condenses to form a biologically functional protein, the protein folding problem, has attracted great interest in recent years.[1] In many proteins, this spontaneous self-organization reaction occurs with high efficiency and fidelity despite a marginal thermodynamic driving force.[2] A full knowledge of the free energy landscape associated with all possible configurations of the polypeptide backbone along the folding coordinate is required for a thorough understanding of this process.[3,4] Unfortunately, calculations and measurements of the energy and entropy of dynamic structures on the molecular level are nontrivial tasks. It is, therefore, beneficial to construct general principles of folding based on simple model systems, such as cytochrome *c*.

Cytochrome *c* (cyt *c*) is a small single-domain protein with a well-defined three-dimensional structure[5] as illustrated in Figure 1. The protoporphyrin IX prosthetic group is covalently linked to the protein moiety through thioether linkages via Cys14 and Cys17. The robust protein architecture is made possible by the sophisticated H-bonding network throughout the molecule and the intramolecular linking between His18 and Met80 mediated by the heme central iron atom.[6] Cyt *c* is an ideal model system for folding studies because it can be unfolded and refolded reversibly without degradation of the protein.[7] Furthermore, the heme prosthetic group stays intact as the protein unfolds due to the covalent linkages between the heme and the protein moiety such that the refolding kinetics are not complicated by bimolecular rebinding of the heme to the polypeptide chain. Another advantage for using cyt *c* as a model system is that the various folding intermediates present during the course of folding can be regulated by a variety of chemical means, such as the pH and the concentration of denaturant.[8−10] This provides the basis for understanding the contributions of the various on- and off-pathway intermediates in the folding energy landscape.[3,4]

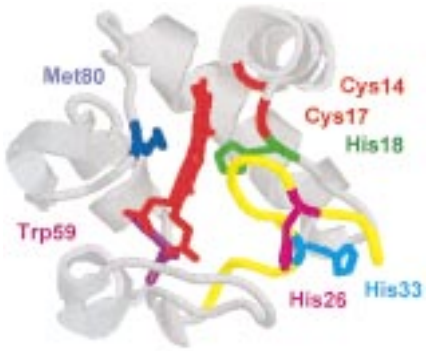


FIGURE 1. Structure of cytochrome *c*.[5] The prosthetic heme group, located in the center of the protein and colored red, is covalently bound to the polypeptide backbone by Cys14 and Cys17, also colored red. The heme iron is coordinated by His18 (green) and Met80 (dark blue) in the native form. Under unfolding conditions His26 (magenta) or His33 (cyan) may coordinate to the heme in place of Met80. The single tryptophan at position 59 is colored purple. In this figure the covalent thioether linkages and the heme coordination bonds are not shown.

There is no universal probe that can provide snapshots of the conformation of each individual polypeptide chain with molecular resolution during the course of folding. However, a great deal of valuable information on specific aspects of cyt *c* folding has been obtained from a variety of spectroscopic techniques.[11] For example, nuclear magnetic resonance (NMR) has been demonstrated to be very useful for probing solvent exposure of exchangeable protons on the polypeptide backbone.[12] Techniques such as circular dichroism (CD)[13] and small-angle X-ray scattering[14] provide information on the secondary structure content and average molecular size of the protein molecule, respectively. Tryptophan fluorescence and optical absorption spectroscopies are informative concerning local interchain interactions by taking advantage of the presence of the heme prosthetic group.[15,16] To open a new window, we sought to employ resonance Raman scattering, a structurally rich technique, to go beyond the phenomenological observations and examine quantitatively the microscopic details of the folding reaction.[8−10,17,18]

To study protein folding dynamics, the folding reaction must be initiated in a pulsed manner, such that all of the

Syun-Ru Yeh received a B.S. from the National Cheng Kung University and an M.S. from the National Tsing Hua University in Taiwan. She received her Ph.D. in chemistry from the University of Maryland in 1993 with Daniel E. Falvey. Subsequently, she became a Research Fellow at Harvard Medical School, a Research Associate at Princeton University, and a Consultant at AT&T Bell Laboratories. In 1995, she joined Albert Einstein College of Medicine where she is now an Instructor in the Department of Physiology and Biophysics.

Sanghwa Han received a B.S. from Seoul National University in Korea and a Ph.D. in Chemistry from Princeton University in 1988 with Thomas G. Spiro. After three years of postdoctoral research with Denis L. Rousseau at AT&T Bell Laboratories, he returned to Korea where he is now an Associate Professor of Biochemistry at Kangwon National University. In 1997 he spent a sabbatical year in the Rousseau group at Albert Einstein College of Medicine.

Denis L. Rousseau received a B.A. from Bowdoin College and an M.S. and a Ph.D. from Princeton University. He spent two years as a Postdoctoral Fellow with S. P. S. Porto at the University of Southern California before joining AT&T Bell Laboratories in 1969. He joined Albert Einstein College of Medicine in 1996 where he is now a Professor of Physiology and Biophysics.

* To whom correspondence should be addressed. Phone: (718) 430-4264. Fax: (718) 430-4230. E-mail: rousseau@aecom.yu.edu.

10.1021/ar970084p CCC: $15.00
Published on Web 11/18/1998

molecules are in phase at the time of observation. This is commonly achieved in a stopped flow apparatus by diluting denaturant-unfolded protein into a solution favoring the formation of the native structure. The earliest events that can be monitored with conventional instrumentation are limited by the mixing deadtime of a few milliseconds by which time a significant fraction of the folding in cyt *c* is already complete, the so-called "burst" phase.[19] To overcome this restriction, ingenious photochemical protocols, such as photodissociation of the CO-bound reduced form,[20] or photoreduction of the oxidized form,[21,22] have been developed taking advantage of the characteristic ligand binding and redox properties of cyt *c*. Despite considerable success in exploring fast folding events in cyt *c*, the application of these strategies is limited to the ferrous protein and the presence of relatively high concentrations of denaturant. To obtain a more general tool, we have developed a rapid solution mixer with a 100 μs deadtime.[8] Combined with resonance Raman scattering it has enabled us to follow the folding of cyt *c* in an unexplored time domain with molecular level sensitivity.[8−10,17,18]

What Do We Learn from Resonance Raman Scattering?

