Transient Dimer in the Refolding Kinetics of Cytochrome *c* Characterized by Small-Angle X-ray Scattering[†]

Daniel J. Segel,[‡] David Eliezer,[§] Vladimir Uversky,[||,⊥] Anthony L. Fink,[||] Keith O. Hodgson,[#] and Sebastian Doniach*[,‡]

Departments of Physics and Chemistry, Stanford University, Stanford, California 94305, Institute for Biological Instrumentation, Russian Academy of Sciences, 142292 Pushchino, Moscow Region, Russia, Department of Chemistry and Biochemistry, University of California, Santa Cruz, California 95064, Stanford Synchrotron Radiation Laboratory, Stanford, California 94309, and Department of Biochemistry, Weill Medical College of Cornell University, 1300 New York Avenue, New York, New York 10021

Received June 10, 1999; Revised Manuscript Received September 10, 1999

ABSTRACT: The equilibrium unfolding and the kinetic refolding of cytochrome *c* (Cyt *c*) in the presence of imidazole were studied with small-angle X-ray scattering (SAXS). The equilibrium unfolding experiments showed the radius of gyration, R_g, of native Cyt *c* to swell ∼1 Å with the addition of imidazole. The thermodynamic parameter *m* also reflects an expansion of the protein as its lower value demonstrates an increase in solvent-accessible surface area over that of native Cyt *c* in the absence of imidazole. Refolding was studied in the presence of imidazole as it prevents misligated intermediate states from forming during the refolding process, simplifying the kinetics, and making them easier to resolve. Time-resolved decreases in the forward scattering amplitude, $I(0)$, demonstrated the transient formation of an aggregated intermediate. Final protein and denaturant concentrations were varied in the refolding kinetics, and the singular value decomposition (SVD) method was employed to characterize the associated state. This state was determined to be a dimer, with properties consistent with a molten globule.

A strong propensity to associate is a general characteristic of non-native proteins and accounts for the fact that protein refolding is often accompanied by aggregation. Normally this has been viewed as a side-effect of working at high protein concentrations, which is the necessary prerequisite to obtain data with good signal-to-noise ratio. Little emphasis has been placed on the study of these associated states formed during protein refolding. However, this situation has changed considerably lately, mainly due to the fact that protein aggregation is believed to be the cause of a number of debilitating diseases including the prion diseases, Alzheimer's and Parkinson's diseases, cataracts, and many others. Understanding the stability and structure of protein aggregates therefore has become an important research goal. Small-angle X-ray scattering (SAXS)[1] potentially represents a powerful tool in the study of protein aggregates. This approach allows characterization of the overall size and shape of a soluble aggregate, as well as determination of the order (dimer, trimer, etc.). Time-resolved SAXS can be used for characterization of transient aggregates, populated during refolding of proteins from the completely unfolded state (*1*). In this study, we focus on characterizing the transient dimers formed during the refolding of cytochrome *c* (Cyt *c*).

Cyt *c* is a small protein (104 residues) with a heme group covalently bound to residues Cys14 and Cys17 (Figure 1). The heme iron has two axial ligands: His18 and Met80 (*2*). Under typical denaturing conditions, such as high guanidine hydrochloride (GdnHCl) or urea concentrations at neutral pH, Met80 dissociates from the heme (*3*). The lack of conformational freedom of His18, which has two neighboring residues (Cys14, Cys17) covalently bound to the heme, prevents it from detaching from the heme iron even under unfolding conditions (*3*−*5*). During refolding, misligated intermediate states are populated, as His26 and His33 can bind to the heme iron (*6*−*8*). A recent study suggests that His33 acts as the predominant non-native ligand (*9*). It has been shown that the non-native histidine ligation also occurs in the unfolded state (*10*, *11*), leading to multiple starting points for refolding. Non-native ligation leads to complicated kinetics, which can be difficult to resolve. The addition of imidazole to the solvent simplifies the kinetics. Imidazole

[†]This work was supported by National Institutes of Health Grant RR-01209 (to K.O.H.) and by National Science Foundation MCB95-07280 (to A.L.F.). Data were collected at Beam Line 4-2 at Stanford Synchrotron Radiation Laboratory (SSRL). SSRL is supported by the US Department of Energy, Office of Basic Energy Sciences, and in part by the National Institutes of Health, National Center for Research Resources, Biomedical Technology Program and by the Department of Energy, Office of Biological and Environmental Research.

* Correspondence should be addressed to Sebastian Doniach at LAM/Applied Physics, McCullough Building, Stanford University, Stanford, CA 94305-4045. Tel: 650 723 4786. Fax: 650 725 2189. E-mail: doniach@drizzle.stanford.edu.

