Kinetics of Cytochrome *c* Folding Examined by Hydrogen Exchange and Mass Spectrometry†

Houjun Yang and David L. Smith*

Department of Chemistry, University of Nebraska−Lincoln, Lincoln, Nebraska 68588-0304

Received July 14, 1997; Revised Manuscript Received September 30, 1997⊗

ABSTRACT: Pulsed hydrogen exchange/mass spectrometry, a new method for studying protein folding, has been used to investigate folding of cytochrome *c* on the 5 ms to 15 s time scale. Cytochrome *c*, unfolded in guanidine hydrochloride/D_2O, was allowed to refold in a high-speed quenched-flow apparatus and pulse-labeled with protium to identify unfolded regions. Intact, labeled cytochrome *c* was digested into fragments which were analyzed by HPLC electrospray ionization mass spectrometry to determine the level of deuterium in each fragment. Bimodal distributions of deuterium were found for most segments, indicating that regions represented by these segments were either unfolded or completely folded in the intact polypeptide prior to labeling. This behavior is consistent with cooperative, localized folding which occurs in less than 10 ms in individual molecules. Deuterium levels found in the fragments were normalized to levels found in the same fragments derived from folded cytochrome *c*, pulse-labeled in the same manner, to indicate the percentage of cytochrome *c* that was folded. These results show that the N/C-terminal regions fold cooperatively on a time scale extending from less than the mixing time of the apparatus (5 ms) to as long as 15 s, and that the other regions also fold cooperatively. However, these regions do not begin to fold until 30 ms after mixing. In addition to providing new information on cytochrome *c* folding, these results demonstrate that pulse-hydrogen exchange/mass spectrometry is complementary to NMR in some respects and advantageous in others. Results of this study form the foundation required to extend the pulsed hydrogen exchange approach to folding studies of proteins too large to be analyzed by NMR.

Protein folding is of immense theoretical and practical importance. With the means of overexpressing proteins highly developed, there is an increasing need to control protein folding, which logically starts by identifying the paths and intermediates that are transiently populated as proteins fold. To meet this need, new analytical methods useful for identifying the steps through which diverse proteins fold will be required. Of particular importance will be analytical methods leading to identification of folding paths and intermediates in large proteins. Intramolecular hydrogen bonding, a characteristic feature of folded proteins but absent in unfolded proteins, dramatically decreases the rates at which hydrogens located at peptide amide linkages undergo isotope exchange. The sensitivity of hydrogen exchange to protein structure has led to using hydrogen exchange in studies of protein folding kinetics (*1*−*3*). Solutions of small proteins, unfolded and equilibrated in D_2O/denaturant, are rapidly diluted to initiate folding. After various folding times, the protein is transferred to rapid-exchange conditions so that unfolded, partially folded, and completely folded forms can be identified by their characteristic hydrogen exchange rates. The isotope labeling process is quenched by allowing the protein to fold to a state suitable for analysis by high-resolution nuclear magnetic resonance (NMR),[1] which is used to determine proton occupancy at peptide amide linkages. Isotope abundances at specific peptide linkages can be determined by this approach, where the abundance is the average of all molecules comprising the sample.

Results obtained in the present study demonstrate for the first time that the protein fragmentation/MS approach (*4*, *5*) is a useful alternative to pulsed spectroscopic and H/D NMR methods for identifying folding paths and intermediates. This approach, which is based on earlier methods that used tritium in place of deuterium (*6*, *7*), uses an acid protease to cleave labeled proteins into fragments whose deuterium levels are determined by directly-coupled HPLC MS. In the present study, the protein fragmentation/MS approach was used to identify cyt *c* folding paths and intermediates. Results of this study are similar to those obtained by NMR (*2*, *8*) for folding of the N/C-terminal regions of cyt *c*, but differ somewhat for folding of other regions of cyt *c*. Relative to NMR, the protein fragmentation approach is attractive because: (1) the structural heterogeneity of folding intermediates can be determined from the isotope patterns in the mass spectra of backbone fragments (*9*, *10*); (2) amide hydrogens along the entire backbone, including the most rapidly exchanging peptide amide hydrogens, can be used to sense protection (*11*); (3) much less material is required; and (4) folding of proteins much too large to be analyzed by high-resolution NMR can be examined (*10*, *12*, *13*).

†This work was supported by a grant from the National Institutes of Health (RO1 GM40384) and the Nebraska Center for Mass Spectrometry.

* Address correspondence to this author at the Department of Chemistry, University of Nebraska−Lincoln, Lincoln, NE 68588-0304. Telephone: 402 472 2794. FAX: 402 472 6892. Email: DLS@UNLINFO.UNL.EDU.

⊗ Abstract published in *Advance ACS Abstracts,* November 15, 1997.

[1] Abbreviations: CD, circular dichroism; cyt *c*, cytochrome c; ESI, electrospray ionization; GuHCl, guanidine hydrochloride; H/D, hydrogen/deuterium; HPLC, high-performance liquid chromatography; MS, mass spectrometry; *m/z*, mass-to-charge; N/C-terminal, amino-/carboxy-terminal; NMR, nuclear magnetic resonance.

S0006-2960(97)01718-2 CCC: $14.00

Results of this study facilitate a critical comparison of NMR and mass spectrometry for investigations of protein folding via hydrogen exchange and provide the foundation required for future use of hydrogen exchange/MS in protein folding studies.