One of the main challenges in addressing the protein folding problem is to visualize folding intermediates along the folding coordinate. Analogous to the disulfide bonds in BPTI,[23] the heme prosthetic group in cyt *c* traps the folding intermediates and makes them visible to numerous spectroscopic probes by coordination with various amino acid residues, such as the native ligands, His18 and Met80, and the non-native ligands, His26 and His33, and a solvent water molecule.[8−10,16−18,24,25] Resonance Raman spectroscopy has been demonstrated to be especially useful in this respect for its high sensitivity and structure specificity.[8−10]

Resonance Raman spectra of heme proteins have been extensively studied, such that the general characteristics of the spectra with respect to the heme ligation states are well established.[26−28] Particularly useful are the modes in the high-frequency region of the spectrum which are sensitive to both the axial coordination and spin state of the iron at the center of the heme such as those in the 1475−1520 cm^{-1} (ν_3) and 1550−1600 cm^{-1} (ν_2) regions as shown in Figure 2. On the basis of this knowledge, four folding intermediates of ferric cyt *c* with differing heme coordination states were identified:[8−10] (1) the native form (HM) in which His18 and Met80 are the axial ligands, (2) a bis-histidine form (HH) where Met80 is replaced by His26 or His33, (3) a histidine−water form (HW) where Met80 is replaced by a water molecule, and (4) a five-coordinated form (5C) where both the Met80- and His18−iron bonds are broken and, instead, a water molecule is ligated to the heme as the single axial ligand. In a similar manner, high- and low-spin coordination states have also been identified in ferrous cyt *c*.[18]

The development of the native tertiary intrachain

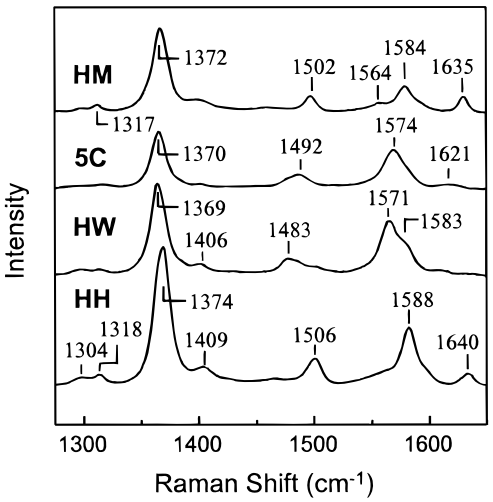


FIGURE 2. Deconvoluted resonance Raman spectra of the various heme coordinated forms of cyt *c*. HM is the native form with histidine and methionine as heme axial ligands. 5C is the five-coordinated form with a single water molecule as its axial ligand. HW and HH are the histidine−water and the histidine−histidine coordinated forms, respectively.

interactions during the folding of cyt *c* can be probed by the low-frequency resonance Raman spectrum (300−800 cm^{-1}), which, in contrast to the high-frequency spectrum, is very sensitive to the polypeptide conformation near the heme.[9,27,28] In the native protein, the heme group is highly ruffled[5] due to the tension exerted on it by the surrounding polypeptide architecture (Figure 1). As a result, the low-frequency spectra of both the ferric and the ferrous hemes display rich resonance Raman spectra, congested with heme-peripheral substituent modes, porphyrin out-of-plane modes, and modes that are normally only infrared active[9,27,28] as shown in Figure 3 for both oxidation states. When the protein is unfolded by strong denaturants, the heme adopts a planar structure due to the relaxation of the tertiary structure associated with the disruption of the heme−ligand interactions. Consequently, the spectrum becomes rather simple with only a few features. Among these many vibrational modes, the Raman lines at 397 and 394 cm^{-1} in the ferric (Figure 3a) and ferrous (Figure 3b) states, respectively, are especially useful, and can be used as marker lines for the formation of the native tertiary structure, since they involve motion of the covalent linkage between the heme and the polypeptide. The lines have been assigned as bending modes involving the sulfur atom of Cys14/17, the β-carbon atom of the heme, and the atom bridging these two.[27]

Unfolded State of Ferric Cytochrome *c*

From resonance Raman studies, we have been able to characterize the heme ligation states of the unfolded protein.[8−10,17] The equilibrium unfolding curves at pH 5.9 measured by CD and tryptophan fluorescence coincide with that from resonance Raman measurements of the native HM state with a transition midpoint of approximately 2.5 M Gdn−HCl[17] (Figure 4a). This indicates that the disruption of the Met80−iron bond occurs concomi-

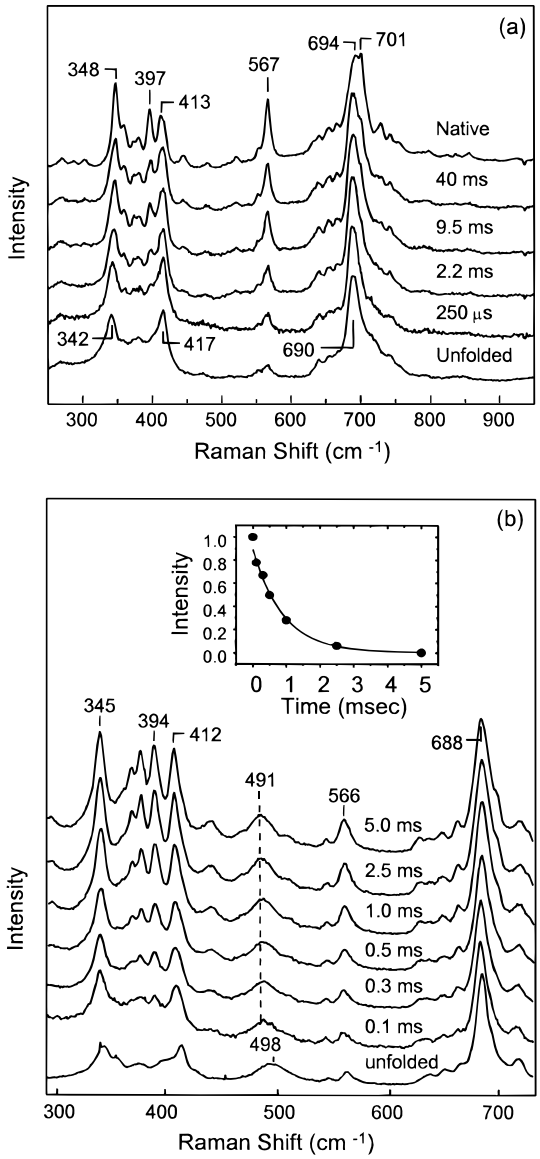


FIGURE 3. Time evolution of the low-frequency resonance Raman spectra of cyt *c* following the initiation of folding. (a) The ferric form of the protein was unfolded at pH 4.7 in 4.4 M Gdn−HCl and refolded to pH 4.9 in 0.7 M Gdn−HCl. (b) The ferrous CO-bound form of the protein was unfolded in 7.0 M Gdn−HCl and refolded in 0.7 M Gdn−HCl at neutral pH. The inset is the decay curve of the unfolded form obtained by monitoring the intensity of the line at 394 cm^{-1}. The line at 491−498 cm^{-1} is assigned as the Fe−CO stretching mode.