[‡]Stanford University.

[§] Weill Medical College of Cornell University.

[||] University of California.

[⊥] Russian Academy of Sciences.

[#] Department of Chemistry, Stanford University and Stanford Synchrotron Radiation Laboratory.

[1] Abbreviations: SAXS, small-angle X-ray scattering; GdnHCl, guanidine hydrochloride; SVD, singular value decomposition; R_g, radius of gyration; N, native; U, unfolded state; C, collapsed state; D, dimer; SEC, size-exclusion chromatography.

10.1021/bi991337k CCC: $18.00 © 1999 American Chemical Society
Published on Web 10/28/1999

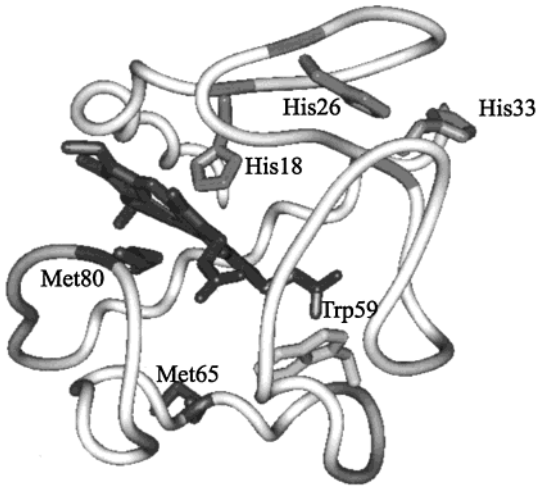

FIGURE 1: Crystal structure of horse heart Cyt *c* (*2*). Residues relevant to the discussion were drawn. The image was created with MidasPlus (USCF Computer Graphics Laboratory, San Francisco).

binds to the heme iron, displacing Met80, even in native Cyt *c* (*3*, *12*). This prevents the misligated states from forming and therefore suppresses these intermediate states during refolding (*7*, *12*−*14*). Lowering the pH can also diminish the population of the misligated intermediate, as the histidine residues become protonated (*7*, *11*). However, a small fraction of the protein molecules will still populate the intermediate state, leading to complex kinetics similar to those observed at neutral pH (*7*, *11*).

The refolding of Cyt *c* at neutral pH in the presence of imidazole can be described by the simple scheme (*13*, *14*)

$$U \rightarrow C \rightarrow N \quad (1)$$

where U is the unfolded state, C is the collapsed state, and N is the native state. Traditional time-resolved methods show the collapsed state to form within the mixing dead-time of the experiments (*7*, *15*), followed by single exponential kinetics that reflect the conversion of the collapsed state to native protein. Recent efforts have sought to better understand the cause of the formation of the collapsed state, which can be viewed as a fast, energetically barrier-free and downhill event (*13*, *15*). However, a recent result is not consistent with this interpretation (*14*). An ultra-rapid mixer has been developed that can probe kinetic time scales as short as ~45 μs (*16*). With this time resolution, the data on Cyt *c* refolding could be represented by a double-exponential fit, suggesting that the collapse involves overcoming an energy barrier (*14*). The initial phase, believed to be due to the U → C transition, has a time constant of 50 μs.

Tryptophan (Trp) fluorescence has long been used as a conformational probe for Cyt *c* (*7*, *12*−*15*, *17*−*19*). For the native protein Trp fluorescence is almost completely quenched due to the nonradiative energy transfer to the heme. The measured degree of quenching can be used to calculate the average heme-Trp distance (*19*). However, this method effectively measures only one distance within the protein and is not necessarily indicative of the overall size of the protein. Therefore, we sought to study the Cyt *c* refolding kinetics with time-resolved SAXS, which directly measures the overall size of the protein. Further, inspection of the forwarding scattering amplitude, $I(0)$, gives information on the association state of the protein. Associated states formed during refolding are often part of a polydisperse system that also includes monomers. The challenge is to characterize the associated state (and determine whether it is unique) in a polydisperse system. Recent developments applying the singular value decomposition (SVD) technique to SAXS spectra have created the ability to characterize individual monomeric conformational states in a polydisperse system (*20*−*23*). We present here a further application of this technique to characterize the overall size and shape information of transiently formed aggregates. By using the information provided by $I(0)$ and an SVD analysis of the time-resolved SAXS spectra, we identified and characterized a transient dimer populated during Cyt *c* refolding. (Note that the fluorescence experiments (*14*) were carried out at micromolar concentrations and therefore would not be expected to lead to transient dimer formation.)