EXPERIMENTAL PROCEDURES

Folding/Pulsed Hydrogen Exchange. A rapid mixing apparatus (BioLogic QFM-5) was used to dilute solutions of unfolded cyt *c*, label unfolded regions of cyt *c*, and quench isotope exchange. The general procedure has been described previously (*1*, *2*). Unfolded cyt *c* in which all exchangeable hydrogens had been replaced with deuterium was prepared by incubating the protein (1 mM) in D_2O (5 mM phosphate, pD 7, 4 M GuDCl, 20 °C) for at least 48 h. All pD measurements are given as read from the pH meter with no adjustment for isotope effects (*14*). The pH was decreased to 5.5 just prior to pulse-labeling experiments. For short folding times, deuterated and unfolded cyt *c* (syringe 1) was diluted 10-fold with H_2O (syringe 2, H_2O, 5 mM phosphate, pH 5.0). The pH of the folding solution was 5.0. Following folding times of 5 ms to 1 s, the pH was increased to 10.1 (syringe 3, H_2O, 0.1 M phosphate, pH 10.9) for 11 ms to label unfolded regions with protium. Isotope exchange was quenched by decreasing the pH to 2.5 (syringe 4, H_2O, 1 M HCl). The entire mixing apparatus was cooled to 8 °C. The flow from the rapid mixing system was collected in a dry ice/acetone bath and stored at −80 °C until analyzed. A similar procedure was used for folding times of 0.22 - 15 s, except folding was initiated by dilution with D_2O (syringe 2, D_2O, 5 mM phosphate, pH 5.0), and the labeling time was increased from 11 to 13 ms.

Isotope Analysis by HPLC ESIMS. Isotope analyses were initiated by rapidly thawing and digesting the labeled cyt *c* with pepsin (substrate:enzyme, 1:1, w/w) at 0 °C for 5 min. The digest was analyzed immediately by HPLC ESIMS, as described elsewhere (*15*). Deuterium levels found in peptic fragments derived from cyt *c*, either nonexchanged or completely exchanged, were used to adjust for small losses of deuterium from peptide amide linkages that occurred during analysis (*4*). Completely exchanged cyt *c* was prepared by denaturing the protein in D_2O, refolding, and decreasing the pH to 2.5 to quench H/D exchange. The deuterated and nondeuterated standards were digested and analyzed by HPLC ESIMS using the same conditions as were used to analyze samples labeled by the pulsed hydrogen exchange procedures. Another control was used as a reference to indicate the deuterium level representative of folded cyt *c* labeled under the pulsed hydrogen exchange conditions described above. Cyt *c* was unfolded and completely exchanged in D_2O (4 M GuDCl), and then allowed to refold by incubating for 1 h under the folding conditions described above for pulsed hydrogen exchange (0.4 M GuDCl, 5 mM phosphate, pD 5.0). Deuterium levels found in each segment following pulsed labeling, quenching, and isotope analysis were used to indicate deuterium levels representing folded cyt *c*.

RESULTS

Intermolecular Distribution of Deuterium Directly Implicates Localized Cooperative Folding. Successful completion of these experiments required careful selection and control of pH at each step. Pulse-labeling experiments were initiated by denaturing oxidized cyt *c* in 4 M GuDCl/D_2O, diluting the solution 10-fold to initiate folding under conditions where amide H/D exchange was relatively slow (pH 5.0, 8 °C), and allowing folding to proceed for 5 ms to 15 s. During the subsequent labeling step, the pH was increased to 10 to facilitate complete replacement of deuterium located at peptide amide linkages in unfolded regions of cyt *c* with protium within the labeling time of 11−13 ms. Labeled cyt *c* was digested with pepsin and analyzed by HPLC MS. During these steps, H/D exchange at peptide amide linkages was quenched by maintaining the pH and temperature at 2.5 and 0 °C, respectively. Under these quench conditions, the half-life for exchange at most peptide linkages was greater than 60 min (*16*), which was far greater than the total time required for proteolytic fragmentation of cyt *c* and HPLC analysis (approximately 10 min). Deuterium levels in short segments of the cyt *c* backbone were determined from the molecular weights of the corresponding peptic fragments. Since the HPLC step was performed in H_2O, rapidly exchanging deuterium located in the side chains was replaced with protium before the fragments reached the mass spectrometer (*16*). Hence, the deuterium levels measured in these experiments refer to deuterium located primarily at peptide amide linkages. The proteolytic conditions used in this study typically gave approximately 21 fragments from which folding along nearly all of the cyt *c* backbone could be sensed.

The ESI mass spectra of the peptic fragment including amide linkages 1−21 of cyt *c* allowed to fold for 5, 31, and 506 ms are given in Figure 1 b−d. Results for the same fragment representing unfolded and folded cyt *c* are presented in Figure 1a,e, respectively. Only results for the region including the +3 charge state of this fragment (*m/z* 750−758) are presented. Although the mass spectrometer resolution used for these measurements was not adequate to separate individual isotope peaks for many ions, as in Figure 1, it was adequate for determining the intermolecular distribution of heavy isotopes in the peptides. Thus, the isotope pattern of the N-terminal peptic fragment of cyt *c*, illustrated in Figure 1, is expected to be an accurate representation of the intermolecular distribution of deuterium in this fragment. Isotope patterns are an important source of information because they can be used to identify and quantify various structural forms of cyt *c* that are populated during folding. However, this use may be limited by incomplete resolution of isotope patterns representing the variously folded forms of cyt *c*. Several factors contribute to the intermolecular distribution of deuterium found in these proteolytic fragments (*10*). Amide linkages with very large or small H/D exchange rates will be either completely protiated or completely deuterated. Although these amide linkages affect the average deuterium level found in a peptide, they contribute little to the peak broadening. In contrast, amide linkages that undergo H/D exchange on the time scale of the experiment are only partially deuterated and make major contributions to peak broadening. For the labeling conditions used in this study, H/D exchange was complete in unfolded regions, but spanned a wide range in folded regions. Artifactual H/D exchange that occurred during analysis also contributed to peak broadening. For the types of measurements described here, broadening of mass spectral peaks due to the random distribution of heavy