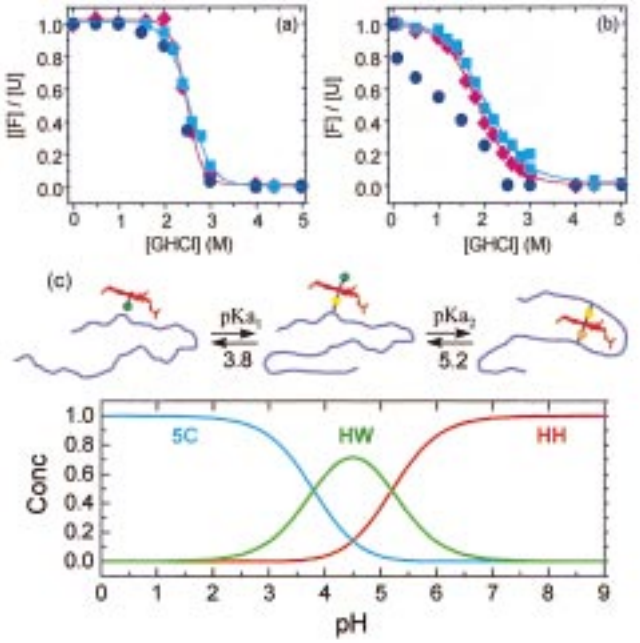


FIGURE 4. Equilibrium titration curves for the unfolding of cyt *c* by Gdn−HCl at pH 5.9 (a) and pH 3.0 (b). The dark blue circles represent the population of the native HM structure determined from resonance Raman measurements of the histidine−methionine coordinated species. The cyan squares were determined from the intensity of the tryptophan fluorescence, and the magenta diamonds were determined from CD measurements. (c) Populations of the various heme ligand coordination states as a function of pH. Above pK_{a2} (5.2) the heme is six-coordinate with its native His18 as one axial ligand and either His26 or His33 as the other ligand (HH). At lower pH, His26/33 is protonated and replaced by a solvent water molecule (HW). Below pK_{a1} (3.8), His18 becomes protonated as well and the heme becomes five-coordinate with a water molecule as its axial ligand (5C). In the schematic illustration in the middle, the covalent linkages between the heme and Cys14 and Cys17 are not shown. The green circle represents water, the yellow circle represents His18, and the orange circle represents His26/33.

tantly with the loss of the secondary structure and the increase in the molecular size under equilibrium unfolding conditions. In contrast, the unfolding curve at pH 3.0, from resonance Raman measurements of the HM state, deviates significantly from that measured by CD and fluorescence in the pretransition region (Figure 4b),[17] indicating that the iron−methionine bond is destabilized at low pH such that it is ruptured more easily than the global unfolding of the protein. Moreover, the transition midpoint of 2.0 M is lower than that at pH 5.9, suggesting that, in addition to the iron−methionine bond, the protein architecture is also destabilized under acidic conditions probably due to the electrostatic repulsion between the polypeptide segments subjected to protonation.[29]

In the presence of 4.4 M Gdn−HCl at near neutral pH, the polypeptide chain attains a random-coil-like structure in which the native His18 ligand is retained but the native Met80 ligand is replaced by either His26 or His33 as the sixth heme ligand (HH) as illustrated in Figure 4c. It was recently shown that His33 is the predominant non-native heme ligand in the denatured state of cyt *c* on the basis of the studies of His mutants.[30] When the pH is reduced, His26/33 is protonated and replaced by a solvent water molecule (HW). At even lower pH, His18 becomes protonated as well, and a species in which the heme is coordinated by only one water molecule (5C) is formed. The populations of these three ligation states are determined by an acid−base equilibrium, in which pK_{a1} is 3.8 and pK_{a2} is 5.2 (Figure 4). Due to the steric constraints imposed on His18 by the nearby thioether bonds between the heme and Cys14/17, the pK_a of the protonated His18 is lower than that of the protonated His26/33. These pH-dependent properties of the unfolded protein provide an excellent means to systematically control the starting point of the folding reaction and the ending point of the unfolding reaction, such that the early folding and unfold-

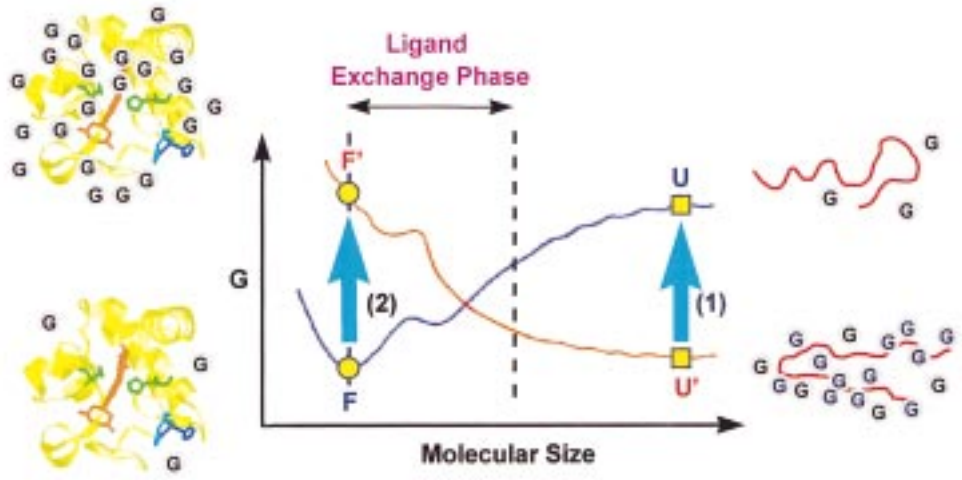


FIGURE 5. Schematic illustration of the free energy surfaces of cytochrome *c* during folding and unfolding. Under native conditions indicated by the blue curve, the energy of the unfolded protein (U) is higher than that of the folded form (F). The free energy is reduced along the U to F trajectory as the unstructured polypeptide chain reduces its molecular size. It reaches an energy minimum when the polypeptide assumes its native conformation. When the size of the polypeptide chain is further reduced, the energy is elevated due to the repulsion imposed by steric interactions between the spatially neighboring amino acid residues. On the other hand, under denaturing conditions indicated by the red curve, the unfolded polypeptide molecule (U′) is more stable than its "folded" counterpart (F′). As the size of the polypeptide chain is reduced, the energy increases correspondingly along the U′ to F′ trajectory. Unlike the unique F state under native conditions, there is no well-defined "folded" state (F′) under denaturing conditions. When the folding reaction is initiated by rapidly diluting out the denaturant, the protein is promoted from the U′ to U state instantaneously as indicated by arrow 1. If the initiation of the reaction is fast enough with respect to the conformational changes of the polypeptide chain, the folding course from the U to F state can be followed. The unfolding reaction from F′ to U′ can also be initiated as indicated by arrow 2, and subsequently followed. For cytochrome *c*, the refolding and unfolding reactions share a common ligand exchange phase as indicated.

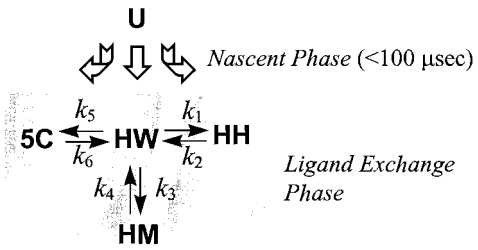
ing events can be unraveled and the influence of these differing unfolded states on the overall folding/unfolding kinetics can be assessed.

A Biphasic Model for Cytochrome *c* Folding

The conventional method to initiate a folding reaction is to rapidly dilute the denaturant-unfolded protein with excess of buffer in rapid mixing instrumentation such as a stopped flow or a continuous flow apparatus. Instantaneously diluting the denaturant can be viewed conceptually as promoting a protein molecule from an unfolding free energy surface (red curve) to a folding free energy surface (blue curve) as illustrated in Figure 5. The energetically downhill folding reaction from U to F, with or without barriers, can then be followed in real time with various spectroscopic techniques. The early part of the folding kinetics of cyt *c* has been lost in the past since the motion of the polypeptide chain is faster than the mixing deadtime of commercially available instrumentation.