It is important to note that the time resolution of stopped-flow SAXS experiments is not great enough to resolve the C → N conversion. This paper focuses on the time scales where this conversion is complete for the fraction of molecules that do not form dimers, and thus, only the much slower conversion from dimer to native state is observed.

MATERIALS AND METHODS

Materials. Horse heart cytochrome *c* was purchased from Sigma Chemical Company (St. Louis, MO). The protein was used without additional purification; 100 mM phosphate, 200 mM imidazole stock buffer solutions were prepared with (4.8 and 3.6 M) and without GdnHCl at pH 7. Stock buffers were filtered with a 0.2 μm acetate filter. Protein was dissolved directly into the GdnHCl solutions; the sample was subsequently filtered with a 0.2 μm filter to remove undissolved protein from the sample. Cytochrome *c* concentrations were measured by UV absorption ($\epsilon_{280} = 8.55$ mL mg^{-1} cm^{-1}). For equilibrium unfolding measurements, protein concentrations of 5−7 mg/mL were used. Protein concentrations for refolding measurements are listed in Table 2.

SAXS. Measurements were made using the SAXS instrument on Beam Line 4−2 at Stanford Synchrotron Radiation Laboratory (*24*). X-ray energy was selected at 8980 eV (Cu edge) by a pair of Mo/B_4C multilayer monochromator crystals (*25*). Scattering patterns were recorded by a linear position-sensitive proportional counter, which was filled with an 80% Xe/20% CO_2 gas mixture. Scattering patterns were normalized by incident X-ray flux, which was measured with a short-length ion chamber before the sample. The sample-to-detector distance was calibrated to be 230 cm, using a cholesterol myristate sample.

To avoid radiation damage of the sample in equilibrium and manual mixing measurements, we continuously passed the protein solution through a 1.3 mm path length observation flow cell with 25 μm mica windows. For manual mix refolding measurements, denatured protein was mixed with dilution buffer using a pipet and then transferred to the flow cell system. Data accumulation began 1.5 min after mixing. Kinetics on the order of seconds were studied using a stopped-flow mixer (*1*, *26*). For both stopped-flow conditions (listed in Table 2), data from 300 mixing events were integrated. Background measurements were performed before and after each protein measurement and then averaged before being used for background subtraction. All SAXS measurements were performed at 23 ± 1 °C.

Guinier Analysis. Radii of gyration were calculated according to the Guinier approximation (*27*):

$$Ln(I(S)) = Ln(I(0)) - \frac{4\pi^2 R_g^2}{3} S^2 \quad (2)$$

where $I(S)$ = scattering profile as a function of S, $S = (2\sin\theta)/\lambda$, 2θ = scattering angle, λ = X-ray wavelength, $I(0)$ = forward scattering amplitude, and R_g = radius of gyration. Guinier fits and errors were calculating using KaleidaGraph (Synergy Software, Reading, PA). The fitting range used was 0.0045−0.0071 Å^{-1} in S.

Thermodynamic Fit. For a two-state system, the measured R_g is an average radius of gyration of all species in solution:

$$R_g^2 = f_N R_N^2 + f_U R_U^2 \quad (3)$$

where f_N = native fraction, f_U = unfolded fraction, R_N = radius of gyration for the native state, and R_U = radius of gyration for the unfolded state (*20*). Since R_g^2 maintains a linear dependence upon the fractional populations, R_g^2 is the relevant quantity to fit (not R_g). The fractional populations can be expressed as

$$f_N = \frac{1}{1 + e^{-\Delta G_U/RT}} \quad (4a)$$

$$f_U = \frac{e^{-\Delta G_U/RT}}{1 + e^{-\Delta G_U/RT}} \quad (4b)$$

where ΔG_U is the change in free energy of state U relative to the native state. In our model, the free energy was assumed to have a linear dependence on the denaturant concentration (*28*):

$$\Delta G_U = \Delta G_U^\circ - m[GdnHCl] \quad (5)$$

where ΔG_U° is the change in free energy of state U relative to the native state at neutral pH and in the absence of denaturants and m is the linear dependence of ΔG_U upon the denaturant concentration.

Pair Distribution Function. The pair distribution functions, $P(R)$, were calculated using the indirect transformation method implemented by the program GNOM (*29*). $P(R)$ measures the density of intramolecular electron pair distance distributions within the scatterer, that is, the protein molecule. The range use for the $P(R)$ calculation was from 0.0045 to 0.0350 Å^{-1} in S.