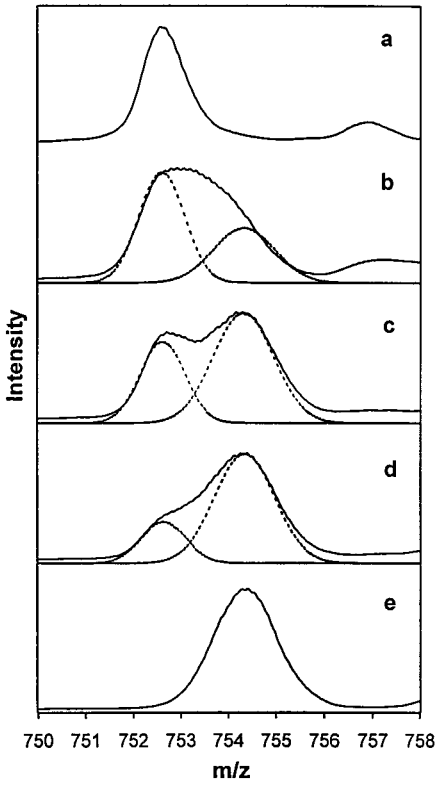

FIGURE 1: ESI mass spectra of the N-terminal segment of cyt *c*, including amide linkages 1−21, following folding for 5, 31, and 506 ms are presented in panels b−d. The areas of the low- and high-mass envelopes of isotope peaks, indicated by dashed lines, represent the fraction of the cyt *c* population in which the N-terminus was unfolded or folded, respectively. Panels a and e indicate isotope patterns of this segment representing unfolded and folded cyt *c*, respectively. The average *m/z* found for this fragment (MH_3^{+3}) derived from unfolded (panel a) or folded (panel e) cyt *c* was 752.5 and 754.3, respectively. Curve-fitting the bimodal isotope patterns in panels b−d with reference isotope patterns for unfolded and folded forms, panels a and e, indicates the extent of folding that occurred after various folding times.

isotopes of carbon, nitrogen, oxygen, and sulfur is relatively unimportant. Likewise, the resolution of the mass spectrometer did not contribute significantly to the broadening of the isotope envelopes. Despite these factors, which may limit detecting cyt *c* forms that differ little in H/D exchange rates, isotope patterns of peptides detected in this study could often be used to identify and quantify folded and unfolded regions of the cyt *c* backbone.

For structurally homogeneous samples, H/D exchange within any segment is the same for all molecules comprising the sample, yielding a random distribution of deuterium and a single envelope of isotope peaks. Structurally heterogeneous samples in which a particular segment is folded in some molecules and unfolded in others yield bimodal isotope patterns, consisting of two envelopes of isotope peaks (*9*, *10*). The isotope patterns presented in Figure 1 for the N-terminal fragment of cyt *c* illustrate how they can be used to detect structural heterogeneity. The peak widths of this fragment representing unfolded and folded reference samples of cyt *c* (Figure 1a,e) are much narrower than those of the same fragment when derived from partially folded cyt *c* (Figure 1b−d). The isotope pattern found for this fragment, when analyzed before cyt *c* folding was complete, suggests a bimodal isotope distribution representing two different structural forms of the N-terminus of cyt *c*, each with a different and characteristic hydrogen exchange rate.

Resolution of the two distributions (dashed lines in Figure 1) comprising the bimodal isotope pattern was enhanced by fitting the reference isotope distributions representing unfolded and folded cyt *c* (Figure 1a,e) to the isotope patterns of the partially folded cyt *c* (Figure 1b−d). Fitting was performed by fixing the centroids and widths of the reference peaks, and varying only their areas. The low-mass envelope of isotope peaks represents cyt *c* that retained little or no deuterium, while the high-mass envelope represents cyt *c* with a deuterium level characteristic of the completely folded form. Since the labeling step was designed to replace all of the deuterium in unfolded regions with protium, the area of the low-mass envelope indicates the fraction of cyt *c* that was unfolded at the end of the various folding times. Correspondingly, the area of the high-mass envelope is a direct measure of the amount of cyt *c* that was folded in the N-terminus at the end of the folding time.

Residuals from curve-fitting were usually small, indicating that the N-terminal segment of most of the cyt *c* population is either unfolded or folded, as assessed by amide hydrogen exchange. The largest residual was found for the shortest folding time (Figure 1b), which may be attributed to the occurrence of significant folding during the labeling time. Finding little evidence for peptides with deuterium levels intermediate between those representative of the unfolded and folded forms suggests that the time required for this region within any molecule to fold is much shorter than the labeling time (approximately 10 ms). Although seconds are required for the N-terminus of the entire cyt *c* population to fold (see below), less than 10 ms is required for this region to fold in individual molecules of cyt *c*, demonstrating that cooperative folding is the dominant process in the N-terminus of cyt *c*.

A similar strategy was used to determine whether cooperative folding dominates in other regions of the cyt *c* backbone. Peaks in the mass spectra of all 21 peptic fragments were broader than expected for a random intermolecular distribution of deuterium throughout the entire cyt *c* population. Isotope distributions of many fragments were clearly bimodal (e.g., linkages 1−21, Figure 1), and could be resolved into their constituent envelopes representative of cyt *c* that was either unfolded or folded in the region from which the peptides were derived. As for the segment including amide linkages 1−21, these results indicate that short segments of cyt *c* backbone fold cooperatively in less than the labeling time (approximately 10 ms). For segments whose isotope distributions representing unfolded and folded forms were poorly resolved, the possibility for noncooperative folding on the millisecond time scale could not be eliminated.