To elucidate the characteristic features of the folding free energy surface in Figure 5, a rapid solution mixer with a 100 μs deadtime was developed.[8] Along with resonance Raman scattering, a cohesive picture defined in Scheme 1 and illustrated schematically in Figure 6 is emerging. In this model, the unstructured polypeptide chain becomes semicompact in less than 100 μs, which we have termed the *nascent phase* of folding. The nascent phase is followed by a *heme ligand exchange phase* in which all the potential heme ligands attached to the polypeptide chain, such as His18, Met80, His26, and His33, as well as

Scheme 1



exogenous water molecules, are exchanging with each other through random structural fluctuations until the final native HM conformation is reached. To demonstrate the salient features of this biphasic model, some of the experiments performed in our laboratory[8−10,17,18] will be illustrated below.

In Figure 7a, time-dependent resonance Raman data are presented in which ferric cyt *c* is refolded from the Gdn−HCl denatured state at an initial pH of 3.6 and a final pH of 5.0. Clear changes are evident in the spectra which indicate a change from a high-spin to a low-spin heme. To extract quantitative information, the spectra were deconvoluted into four possible heme coordination states, HM, HH, HW, and 5C, on the basis of the standard spectra shown in Figure 2. The resulting populations of these species are plotted in Figure 8a. The solid lines are a best fit to the data with the ligand exchange model described in Scheme 1. The rates are listed in Table 1. The significant difference between the extrapolated data at $t = 0$ and those obtained from equilibrium unfolding conditions, as indicated by the arrows, verifies the exist-

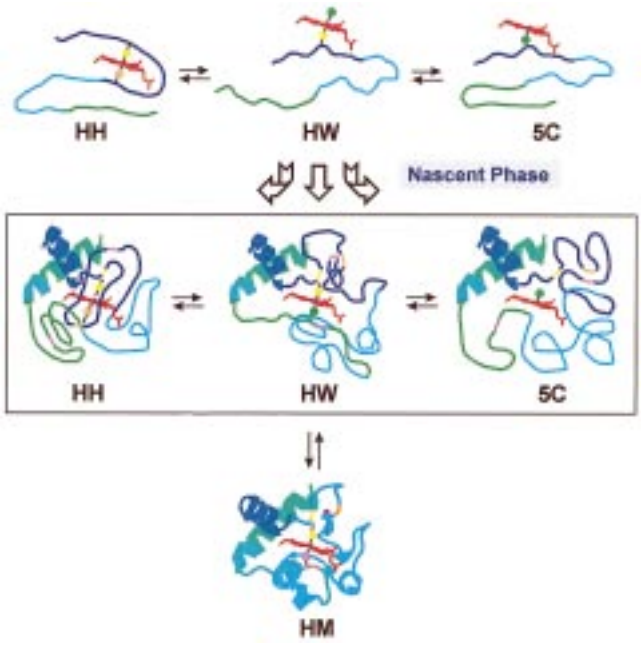


FIGURE 6. Schematic illustration of the folding dynamics of cytochrome *c*. Under unfolding conditions, the bonds between the heme iron atom and Met80 (and/or His18) can be broken and the protein assumes a random-coil-like structure. The populations of the three heme coordination states, HH, HW, and 5C, are controlled by a fast equilibrium. Following the initiation of folding, a redistribution of these heme ligation states is accompanied by the collapse of the unstructured polypeptide chain in the *nascent phase of folding*. In the subsequent *ligand exchange phase of folding*, a slower change in the population of these heme ligation states takes place, ultimately leading to the native state (HM). The water ligand for 5C can be on either side of the heme plane.

ence of the nascent phase occurring within the mixing deadtime of the instrument (100 μs). It is important to point out that the nascent phase is not simply reflecting the pH jump from pH 3.6 to pH 5.0; otherwise a significant population of HH would be present at the end of this phase since ~40% of HH is expected at pH 5.0 on the basis of Figure 4c.

Similar results were observed during the folding of ferrous cyt *c*[18] as shown in Figure 7b. Large time-dependent changes, which are essentially complete within 3 ms, indicate a conversion from a five-coordinated species to a six-coordinated species. At longer times additional small changes lead to the native HM spectrum. Unlike ferric cyt *c*, it is hard to quantitatively analyze these data. Careful assignments of the six-coordinate states, which are currently in progress, are required for more detailed understanding of this reaction. Qualitatively, these data demonstrate the existence of the nascent phase during the folding of ferrous cyt *c*, since a substantial change in the ligand coordination states (from five- to six-coordinate) has occurred at the earliest time point (100 μs) with respect to the equilibrium unfolded protein.

Nascent Phase of Folding

The nascent phase of folding is kinetically distinct from the ligand exchange phase. This is evident in the folding

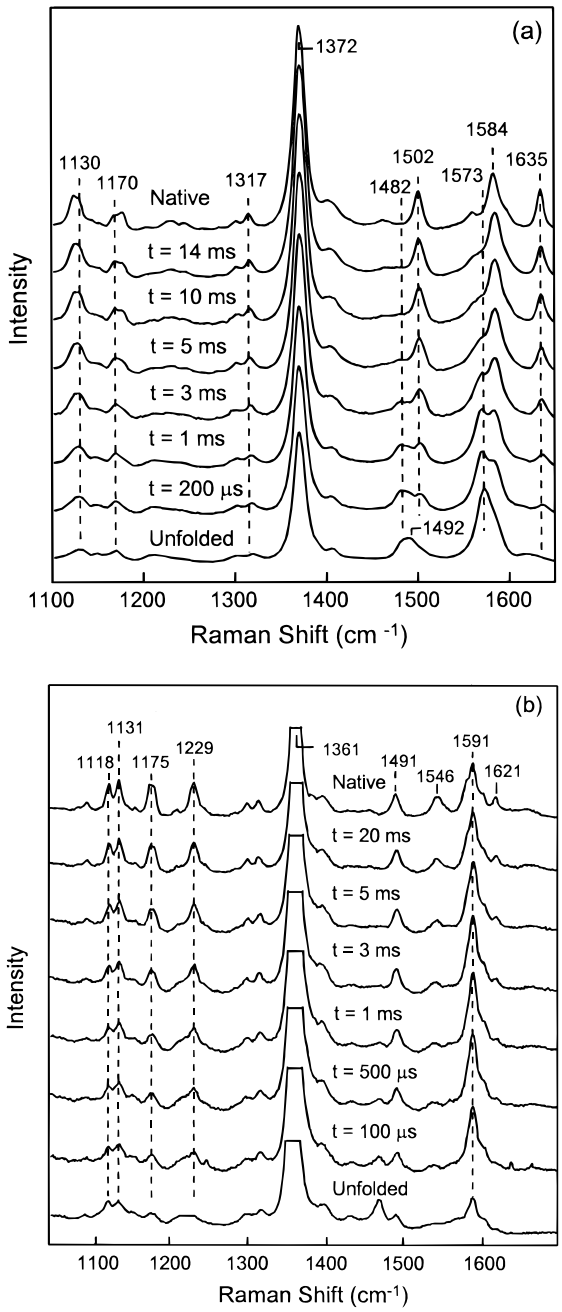


FIGURE 7. Time-dependent resonance Raman spectra of cytochrome *c* in the high-frequency region during the course of folding. (a) Ferric cyt *c* was unfolded at pH 3.6 in 4.4 M Gdn−HCl and refolded at pH 5.0 to a final Gdn−HCl concentration of 0.7 M. (b) Ferrous cyt *c* was unfolded in 7.0 M Gdn−HCl and refolded in 1.2 M Gdn−HCl.