Size Exclusion Liquid Chromatography. Gel-filtration measurements were carried out on a Superose-12 column using a Pharmacia FPLC apparatus. To determine the molecular mass of Cyt *c*, we calibrated the gel-filtration column using standard procedures (*30*, *31*). Proteins from a standard molecular mass (M) marker set were passed through the column, and the retention coefficient (K_d) for each individual protein was determined:

$$K_d = (V - V_o)/(V_t - V_o) \quad (6)$$

where V_o is the column void volume determined as the elution volume of blue dextran; V_t is the total solvent-accessible column volume determined as the elution volume of acetone; and V is the elution volume of the given protein.

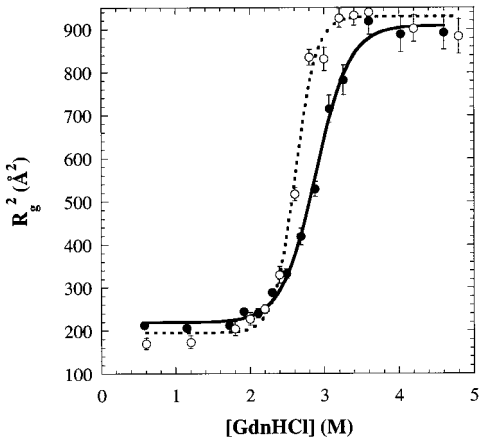

FIGURE 2: The radius of gyration (R_g) dependence upon GdnHCl at pH 7. The open circles (○) and closed circles (●) represent the measurements taken in the absence and presence (200 mM) of imidazole, repectively. The dashed and solid curves are two-state thermodynamic fits (see Methods) to the data. The thermodynamic parameters resulting from the fits are listed in Table 1.

Least-squares analysis was used to fit the data to generate the calibration curve. Kinetic experiments were carried out as follows. Cyt *c* was dissolved in phosphate buffer, pH 7.5, containing 4.8 M GdnHCl in the presence or absence of 100 mM imidazole. For final protein concentrations of 6.5, 3.3, and 1.6 mg/mL, the initial solutions contained 52, 26.4, and 13 mg/mL Cyt *c*, respectively. Immediately prior to injection into the gel-filtration column, a small amount of the concentrated Cyt *c* solution was diluted 8-fold by phosphate buffer (pH 7.5), containing 0 or 100 mM imidazole. The flow rate was 1 mL/min.

RESULTS

Equilibrium Unfolding of Cyt c Monitored by SAXS. The size of a protein can be determined from X-ray scattering in the small angle region. The protein scattering in this region, known as the Guinier region, is given by

$$I(S) = I(0)e^{-4\pi^2 R_g^2 S^2/3} \quad (7)$$

where R_g = radius of gyration, $I(S)$ = scattering profile as a function of S, $S = (2\sin\theta)/\lambda$, 2θ = scattering angle, λ = X-ray wavelength, and $I(0)$ = forward scattering amplitude. Radii of gyration can be obtained from the slope of the curve $\ln(I(S))$ vs S^2 in the Guinier region (see Materials & Methods Section).

Data on the equilibrium GdnHCl-induced unfolding of Cyt *c* at pH 7.0 in the presence and absence of imidazole, monitored by the change in R_g^2, are presented in Figure 2. Both cases show a cooperative unfolding transition on addition of denaturant. From a two-state thermodynamic fit of the experimental data in the presence of imidazole to R_g^2 (see Methods), the R_g values for native and unfolded protein were determined to be 14.7 ± 0.3 Å and 29.9 ± 0.2 Å, respectively. The value for unfolded Cyt *c* is in excellent agreement with the R_g value measured in the absence of imidazole, 30.1 ± 0.2 Å (*21*). Further, the entire scattering spectra were nearly identical (data not shown). However, the native R_g value is noticeably larger (∼1 Å) than the value measured in the absence of imidazole, 13.8 ± 0.3 Å (*21*). The increased R_g demonstrates that displacement of Met80

Table 1. Thermodynamic Parameters and Radii of Gyration (R_g) for the Native and Unfolded States Obtained from a Two-State Thermodynamic Fit to R_g^2 (See Methods)

buffer	ΔG_U° (kcal/mol)	m (kcal mol^{-1} M^{-1})	native R_g (Å)	denatured R_g (Å)
no imidazole	10.7 ± 0.9	4.1 ± 0.3	13.8 ± 0.3	30.1 ± 0.2
200 mM imidazole	7.2 ± 0.5	2.5 ± 0.2	14.7 ± 0.3	29.9 ± 0.2