The areas of the two envelopes of isotope peaks indicate the relative concentrations of the unfolded and folded forms of the segments present at the end of the folding time. The fraction of the cyt *c* population that remained unfolded in a particular region of the backbone decreased with increasing folding time, as illustrated in Figure 1 for the N-terminus. Plots of the natural logarithm of the fraction of cyt *c* population that was unfolded in the N-terminus versus folding time for most peptic fragments were linear through the 5−120 ms range of folding times. Rate constants describing folding in the N/C-terminal regions of cyt *c* (Table 1) determined from the slopes of these plots show that the N/C-terminal regions of 40−50% of cyt *c* folded at similar rates (10−30 s^{-1}). Furthermore, the *y*-intercepts of these

Table 1: Rate Constants and Subpopulations for Folding of the N/C-Terminal Regions of Cyt *c* Determined from the Areas of the Low-Mass Envelope of Isotope Peaks, As Illustrated in Figure 1

segment[a]	% folded[b]	k (s^{-1})
1−10	50	30
1−21	40	10
82−96	41	10
96−104	51	10

[a] Amide linkages of the cyt *c* backbone. [b] Population of cyt *c* that is folded in a particular region.

plots indicate that the N/C-terminal regions in 40−50% of cyt *c* folded in less than 5 ms.

Rates of Localized Folding Determined from Deuterium Levels in Peptic Fragments. The approach described above for determining folding rate constants used hydrogen exchange to label unfolded and folded regions, and the areas of the ensuing isotope envelopes to quantify the fraction of cyt *c* that was unfolded in specific regions of the protein. This approach is arguably the most direct of all methods for investigating protein folding. However, it is useful only for segments in which the isotope envelopes are resolved. When considerable hydrogen exchange occurred in the folded form (i.e., H replaced D), the low- and high-mass envelopes were not resolved, and the relative concentrations of unfolded and folded forms of cyt *c* could not be determined accurately. An alternate approach for determining folding rate constants from the same mass spectra is based on the deuterium levels found in the peptic fragments. The average molecular weight of a peptic fragment, determined from the centroid of all of its isotope peaks, gives the deuterium level in this segment. Finding no deuterium in a segment derived from cyt *c* indicates that this segment was unfolded at the end of the folding time; finding a deuterium level equal to that found in the same segment when derived from folded cyt *c* indicates that this segment was folded. The standard for the deuterium level representative of folded cyt *c* was determined for each segment from its molecular weight when the segment was derived from folded cyt *c* labeled under the normal pulsed hydrogen exchange conditions (label time 11−13 ms, pH 10). For example, 4.3 deuteriums were found in the N-terminal segment comprising amide linkages 1−10. Deuterium levels between 0 and 4.3 in this fragment indicate the fraction of molecules that had folded prior to labeling.

Comparing the deuterium levels found in peptic fragments at a particular folding time with those found in the same fragments derived from reference folded and unfolded cyt *c* provides a basis for expressing these deuterium levels as the percentage of cyt *c* that is folded at this time in a particular region. Results for 16 segments representing nearly all of the backbone of cyt *c* are presented in Figure 2. Folding kinetics for the N/C-terminal regions of cyt *c* are illustrated by deuterium levels found in the four segments including amide linkages 1−10, 11−21, 95−96, and 96−104 (Figure 2a). Although peptides specific for segments 11−21 and 95−96 were not detected, folding in these regions was determined from differences in deuterium levels found in overlapping peptides (*4*). For example, folding of amide linkages 11−21 was determined from the difference in deuterium levels found in segments 1−10 and 1−21. These results show that the N/C-terminal regions of cyt *c* have similar folding kinetics, and that these regions in approximately 50% of cyt *c* are folded within 5 ms. It is also evident that the N/C-terminal regions in a subpopulation of cyt *c* require many seconds to fold.

Folding of much of the cyt *c* backbone joining the N/C-terminal regions, indicated by deuterium levels found in four peptides including amide linkages 23−32, 33−36, 38−46, and 49−64, is illustrated in Figure 2b. Deuterium levels in these four peptic fragments indicate that much of this part of the cyt *c* backbone has similar folding kinetics, which differ markedly from the folding kinetics of the N/C-terminal regions. Folding in this region did not start until 20−30 ms after dilution, while the N/C-terminal regions in approximately half of cyt *c* folded within 5 ms. Like the N/C-terminal regions, the interior region of part of the cyt *c* population required several seconds to fold. Results for eight overlapping peptides derived from two regions including linkages 66−82 (Figure 2c) and linkages 81−96 (Figure 2d) indicate the internal consistency typical of these measurements.

Modeling the folding results presented in Figure 2 using first-order kinetics has been used to compare folding kinetics of different regions of the cyt *c* backbone (Figure 3). For example, the eight peptic fragments including amide linkages 1−21 and 81−104 of the N/C-terminal regions displayed similar folding kinetics. Data for these fragments, when combined and fitted to an equation of three exponential terms, were used to draw the solid line in Figure 3. Likewise, the folding kinetics of the cyt *c* backbone regions represented by the four fragments including amide linkages 23−64 were similar. Results for these four fragments, when combined and fitted to an equation of two exponential terms, were used to draw the dotted line in Figure 3. This presentation illustrates that the folding kinetics for the N/C-terminal regions are different from those of the interior of cyt *c* backbone. Similar presentation of results for amide linkages 66−82, represented by the dashed line in Figure 3, suggests folding kinetics intermediate between those of regions located on either side.