experiments from pH 3.6 to pH 5.0 (Figure 8a) where approximately 50% of 5C is converted to HW during the nascent phase. The associated His18 ligation reaction is therefore faster than 10,000 s^{-1} since it is kinetically unresolved in our rapid mixing instrumentation that has a 100 μs deadtime. This rate is much faster than that of the same reaction occurring during the subsequent ligand exchange phase, 1500 s^{-1} (k_6 in Table 1). Analogous results were observed when folding was initiated at an initial pH of 4.7 and refolded at a final pH of 4.5 as shown in Figure 8b. In this experiment, the population of HH

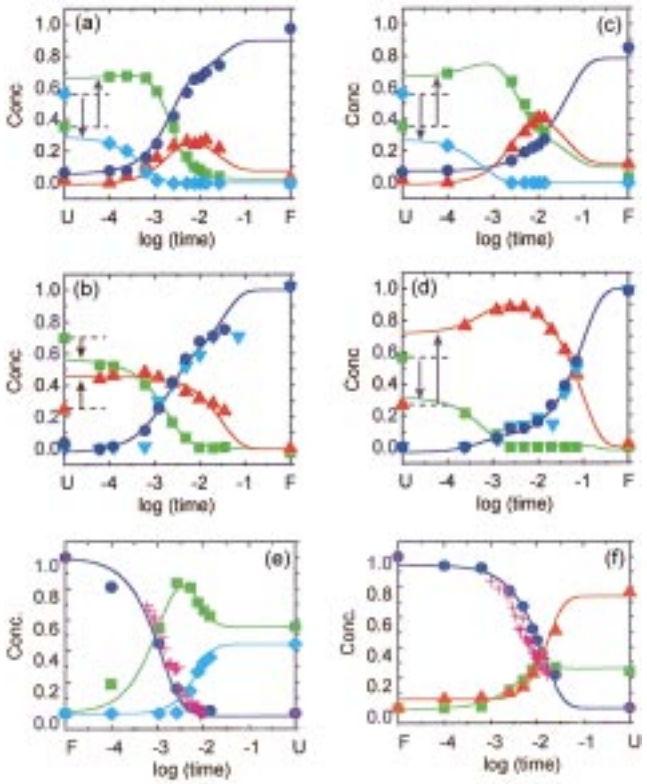


FIGURE 8. (a–d) Time course for the populations of the heme–ligand coordinated forms during the folding of ferric cyt *c* under various conditions. (a) The unfolded protein at pH 3.6 in 4.4 M Gdn–HCl was refolded at pH 5.0 in 0.7 M Gdn–HCl. (b) The unfolded protein at pH 4.7 in 4.4 M Gdn–HCl was refolded at pH 4.5 in 0.7 M Gdn–HCl. (c) The unfolded protein at pH 3.6 in 4.4 M Gdn–HCl was refolded at pH 5.0 in 1.9 M Gdn–HCl. (d) The unfolded protein at pH 4.7 in 4.4 M Gdn–HCl was refolded at pH 5.9 in 0.7 M Gdn–HCl. The green squares, the red triangles, the dark blue circles, and the cyan diamonds are the populations of the HW, HH, HM, and 5C forms, respectively. The cyan wedges in plots b and d represent the intensity of the 397 cm^{-1} line. The solid lines are fits to the data using the model presented in Scheme 1. The arrows show the changes that occur in the nascent phase of folding. On the *x*-axes, U and F represent the unfolded and the native states, respectively. (e, f) Time course for the populations of the heme–ligand coordinated forms during the unfolding of cyt *c*. The folded protein at pH 5.9 was unfolded at pH 3.8 (e, bottom left) and at pH 5.8 (f, bottom right) in 4.4 M Gdn–HCl. The magenta crosses were determined from the intensity of the tryptophan fluorescence. See (a–d) for the definition of the other symbols. The solid lines are fits to the data using the model presented in Scheme 2. The rate constants are listed in Table 1.

almost doubles during the nascent phase of folding. The associated His26/33 ligation rate decreases from >10,000 s^{-1} in the nascent phase to 9 s^{-1} in the ligand exchange phase (k_1 in Table 1). This tremendous discrepancy in the histidine ligation rates demonstrates that the two folding phases are very different kinetically. The histidine ligation reactions occurring during the nascent phase are accompanied by compaction of the polypeptide backbone due to hydrophobic collapse.[31–35] This assumption is supported by tryptophan fluorescence quenching experiments performed in our laboratory as well as in others,[31–35] in which the tryptophan fluorescence is significantly quenched within 50–100 μs. The compaction of the polypeptide chain raises the effective concentration of histidine relative to the heme and therefore facilitates histidine ligation during the nascent phase. On the other hand, the condensation of the polypeptide backbone also causes it to be stiffer and thus retards histidine ligation during the subsequent ligand exchange phase.

Table 1. Rate Constants (Schemes 1 and 2) Determined from Fitting the Time-Resolved Resonance Raman Data for Folding and Unfolding Ferric Cytochrome *c* under Various Conditions

	k_1 (s^{-1})	k_2 (s^{-1})	k_3 (s^{-1})	k_4 (s^{-1})	k_5 (s^{-1})	k_6 (s^{-1})
U (pH 3.6) → F(pH 5.0)[a]	160	50	300	8	0	1500
U (pH 4.7) → F(pH 4.5)[a]	9	33	503	0	0	0
U (pH 3.8) → F(pH 5.0)[b]	160	130	40	5	0	1500
U (pH 4.7) → F(pH 5.9)[a]	1034	33	503	0	0	0
F (pH 5.9) → U(pH 5.8)[a]	170	60	0	70	0	0
F (pH 5.9) → U(pH 3.8)[a]	0	0	0	800	65	81

[a] The concentration of Gdn–HCl in the refolding solution was 0.7 M. [b] The concentration of Gdn–HCl in the refolding solution was 1.9 M.

Despite the fact that the pH of the refolding solution (5.0) in Figure 8a is slightly higher than that (4.5) in Figure 8b, none of the protein molecules are in the HH state at the onset of the ligand exchange phase in the former, in contrast to a 40% population of HH in Figure 8b. Similarly, the five-coordinated species is present at the onset of the ligand exchange phase in Figure 8a but absent in Figure 8b. These observations indicate that the reaction that occurs during the nascent phase is not simply determined by the refolding conditions, but also depends on the starting conditions. This suggests that the reactions occurring during the nascent phase are subject to kinetic control rather than thermodynamic control; otherwise the ratio of the reaction products would simply reflect the final refolding conditions employed. It is likely that *each unstructured polypeptide chain folds uniquely depending on its starting conformation* during the nascent phase of folding.