Table 2. Protein Refolding Kinetics as a Function of Final Denaturant Concentration and Protein Concentration

condition	$I(0, t = 0)/I_N(0)$	time constant
MM: 4.8 M → 0.6 M GdnHCl, 6.5 mg/mL	1.8	>10 min
MM: 4.8 M → 0.6 M GdnHCl, 3.3 mg/mL	1.5	>10 min
MM: 4.8 M → 0.6 M GdnHCl, 1.6 mg/mL	1.3	>10 min
MM: 4.8 M → 1.2 M GdnHCl, 6.5 mg/mL	1.3	3.3 ± 0.6 min
MM: 4.8 M → 1.2 M GdnHCl, 3.3 mg/mL	1.2	4.9 ± 3.3 min
SF: 3.6 M → 1.8 M GdnHCl, ~11 mg/mL	1.3	13.8 ± 1.3 s
SF: 4.8 M → 2.4 M GdnHCl, ~11 mg/mL	1.2	0.6 ± 0.1 s

[a] $I(0, t = 0)$ is the forward scattering amplitude at zero time. This value was obtained by a single-exponential fit to the data. The time constants were also obtained from these fits. For the 0.6 M GdnHCl conditions, no significant time dependence in $I(0)$ was observed, and therefore, the data was simply fit to a straight line to obtain the $I(0, t = 0)$ value. MM = manual mixing; SF = stopped-flow experiment.

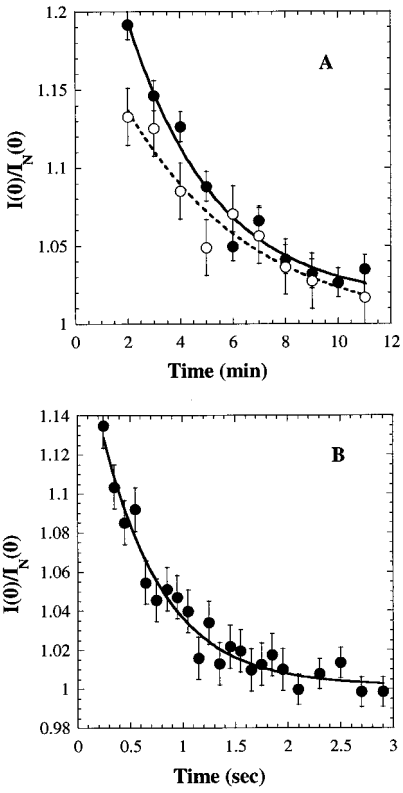

FIGURE 3: Time-resolved changes in the forward scattering amplitude, $I(0)$, during refolding. Panel A shows refolding in 1.2 M GdnHCl (○, 3.3 mg/mL; ●, 6.5 mg/mL). Panel B shows refolding in 2.4 M GdnHCl. The data are normalized by the native value, $I_N(0)$. The curves are single-exponential fits to the data, from which the scattering amplitude at zero time, $I(0, t = 0)$, and the time constant of dissociation can be determined. These parameters are listed in Table 2 for all of the refolding conditions.

from the heme iron by imidazole causes measurable structural changes in the protein. These changes are reinforced by 2D NMR data, which shows a "swelling" of the protein on the Met80 side of the heme with addition of imidazole (*32*).

The thermodynamic parameters from the two-state fit of the experimental data also indicate a structural difference between the native state in the absence and presence of imidazole (Table 1). In the presence of imidazole, our values, $\Delta G_U^\circ = 7.3 \pm 0.5$ kcal/mol and $m = 2.5 \pm 0.2$ kcal mol^{-1} M^{-1}, were similar to those measured with fluorescence: $\Delta G_U^\circ = 8.2$ kcal/mol and $m = 2.85$ kcal mol^{-1} M^{-1} (*13*). The m value is roughly proportional to the difference in solvent-accessible surface area between the native state (N) and the unfolded state (U) (*33*, *34*). Since the unfolded states in the presence and absence of imidazole are indistinguishable (as judged by R_g and nearly identical SAXS profiles), they are presumed to have the same solvent-accessible surface area. Hence, the smaller m value (2.5 ± 0.2 vs 4.1 ± 0.3 kcal mol^{-1} M^{-1} in the absence of imidazole) indicates that in the presence of imidazole Cyt *c* native state has more exposed solvent-accessible surface area than that in the absence of imidazole. Further, the lower ΔG_U° value implies that the native state in the presence of imidazole is less stable than in the absence of imidazole.