Modeling of the folding results summarized in Figure 3 also demonstrates that each of these regions may fold at different rates in different subpopulations of cyt *c*. Subpopulations and their folding rate constants are given in Table 2. This analysis indicates that the N/C-terminal regions in 49% of the cyt *c* population fold within the dead time of the experiment ($k > 200$ s^{-1}), 37% fold with a rate constant of 16 s^{-1}, and 14% fold with a rate constant of 0.28 s^{-1}. In contrast to these regions, folding of the four peptic fragments derived from the region including amide linkages 23−64 is adequately fitted using two exponential terms. The compartment folding most rapidly for this region includes 70% of the cyt *c* population, which folds with a rate constant of 9 s^{-1}. Folding of the remaining 30% of cyt *c* in this region is much slower ($k = 0.07$ s^{-1}). It is significant that localized folding kinetics based on the deuterium levels found for the N/C-terminal regions (Table 2) agree with those based on the intermolecular distribution of deuterium found for fragments derived from these regions (Table 1).

DISCUSSION

Much of our understanding of protein folding is based on structural studies of cyt *c* as it is rapidly transferred from mildly denaturing conditions to conditions favoring the folded state. Under mildly denaturing conditions (e.g., 4 M

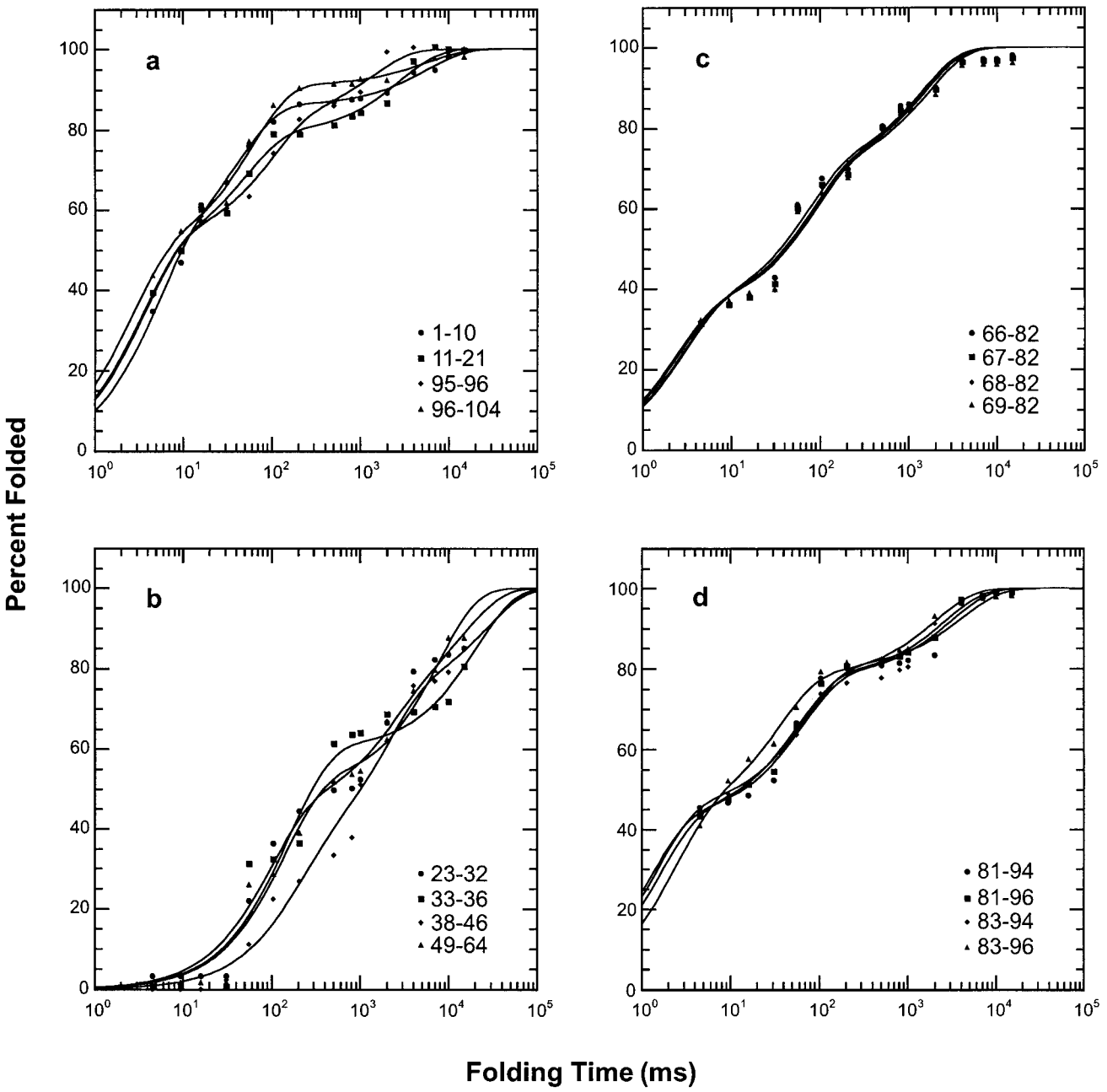

FIGURE 2: Extent of localized folding, as assessed by amide hydrogen exchange, in 16 segments of the cyt *c* backbone vs folding time. Segments displaying similar folding kinetics are grouped in panels a−d.