In summary, the condensation of the polypeptide chain occurring during the nascent phase follows a kinetically controlled multiple-pathway mechanism. At the onset of the ligand exchange phase, the heterogeneous pathways start to merge. This multiple-pathway compaction reaction reduces the size of the conformational space and is associated with the formation of some secondary structure[36] but does not involve global formation of the specific tertiary contacts which stabilize the native structure. Once the conformational space is reduced, the native conformation can be reached in the subsequent ligand exchange phase through thermodynamically controlled structural fluctuations within a reasonable time frame as discussed below.

Ligand Exchange Phase of Folding

In the biphasic folding model shown in Scheme 1, the constructive folding pathway is from U to HW to HM. Unlike HW, the obligatory folding intermediate, HH and 5C are off-pathway intermediates. The conversion from

HH to HW during the ligand exchange phase is slow with a rate of approximately 30−50 s^{-1} at room temperature (Table 1) due to the stability of the histidine−iron bond and the associated misorganized structure. The HH intermediate thus forms a deep trap during the course of folding. The 5C intermediate, on the contrary, contains a more open structure due to the absence of the His18−iron bond[10] and can be converted to the HW form efficiently with a rate of 1500 s^{-1} (Table 1), which is 30−50 times faster than the conversion from HH to HW.

The validity of the various on- and off-pathway processes, illustrated in Scheme 1, is best demonstrated by the data shown in Figure 8a, where the pH of the unfolding solution is 3.6 and that of the folding solution is 5.0. In this experiment, HH is transiently populated, reaching a maximum at approximately 10 ms following the initiation of folding, even though no HH is present at the early stage of folding. Both the HH and 5C intermediates decay to HW which subsequently decays to the native HM state. The trapping effect of HH is best illustrated in Figure 8d in which the pH values of the unfolding and refolding solutions are 4.7 and 5.9, respectively. The high population of HH and the slow conversion rate from HH to the obligatory folding intermediate, HW, in the ligand exchange phase cause the overall folding reaction to be slow and heterogeneous. This can be improved by lowering the pH of the refolding solution such that the coordination of the "wrong" histidine ligands to the heme can be prevented by their protonation. An example of this is shown by the comparison between the data in Figure 8d to those in Figure 8b, in which the histidine on-rate, k_1, is reduced from 1034 to 9 s^{-1} by reducing the pH of the refolding solution from 5.9 to 4.5. The kinetic trapping effect of HH can also be reduced by increasing the concentration of the denaturant present in the refolding solution such that the misfolded structure associated with HH can be untangled more easily.[10] This is demonstrated in Figure 8c, in which the conversion rate from HH to HW (k_2) increased from 50 to 130 s^{-1} as the Gdn−HCl concentration in the refolding solution increased from 0.7 M (Figure 8a) to 1.9 M (Figure 8c). It is important to point out that although the rate from HH to HW is accelerated in the presence of high concentration of Gdn−HCl, the overall folding rate (rate of formation of HM) is retarded due to a decrease in the rate from HW to HM (k_3), which causes an accumulation of HW and in turn populates more HH during the folding reaction.

The reactions occurring during the ligand exchange phase are all reversible with rates in the range of 10−1500 s^{-1} (Table 1); thus, it is evident that the reactions occurring during the ligand exchange phase are subject to thermodynamic control, in which the final populations of the heme ligation states are determined by their relative free energies. Since all the molecules are funneled into the HM state, HM must be the lowest free energy state under these folding conditions! The thermodynamic parameters for these ligand exchange reactions can be derived from Arrhenius plots based on temperature-dependent studies[9] as shown in Figure 9. Three interesting issues arise here:

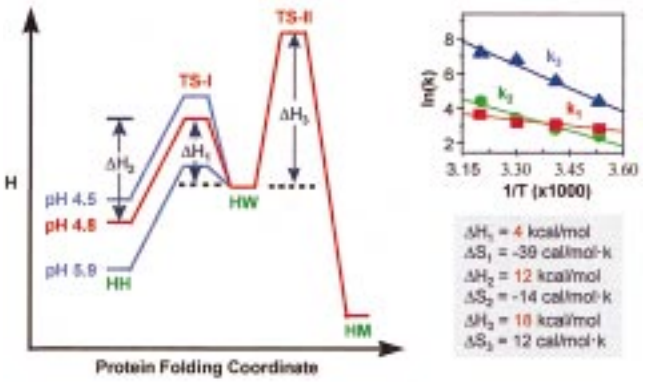


FIGURE 9. Arrhenius plots (pH 4.8) and the resulting potential energy diagram showing the activation energy barriers for the heme−ligand exchange reactions present during cyt *c* folding. In the potential energy diagram, the transition state between the HW and the HH form is designated by TS-I and that between the HW and the HM form is designated as TS-II. The red curve represents the potential energy curve of folding at pH 4.8. The corresponding changes in enthalpy and entropy are listed on the bottom right.

First, the linear relationship between $\ln(k_i)$ and $1/T$ suggests that the ligand exchange reactions are not accompanied by global conformational changes and therefore the Arrhenius equation can hold. This confirms the proposal that the protein molecules are already in a semicompact structure at the onset of the ligand exchange phase. Second, the enthalpic and entropic barriers for forming HH from HW are 4 kcal/mol and −39 cal/mol·K, respectively. Those for the reverse reaction are 12 kcal/mol and −14 cal/mol·K, respectively. HH is therefore 8 kcal/mol more stable than HW enthalpically; however, entropically it is 7 kcal/mol less stable than the latter at room temperature. The free energy change for the conversion from HW to HH is therefore close to zero. Consequently, the equilibrium between HW and HH can be easily shifted by controlling the ambient temperature. Last, the enthalpic barriers for forming HH and HM from HW are 4 and 18 kcal/mol, respectively. The corresponding entropic barriers are −39 and 12 cal/mol·K, respectively. The kinetic barrier for the folding to the native HM state is therefore mostly enthalpic rather than entropic, in contrast to that for forming the misfolded HH state. Due to the differences in the entropic barrier (−39 vs 12 cal/mol·K), the ligand exchange equilibrium can again be shifted by tuning the ambient temperature. The same effect can also be achieved by controlling the pH of the refolding solution, which significantly changes the enthalpic barrier[9] for forming HH from HW as illustrated in Figure 9.

Formation of the Native Polypeptide Conformation

The low-frequency resonance Raman spectrum, in addition to being sensitive to the ligand coordination, also yields information concerning the tertiary interactions between the heme and its surroundings. A direct com-

parison of the low-frequency data with those from high-frequency Raman measurements is informative.[9] The development of the 397 cm^{-1} line in ferric cyt *c* (e.g., Figure 3a) is especially useful and can be used as a marker line for the ruffled heme since it is a mode involving motion of the thioether linkage between the heme and the protein moiety. The excellent agreement between the formation of the 397 cm^{-1} line and the formation of the Met80 ligation, as demonstrated in Figures 8b,d, indicates that the development of the native tertiary structure near the heme pocket is concerted with the formation of HM.