Kinetics of Cyt c Refolding: Guinier Analysis. The refolding kinetics of Cyt *c* were studied at neutral pH in the presence of (200 mM) imidazole under a variety of final denaturant concentrations (Table 2). Most significant changes were observed in the forward scattering amplitude, $I(0)$ (see Methods). $I(0)$ is proportional to $n\rho_c^2V^2$, where n is the number of scatterers (protein molecules) in solution, ρ_c is the electron density difference between the scatterer and the solvent, and V is the volume of the scatterer (*27*). Thus, changes in $I(0)$ indicate changes in the association state of a protein. For example, if a protein undergoes transition from a fully monomeric to a fully dimeric state, the number of scatterers, n, would decrease by a factor of 2, and the volume of each scatterer, V, would roughly increase by a factor of 2; the net effect would be a 2-fold increase in $I(0)$.

The time-resolved $I(0)$ courses for two of the refolding conditions are shown in Figure 3. For all conditions studied (listed in Table 2), kinetic decreases in $I(0)$ are observed, suggesting dissociation of an associated state (denoted *A*). However, not all protein molecules become involved in associated states. The smallest-order aggregate possible is a dimer, and as discussed above, full dimerization of the sample would lead to a 2-fold increase in $I(0)$. Since $I(0, t = 0)/I_N(0)$ does not reach the value 2 (see Table 2), some of the protein molecules fold directly, without becoming involved in an associated state.

The magnitude of the association, as judged by the initial $I(0)$ value, depends on the final denaturant concentration and the protein concentration (Table 2). A 2-fold decrease in $I(0)$ has been reported for the refolding of myoglobin (*1*), where the results have been interpreted as the dissociation of a transiently populated dimer. However, recent publications have pointed out that changes in $I(0)$, as large as 20%, can be caused by conformational changes in monomeric protein samples (*22*, *23*). This effect is believed to be due to changes

in the volume of the hydration shell. Protein conformations with increased solvent-accessible surface area presumably will have a hydration shell with a larger volume. This larger volume contributes to the scattering, leading to a larger $I(0)$ value. To ensure that $I(0)$ changes were caused by kinetic dissociation of transient aggregates rather than by conformational changes in the monomeric protein, we used the following approach. The characteristic time of the monomeric refolding of Cyt *c* has been determined by fluorescence (*13*, *14*). In our analysis, time points shorter than 4τ for folding (τ = time constant) from the time of mixing were discarded (τ = ~1, ~4, ~12, and ~25 ms for final denaturant concentrations of 0.6, 1.2, 1.8, and 2.4 M GdnHCl, respectively (*13*)). Therefore we believe that the observed kinetic changes presented here represent dissociation of an associated state and not conformational changes of monomeric protein.

Figure 3 and Table 2 document the effect of denaturant concentration on the refolding kinetics. Increasing the final denaturant concentration accelerates the kinetics of dissociation and reduces the magnitude of the association (as judged by the $I(0; t = 0)$). This indicates that denaturant destabilizes the associated state and impedes its formation. Interestingly, formation of the associated state does not prevent later formation of the native state (denoted N), as all conditions (with the exception of 0.6 M GdnHCl) reached the native baseline, as judged by $I(0)$ and R_g. The kinetics at 0.6 M GdnHCl are very slow, as the signal does not reach the native baseline, even after an hour.

Singular Value Decomposition. The minimum number of conformational states present during the observed kinetics can be determined using the singular value decomposition (SVD) method (*35*). This analysis takes a column matrix of scattering profiles and represents them as a linear superposition of orthogonal basis curves:

$$I(S,t) = \sum_{j=1}^{M} b_j(t) u_j(S) \tag{8}$$

where $I(S,t)$ are the time-resolved scattering profiles; S is the momentum transfer; $b_j(t)$ are the time-dependent basis coefficients; $u_j(S)$ are the basis functions; and M is the number of time-resolved scattering profiles. Each SAXS profile is a linear superposition of the profiles of the distinct states that can be populated by the protein. For the states to be distinct, their scattering profiles must be linearly independent. The number of states therefore indicates the number of linearly independent scattering curves required to construct the scattering profile at any time point. The number of significant SVD components (M) determines the minimum number of linearly independent scattering profiles needed to construct the series of scattering patterns, $I(S,t)$. Therefore, the number of significant SVD components reveals the number of distinguishable conformational states.