GuHCl), cyt *c* exists as a nearly random coil in which Cys 14/17 are covalently bound to the heme (*8*). Because of its close proximity to Cys 14/17, His 18 is probably ligated to the heme iron (*17*). Electron-rich ligands, such as His 26, His 33, Met 80, and water, may be transiently ligated to the remaining iron binding site. The goal of most folding studies is to describe events linking the near random coil of the unfolded state to the highly compact, folded state. Folding rates for cyt *c* regions that contain chromophores have been determined using stopped-flow techniques where the extent of folding was monitored by tryptophan fluorescence (*8*, *18*, *19*), CD (*18*, *20*), and resonance Raman scattering (*19*, *21*, *22*) spectroscopies. Upon rapid dilution of the solution, the nearly random coil of cyt *c* contracts to a more compact form within 0.1 ms (*19*). This compact form may have His 26/33, Met 80, or water bound to the heme. If Met 80 is bound, folding to native-like states may proceed within a few milliseconds. However, if His 26/33 is bound, it must be replaced with water before folding to the native conformation can continue. Likewise, isomerization of non-native cis forms of Pro 71/76 to native trans forms must occur before folding is complete. These processes, as well as their combinations, appear as kinetic traps that may delay folding for seconds (*8*, *22*, *23*).

The rate at which hydrogen bonds form has been used to examine folding events that do not directly involve cyt *c* chromophores. The extent of folding determined by hydrogen exchange/NMR is expressed as proton occupancy levels, which indicate the extent to which the folded structure decreases the rate of hydrogen exchange. Since the level of protection against H/D exchange is averaged over the entire sample, a small number of completely folded molecules may give the same result as a large number of partially folded molecules. In contrast, hydrogen exchange/MS gives the distribution of deuterium among all molecules, from which the structural heterogeneity of the backbone can be detected,

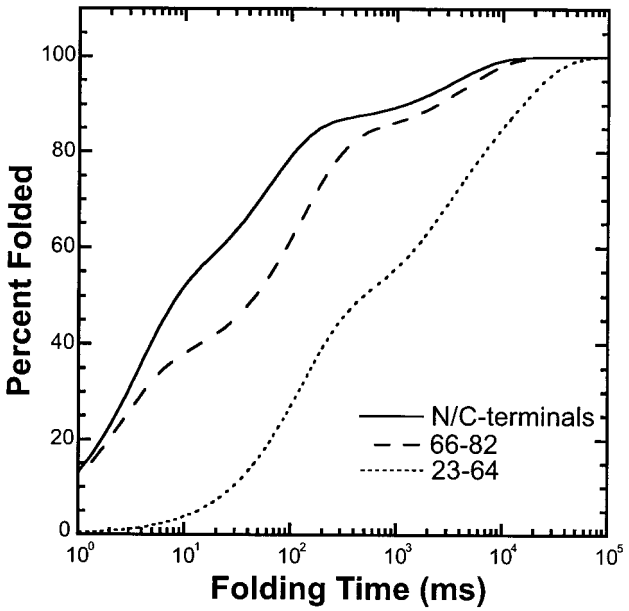

FIGURE 3: Comparison of folding kinetics in different regions of cyt *c*. The lines were fitted to data for all segments derived from the three regions exhibiting similar folding behavior.

Table 2: Rate Constants and Subpopulations for Localized Folding of Cyt *c* Determined by Fitting Data Presented in Figure 3 to Three Exponential Terms[a]

segment[b]	% folded[c]	k (s^{-1})	% folded[c]	k (s^{-1})	% folded[c]	k (s^{-1})
1−21 and 81−104	49	>200	37	16	14	0.28
23−64	−	−	70	9	30	0.07
66−82	35	>200	48	8	17	0.22

[a] % folded = $100 - \{A_1 \exp(-k_1 t) + A_2 \exp(-k_2 t) + A_3 \exp(-k_3 t)\}$. [b] Amide linkages of the cyt *c* backbone. [c] Population of cyt *c* that is folded in a particular region.

quantified, and structurally characterized. In the present study, the rates at which small regions of cyt *c* folded following rapid dilution of the denaturant were indicated by the decreased hydrogen exchange rates of peptic fragments derived from these regions. Mass spectral isotope patterns for most of the peptic fragments were bimodal, indicating the presence of two structural forms, one unfolded and one folded (*9*, *10*). The areas of the envelopes of isotope peaks comprising the bimodal isotope patterns are direct measures of the relative abundances of the unfolded and folded forms.

Modeling of the rate at which cyt *c* became folded in a particular region was used to determine the number and size of cyt *c* folding subpopulations (Table 2). Three subpopulations were required to fit folding of the N/C-terminal regions of cyt *c* (Figure 3). Finding approximately 50% of the N/C-termini folded within the dead time of these experiments (5 ms) is consistent with approximately half of the cyt *c* population collapsing with Met 80 joined to the heme. Modeling shows that up to 10 s is required for the N/C-terminal regions of all cyt *c* molecules to fold. Finding similar folding rate constants and subpopulation sizes for the eight N/C-terminal segments sensed in this study suggests that both termini undergo cooperative folding to form a folded unit.

Only two subpopulations were required to fit the folding data of all peptic fragments derived from the cyt *c* backbone regions including linkages 23−64, of cyt *c* (Figure 3). Unlike the N/C-terminal regions of approximately 50% of the cyt *c* population, this region does not begin to fold for 20−30 ms, as assessed by amide hydrogen exchange. The fraction of cyt *c* that was folded in the N/C-terminal regions was always greater than that of the backbone interior, which may indicate that folding of the N/C-terminal regions precedes folding of the interior. Likewise, the four peptic fragments derived from the interior of the backbone displayed similar folding kinetics, subpopulation sizes, and rate constants, suggesting that the entire region folds cooperatively. These results imply that cyt *c* folds in two steps, with the N/C-terminal regions folding first, followed by folding of the backbone joining the N/C-terminal regions.

The apparent folding behavior of segments including linkages 66−82 is intermediate between those of the N/C-terminal regions and the interior regions. Whether this region forms a separate folding domain, or the transition between two folding domains, is not evident from the present results. The accuracy with which the sizes and locations of unfolding units is determined depends primarily on the sites at which pepsin cleaves cyt *c*. These cleavage sites are not expected to coincide with transitions between folding units, so these transition regions likely fall within specific peptic fragments. The relative sizes of the folding domains found in this study and the peptic fragments (2−15 residue) indicates that the spatial resolution is sufficient to determine the sizes and locations of folding domains, but generally not sufficient to locate specific residues marking the transitions between folding domains.