Similar results were observed during the folding of ferrous cyt *c* in the presence of CO as shown in Figure 3b.[18] CO-bound unfolded ferrous cyt *c*, in which the sixth heme ligation site normally occupied by Met80 is blocked by the exogenous CO ligand, was studied to simplify the folding reaction by preventing off-pathway processes. A steady progression to the native (CO-bound) conformation was observed. It can be quantified by plotting the intensity of the line at 394 cm^{-1}, a mode identical to the 397 cm^{-1} mode in the ferric protein, as a function of time. The resulting kinetic trace is readily fit by an exponential function with a rate of 1600 s^{-1} (inset in Figure 3b). This indicates that, in the absence of any off-pathway reactions, folding follows first-order kinetics. A bonus of studying the CO-bound protein is that the solvent accessibility of the heme pocket during folding can also be monitored by following the Fe−CO stretching mode. This mode shifts from 498 cm^{-1} in the unfolded protein to ~493 cm^{-1} within the deadtime of our mixer due to the dilution of the denaturant. Additional changes of this mode are sensitive to the development of the heme environment and are discussed elsewhere.[18]

The folding of the cyt *c* polypeptide backbone in the absence of exogenous ligands is expected to be coupled with the heme coordination reaction as illustrated in Figure 10, unless one process is much faster than the other. In pathway A, Met80 heme coordination guides the development of the polypeptide conformation. This pathway is thermodynamically indistinguishable from the other extreme illustrated as pathway B, where the formation of the native polypeptide conformation guides the coordination of Met80 to the heme. The low-frequency resonance Raman data suggest that Met80 ligation is concerted with the formation of the native polypeptide conformation in ferric cyt *c*. The rate-limiting step in the folding of the ferric protein in the absence of off-pathway reactions is the conversion from HW, the obligatory intermediate, to the native HM state. The decay rate of ~500 s^{-1} from the HW to HM state therefore places an upper limit on the folding rate of the ferric protein. It is interesting that this rate is comparable to the folding rate of ~1600 s^{-1} for the CO-bound ferrous cyt *c*[18] and the imidazole-bound ferric cyt *c*.[31] This suggests that the folding of the bare polypeptide backbone is independent of the redox state of the heme and the identity of the sixth axial heme ligand.

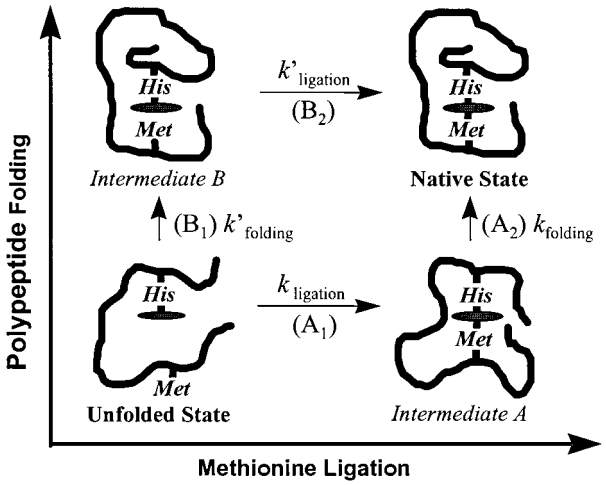


FIGURE 10. Two extreme folding pathways of cytochrome *c*. In pathway A, the native heme coordination by methionine-80 precedes the polypeptide backbone folding, whereas in pathway B, the polypeptide folding precedes the axial heme ligand coordination. Intermediates A and B are the hypothetical folding intermediates for pathways A and B, respectively. For ferric cyt *c*, the methionine to heme coordination and the folding of the polypeptide backbone into its native conformation are concerted.

Unfolding Reaction of Ferric Cytochrome *c*

To extract more information concerning the biphasic folding mechanism of cyt *c*, we also examined its *unfolding* reaction,[17] in which the native protein molecule diverges to form the many unfolded polypeptide chains with differing heme ligation states as formulated in eq 1.

$$\mathrm{HM} \rightarrow \mathrm{HW} + \mathrm{HH} + 5\mathrm{C} \qquad (1)$$

Heme ligand exchange reactions during unfolding of ferric cyt *c* were found to be strikingly similar to those identified during the folding reaction, although with a reverse sequence. Under acidic conditions (unfolding from pH 5.9 to pH 3.8), approximately 50% of the native His−Met coordinated form (HM) was converted to a His−water form (HW) within ~1 ms as shown in Figure 8e. The population of HW reaches a maximum at about 5 ms, at which point a five-coordinated form (5C) starts to be populated at the expense of HW. During unfolding to a final pH of 5.8 (Figure 8f), the decay of the native HM coordinated form is slower with a half-life of ~10 ms; in addition, a bis-His form (HH) instead of the 5C form is generated. A ligand exchange model, similar to that of the folding reaction, was proposed on the basis of these data as illustrated in Scheme 2. The solid lines in Figure 8e,f show the best fit curves. The rate constants are listed in Table 1.

It is noteworthy that the ligand exchange reaction does not reach completion until cyt *c* reaches its totally unfolded state as indicated by tryptophan fluorescence results shown in Figure 8e,f. This suggests that the ligand exchange reactions appear only when the polypeptide chains are still in relatively compact structures. The ligand exchange reactions are then followed by global relaxation of the polypeptide chain, a reverse sequence of the folding reaction. These results demonstrate that the ligand

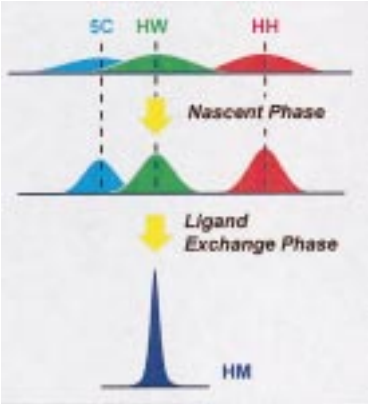
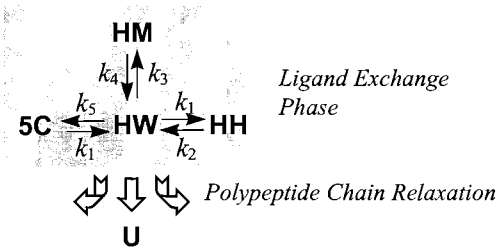


FIGURE 11. Schematic representation of the changes in the conformational space of the various heme coordination states during the folding of cytochrome *c*. Each heme coordination state reflects an ensemble of molecules with a common polypeptide backbone topology rather than a state with a fixed microscopic polypeptide conformation. In the unfolded state, the conformational space of each coordination form is large since the polypeptide chain is totally unstructured. During the nascent phase, the conformational space of each coordination form is dramatically reduced associated with the collapsing of the polypeptide chain. It is further reduced during the subsequent ligand exchange phase until the final native conformation, which has a well-defined architecture, is reached. HM, 5C, HW, and HH are defined in Figure 2.

Scheme 2



exchange phase observed during both the folding and unfolding reactions needs to take place while the protein is still in an entropically accessible conformational space (Figure 5).