An SVD analysis of the time-resolved scattering profiles for each refolding condition shows that at least two basis functions are required to represent the data (Figure 4), and therefore, two conformational states of the protein exist for each refolding condition, although the basis functions for each condition differ slightly (the basis functions for one of these conditions is shown in Figure 5). One of the conformational states is the native state (N) and the other state

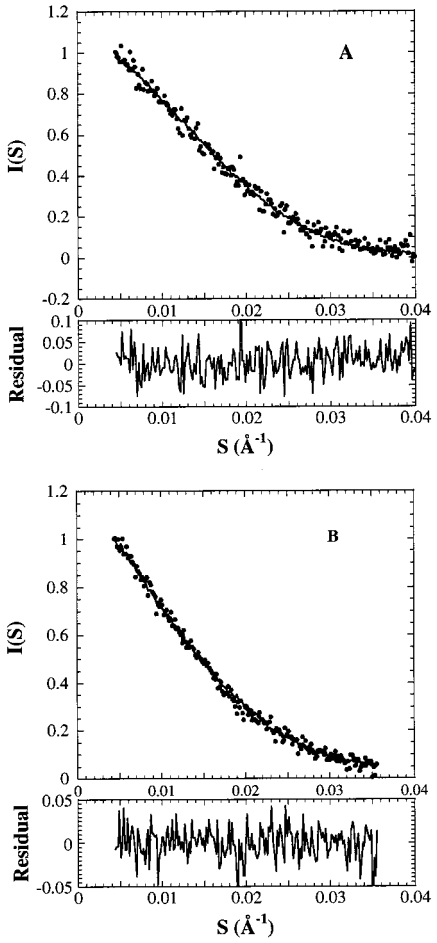

FIGURE 4: Two component SVD reconstruction of time-resolved scattering profiles. Panels A and B represent refolding in 1.2 M GdnHCl (5 min time point) and 2.4 M GdnHCl (500 ms time point). The solid dots are original data and the solid lines are the 2 component SVD reconstructions. Below each scattering profile, the residual of the fit is plotted.

corresponds to the associated state (A). (As discussed earlier, the kinetics studied in this paper do not include the early time scales where the collapsed state is populated.) The challenge is to characterize *A* and determine its order (dimer, trimer, etc.).

Kinetic Model. If the refolding kinetics are a two state system, then the scattering profile measured at any point in time is

$$I(S,t) = f_N(t) I_N(S) + f_A(t) I_A(S) \tag{9}$$

where $f_N(t)$, $f_A(t)$ are the fractional populations and $I_N(S)$, $I_A(S)$ are the scattering profiles of N and A, respectively. The native scattering profile can be easily determined with an equilibrium measurement. (At 2.4 M GdnHCl, ~10% of the protein molecules are unfolded. For these refolding conditions the 2.4 M GdnHCl equilibrium scattering profile can be used as the "native" profile.) Just as the time-resolved scattering profiles, $I(S,t)$, can be represented by a linear superposition of the basis scattering curves, so can the scattering profile of *A*:

$$I_A(S) = b_1^A u_1(S) + b_2^A u_2(S) \tag{10}$$

where b_1^A and b_2^A are the basis coefficients for *A*, which must be determined in order to construct a scattering profile for *A*.

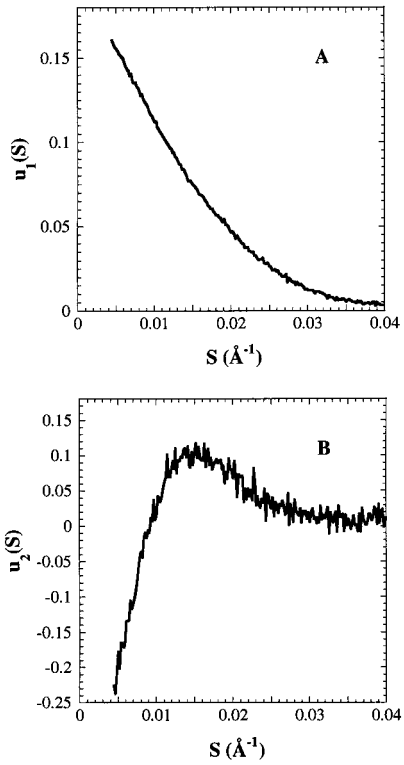

FIGURE 5: Basis functions from the 4.8 M → 2.4 M GdnHCl stopped-flow data: (A) $u_1(S)$; (B) $u_2(S)$.

For each refolding condition, the associated state basis coefficients, b_1^A and b_2^A, as well as time-dependent populations of N and A, $f_N(t)$ and $f_A(t)$, were determined by minimizing the following equation:

$$\chi^2 = \sum_{i=1}^{M}\sum_{j=1}^{N}\left(\frac{I_{exp}(S_j,t_i) - I_{fit}(S_j,t_i)}{\sigma(S_j,t_i)}\right)^2 \qquad (11)$$

where $I_{exp}(S,t)$ are the measured time-resolved scattering profiles, $I_{fit}(S,t)$ are the scattering profiles constructed using eq 9, and $\sigma(S,t)$ are the standard deviations for the scattering profiles. The scattering profile of the *A* state for each refolding condition can be determined by inserting the fitted b_1^A and b_2^A values into eq 10. These scattering profiles are plotted in Figure 6, and the results of the Guinier analysis of these profiles are presented in Table 3. Looking at Table 3, it is important to note that, for all conditions, the scattering profile for *A* yields an $I(0)$ value twice that of *N*. As mentioned earlier, a fully dimerized sample would result in a 2-fold increase in $I(0)$. Therefore, the *A* state appears to be a dimer (denoted D) for all of the refolding conditions studied.