Experimental conditions for folding and labeling cyt *c* were chosen to simulate those used by Elöve et al. (*8*), so results obtained by NMR and MS could be compared. However, the pH of the solution used to equilibrate unfolded cyt *c* just before dilution to initiate folding in the present study was 5.5, but 5.0 in the previous study. The extent of folding of the N/C-terminal regions, as assessed by hydrogen exchange using NMR and MS to follow hydrogen exchange, is nearly identical (Figure 4a). Results from both approaches indicate approximately 50% folding in less than 5 ms, while as long as 10 s is required for folding of the final 10−20%. Folding of amide linkages 23−64 of the cyt *c* backbone, sensed by NMR and MS, is illustrated in Figure 4b. Although results for both methods indicate that folding requires as long as 10 s for completion, the extent of folding that occurs during the shortest period (t < 10 ms) appears different. NMR analysis indicates that 30−40% of the interior region of the cyt *c* backbone folds in less than 10 ms, while analysis by MS indicates that this region begins to fold only after approximately 20−30 ms. The cause for this difference is not evident, but may be related to what is actually measured by the two methods. The NMR results are based on measurements at several specific peptide amide linkages, all of which have similar kinetics for protection from hydrogen exchange. The MS results are based on the deuterium levels found in short segments of the backbone. Isotope measurements by NMR and MS also differ in the way hydrogen exchange was quenched after the labeling step. In the NMR studies, hydrogen exchange was quenched by decreasing the pH to 5 where folding was completed. Isotope analysis by NMR is limited to those amide hydrogens whose half-lives for exchange in the folded state are greater than the several hours usually required for analysis by NMR. In the MS studies, hydrogen exchange was quenched by decreasing the

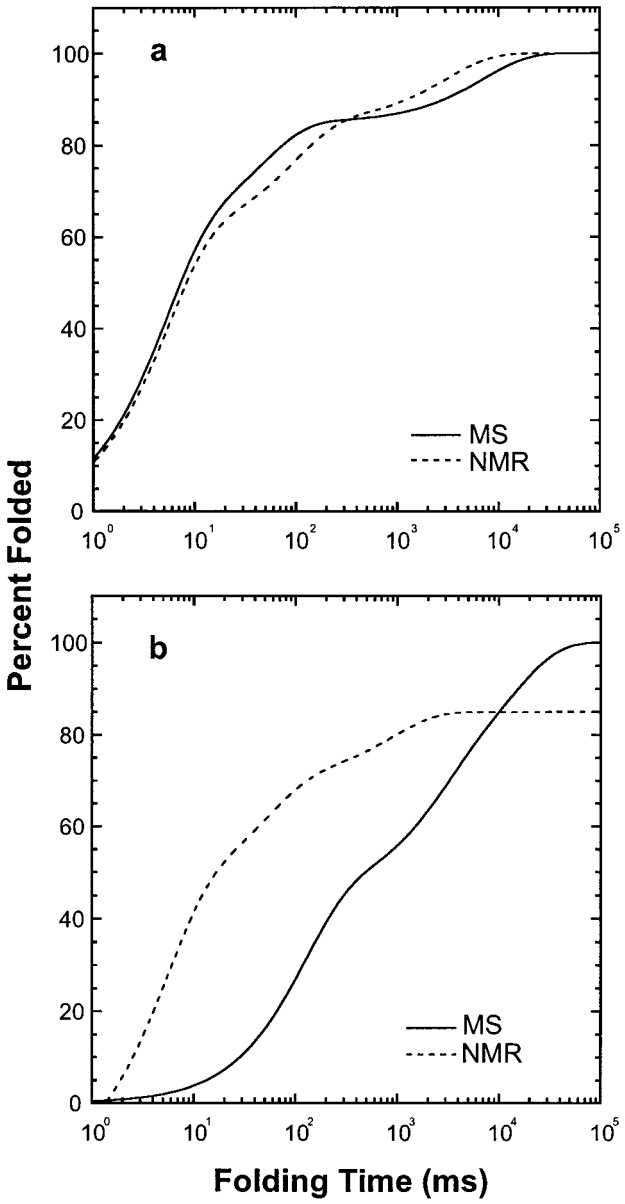

FIGURE 4: Folding of N/C-terminal regions, panel a, and interior backbone regions, panel b, of cyt *c* at pH 5 when hydrogen exchange was determined by NMR and mass spectrometry. This presentation of the NMR data was adapted from Elöve et al. (*8*).

pH to 2.5 and 0 °C following labeling. Under these conditions, the half-life for hydrogen exchange at most peptide amide linkages is greater than 60 min, a time that is much longer than the digestion/analysis time (approximately 10 min). As a result, amide hydrogen exchange rates can be measured for even the most rapidly exchanging amide hydrogens by combining acid quench and mass spectrometry (*11*). Although similar isotope exchange rates were found for model peptides when analyzed by both NMR and MS (*24*), results for hydrogen exchange in proteins may differ because NMR and MS may be sensing exchange at different amide linkages. The excellent correlation between the NMR and MS results for folding of the N/C-terminal regions, and similar results for folding of the internal regions, demonstrates that either method can be used effectively to examine protein folding through H/D exchange.