Conclusions

Using resonance Raman scattering, we are able to map out protein conformation changes along the folding coordinate on the basis of the changes in the population of the various heme ligation states since all the potential heme ligands except the solvent water molecule are attached to the polypeptide backbone. It is important to point out that a specific heme ligation state reflects a group of molecules with a common polypeptide topology rather than a state with a fixed microscopic conformation. For example, HH is an ensemble of molecules in which the peptide segment holding His26/33 (yellow segment in Figure 1) is placed on the "wrong" side of the heme. The conformational space of each heme ligation state is gradually reduced as the protein molecule approaches the unique native conformation during the course of folding as illustrated in Figure 11.

Using resonance Raman scattering and our custom-designed 100 μs rapid solution mixer, we opened a new window to examine the microscopic details of the folding reaction in real time. It is concluded that the folding reaction of ferric cyt *c* includes two distinct phases, a kinetically controlled nascent phase and a thermodynamically controlled ligand exchange phase. This biphasic reaction guarantees that the protein folds into its unique native conformation with high efficiency and fidelity. The high efficiency is made possible by the kinetically controlled nascent phase, in which the conformational space is dramatically reduced by nonspecific condensation of the polypeptide chain without fine-tuning of the protein structure; the high fidelity is achieved through the thermodynamically controlled ligand exchange equilibrium, in which the energy is minimized through random structural fluctuations. The unfolding reaction of ferric cyt *c* is strikingly similar to that observed during the folding, although with a reverse sequence, which further confirms the validity and generality of the biphasic model. Both processes can be understood energetically by the free energy surfaces shown in Figure 5 in which they share similar ligand exchange phases, although for the folding reaction the global reduction of the molecular size appears prior to the ligand exchange phase and vice versa for the unfolding reaction.

This work was supported by National Institutes of Health Research Grants GM-54806 and GM-54812.

References

(1) Matthews, C. R. *Annu. Rev. Biochem.* **1993**, *62*, 653−683.
(2) Privalov, P. L. *Adv. Protein Chem.* **1979**, *33*, 167−241.
(3) Onuchic, J. N.; Wolynes, P. G.; Luthey-Schulten, Z.; Socci, N. D. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 3626−3630.
(4) Dill, K. A.; Chan, H. S. *Nature Struct. Biol.* **1997**, *4*, 10−19.
(5) Bushnell, G. W.; Louie, G. V.; Brayer, G. D. *J. Mol. Biol.* **1990**, *214*, 585−595.
(6) Scott, R. A.; Mauk, A. G. *Cytochrome c*: *A Multidisciplinary Approach*; University Science Books: Sausalito, CA, 1996.
(7) Nall, B. T. In *Cytochrome c: A Multidisciplinary Approach*; Scott, R. A., Mauk, A. G., Eds.; University Science Books: Sausalito, CA, 1996; pp 167−200.
(8) Takahashi, S.; Yeh, S.-R.; Das, T.; Chan, C.-K.; Gottfried, D. S.; Rousseau, D. L. *Nature Struct. Biol.* **1997**, *4*, 44−50.
(9) Yeh, S.-R.; Takahashi, S.; Fan, B.; Rousseau, D. L. *Nature Struct. Biol.* **1997**, *4*, 51−56.
(10) Yeh, S.-R.; Rousseau, D. L. *Nature Struct. Bio*l. **1998**, 5, 222−228.
(11) Evans, P. A.; Radford, S. E. *Curr. Opin. Struct. Biol.* **1994**, *4*, 100−106.
(12) Roder, H.; Elove, G. A.; Englander, W. *Nature* **1988**, *335*, 700−704.
(13) Johnson, W. C., Jr. *Annu. Rev. Biophys. Biophys. Chem.* **1988**, *17*, 145−166.
(14) Lattman, E. E. *Curr. Opin. Struct. Biol.* **1994**, *4*, 87−92.
(15) Elove, G. A.; Chaffotte, A. F.; Roder, H.; Goldberg, M. E. *Biochemistry* **1992**, *31*, 6876−6883.
(16) Elove, G. A.; Bhuyan, A. K.; Roder, H. *Biochemistry* **1994**, *33*, 6925−6935.

(17) Yeh, S.-R.; Rousseau, D. L. To be submitted.
(18) Han, S.; Rousseau, D. L. To be submitted.
(19) Roder, H.; Elove, G. A. In *Mechanisms of Protein Folding*; Pain, R. H., Ed.; Oxford University Press: New York, 1994; pp 26−54.
(20) Jones, C. M.; Henry, E. R.; Hu, Y.; Chan, C.-K.; Luck, S. D.; Bhuyan, A.; Roder, H.; Hofrichter, J.; Eaton, W. A. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 11860−11864.
(21) Pascher, T.; Chesick, J. P.; Winkler, J. R.; Gray, H. B. *Science* **1996**, *271*, 1558−1560.
(22) Mines, G. A.; Pascher, T.; Lee, S. C.; Winkler, J. R.; Gray, H. B. *Chem. Biol.* **1996**, *3*, 491−497.
(23) Creighton, T. E., Ed. In *Protein Folding*; W. H. Freeman & Co.: New York, 1992; pp 301−351.
(24) Sosnick, T. R.; Mayne, L.; Hiller, R.; Englander, S. W. *Nature Struct. Biol.* **1994**, *1*, 149−156.
(25) Pierce, M. M.; Nall, B. T. *Protein Sci.* **1997**, *6*, 618−627.
(26) Boffi, A.; Takahashi, S.; Spagnuolo, C.; Rousseau, D. L.; Chiancone, E. *J. Biol. Chem.* **1994**, *269*, 20437−20440.
(27) Hu, S.; Morris, I. K.; Singh, J. P.; Smith, K. M.; Spiro, T. G. *J. Am. Chem. Soc.* **1993**, *115*, 12446−12458.
(28) Jordan, T.; Eads, J. C.; Spiro, T. G. *Protein Sci.* **1995**, *4*, 716−728.
(29) Goto, Y.; Calciano, L. J.; Fink, A. L. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 573−577.
(30) Colon, W.; Wakem, L. P.; Sherman, F.; Roder, H. *Biochemistry* **1997**, *36*, 12535−12541.
(31) Chan, C.-K.; Hu, Y.; Takahashi, S.; Rousseau, D. L.; Eaton, W. A.; Hofrichter, J. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 1779−1784.
(32) Eaton, E. A.; Munoz, V.; Thompson, P. A.; Chan, C.-K.; Hofrichter, J. *Curr. Opin. Struct. Biol.* **1997**, *7*, 10−14.
(33) Roder, H.; Colon, W. *Curr. Opin. Struct. Biol.* **1997**, *7*, 15−28.
(34) Sosnick, T. R.; Shtilerman, M. D.; Mayne, L.; Englander, S. W. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 8545−8550.
(35) Shastry, M. C.; Roder, H. *Nature Struct. Biol.* **1998**, 5, 385−392.
(36) Colon, W.; Elove, G. A.; Wakem, L. P.; Sherman, F.; Roder, H. *Biochemistry* **1996**, *35*, 5538−5549.

AR970084P