Recent advances in experiment and theory have led to a "New View" of protein folding kinetics (*25, 26*). This new view emphasizes subpopulations of unfolded protein following a multitude of paths along the free energy folding surface. Experiments designed to identify subpopulations defined by intramolecular correlated changes in structure were at the top of the "Wish list for experimentalists" described by Dill and Chan (*25*). In the present study, the folding rates of various segments of the cyt *c* backbone required two or three exponential terms, indicating as many subpopulations. In addition, the well-defined bimodal isotope patterns found for several segments representing each of the subpopulations show that folding within these segments is cooperative. That is, all amide hydrogens within these segments achieve protection from H/D exchange within the labeling time (approximately 10 ms). As used here, "cooperative change" refers to simultaneous H/D exchange of all peptide amide hydrogens within a defined backbone region (peptic fragment) and not simultaneous H/D exchange in the entire cyt *c* backbone. The present results are consistent with a free energy folding surface with barriers capable of trapping partially folded subpopulations. Some subpopulations are trapped for as long as 10 s. However, when the members of the subpopulations escape, less than 10 ms is required for the ensuing folding step. Other subpopulations are not trapped and fold to the native state in less than 10 ms.

As the emphasis of protein folding studies shifts to large, multidomain proteins, new methods for following folding kinetics will be required. Cyt *c* has been a good model for folding studies because, with heme and only one tryptophan, it is especially suited to high-speed spectroscopic measurements. Fluorescence measurements of larger proteins with multiple tryptophan residues will be less useful because signals from specific tryptophan residues may be difficut to identify. Stopped-flow CD will be useful for detecting secondary and tertiary structural features in large proteins, but is unlikely to lead to identification of cooperative folding units or the order in which these units fold. As for small proteins, amide hydrogen exchange will be useful for studies of folding in large proteins because it provides a mechanism for sensing folding of the entire backbone of a protein. While the spatial resolution offered by MS is less than that of NMR, it will be adequate to identify cooperative folding units in large proteins. For example, regions undergoing thermal-induced unfolding in aldoalse, M_r 157 000, have been identified by hydrogen exchange/MS (*10*). Localized hydrogen exchange measured by mass spectrometry opens the possibility for new and highly detailed folding studies of large proteins

ACKNOWLEDGMENT

We are pleased to acknowledge helpful discussions with S. W. Englander.

REFERENCES

1. Udgaonkar, J. B., and Baldwin, R. L. (1988) *Nature (London) 335*, 694−699.
2. Roder, H., Elöve, G. A., and Englander, S. W. (1988) *Nature (London) 335*, 700−704.
3. Englander, S. W., and Mayne, L. (1992) *Annu. Rev. Biophys. Biomol. Struct. 21*, 243−265.
4. Zhang, Z., and Smith, D. L. (1993) *Protein Sci. 2*, 522−531.
5. Smith, D. L., Deng, Y., and Zhang, Z. (1997) *J. Mass Spectrom. 32*, 135−146.
6. Rosa, J. J., and Richards, F. M. (1979) *J. Mol. Biol. 133*, 399−416.

7. Englander, J. J., Rogero, J. R., and Englander, S. W. (1985) *Anal. Biochem. 147*, 234−244.
8. Elöve, G. A., Bhuyan, A. K., and Roder, H. (1994) *Biochemistry 33*, 6925−6935.
9. Miranker, A., Robinson, C. V., Radford, S. E., Aplin, R. T., and Dobson, C. M. (1993) *Science 262*, 896−900.
10. Zhang, Z., Post, C. B., and Smith, D. L. (1996) *Biochemistry 35*, 779−791.
11. Dharmasiri, K., and Smith, D. L. (1996) *Anal. Chem. 68*, 2340−2344.
12. Liu, Y., and Smith, D. L. (1994) *J. Am. Soc. Mass Spectrom. 5*, 19−28.
13. Zhang, Z., and Smith, D. L. (1996) *Protein Sci. 5*, 1282−1289.
14. Connelly, G. P., Bai, Y., Jeng, M.-F., and Englander, S. W. (1993) *Proteins: Struct., Funct., Genet. 17*, 87−92.
15. Engen, J. R., Smithgall, T. E., Gmeiner, W. H., and Smith, D. L. (1997) *Biochemistry 36*, 14384−14391.
16. Bai, Y., Milne, J. S., Mayne, L., and Englander, S. W. (1993) *Proteins: Struct., Funct., Genet. 17*, 75−86.
17. Muthurkrishnan, K., and Nall, B. T. (1991) *Biochemistry 30*, 4706−4710.
18. Sosnick, T. R., Mayne, L., Hiller, R., and Englander, S. W. (1994) *Nat. Struct. Biol. 1*, 149−156.
19. Chan, C.-K., Hu, Y., Takahashi, S., Rousseau, D. L., Eaton, W. A., and Hofrichter, J. (1997) *Proc. Natl. Acad. Sci. U.S.A. 94*, 1779−1787.
20. Elöve, G. A., Chaffotte, A. F., Roder, H., and Goldberg, M. E. (1992) *Biochemistry 31*, 6876−6883.
21. Takahashi, S., Yeh, S.-R., Das, T. K., Chan, C.-K., Gottried, D. S., and Rousseau, D. L. (1997) *Nat. Struct. Biol. 4*, 44−50.
22. Yeh, S.-R., Takahashi, S., Fan, B., and Rousseau, D. L. (1997) *Nat. Struct. Biol. 4*, 51−56.
23. Ridge, J. A., Baldwin, R. L., and Labhardt, A. M. (1981) *Biochemistry 20*, 1622−1630.
24. Thévenon-Emeric, G., Kozlowski, J., Zhang, Z., and Smith, D. L. (1992) *Anal. Chem. 64*, 2456−2458.
25. Baldwin, R. L. (1994) *Nature 369*, 183−184.
26. Dill, K. A., and Chan, H. S. (1997) *Nat. Struct. Biol. 4*, 10−19.

BI9717183