Size Exclusion Chromotography. Size exclusion chromatography (SEC) was used for independent verification that *A* was in fact a dimer. This technique is usually applied for the investigation of slow protein refolding kinetics (*36*−*41*). As the characteristic time of SEC in FPLC configuration was about 20 min, we were unable to investigate Cyt *c* refolding kinetics by this technique with final GdnHCl higher than 0.6 M (see Table 2). Figure 7 represents typical elution profiles of Cyt *c* for 0.6 M GdnHCl measured at different final protein concentrations in the absence and presence of imidazole. The figure reflects the presence of two major components in the reaction mixture. The component with elution volume of ∼18.4 mL (marked as *M*) corresponds to native monomeric Cyt *c*, as its apparent molecular mass determined by SEC (∼11.0 kDa) is very close to the real molecular mass of the protein (11.7 kDa). Another major component, with elution volume of ∼ 16.4 mL (marked as

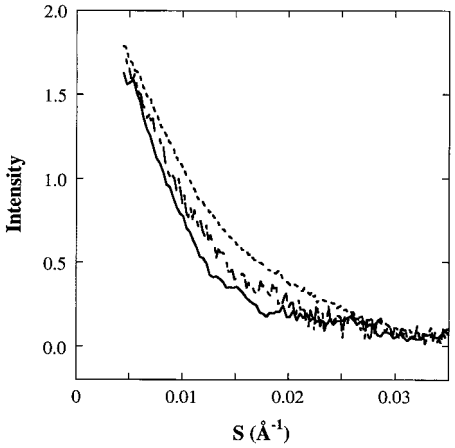

FIGURE 6: Denaturant dependence of constructed dimer scattering profiles. The solid, broken, and dashed lines are dimer scattering profiles at 2.4 M, 1.8 M, and low (see Table 3 caption) GdnHCl. The results of a Guinier analysis of these profiles are presented in Table 3.

Table 3. Parameters Obtained from Guinier Analysis of the Model Dimer Scattering Profiles Obtained Using SVD[a]

condition	R_g (Å)	$I_A(0)/I_N(0)$
MM[b]: 4.8 M → 0.6 M GdnHCl	23.4 ± 0.2	2.08 ± 0.10
MM[b]: 4.8 M → 1.2 M GdnHCl	23.4 ± 0.2	2.08 ± 0.10
SF: 3.6 M →1.8 M GdnHCl	25.1 ± 0.3	2.01 ± 0.10
SF: 4.8 M → 2.4 M GdnHCl	26.7 ± 0.5	2.10 ± 0.10

[a] The forward scattering amplitude, $I_A(0)$, is normalized by the native value, $I_N(0)$. The 2-fold increase over the native value suggests that the associated state is a dimer: MM = manual mixing experiment, SF = stopped-flow experiment. [b] The data from each manual mix condition were too noisy to fit independently for a dimer scattering profile. Therefore, the data sets were combined prior to determination of the dimer scattering profile. Hence, the same parameters are listed for each manual mix condition.

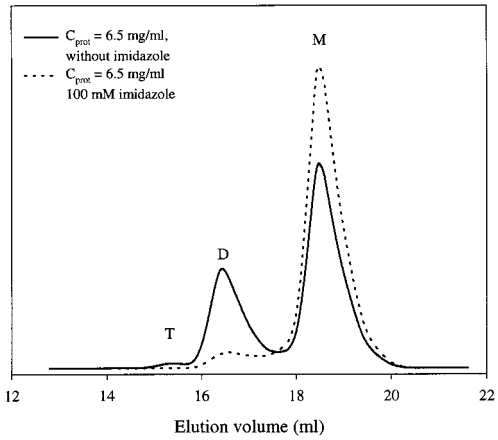

FIGURE 7: FPLC gel-filtration studies of Cyt *c* refolding kinetics. A small amount of GdnHCl-unfolded Cyt c (4.8 M GdnHCl) was 8-fold diluted by phosphate buffer (pH 7.5). Sample was injected into the gel-filtration column without delay. Plot represents dependence of the time "zero" elution profile on the presence of imidazole. M corresponds to the monomeric species, D represents the transient dimers, and T refers to tetramers.

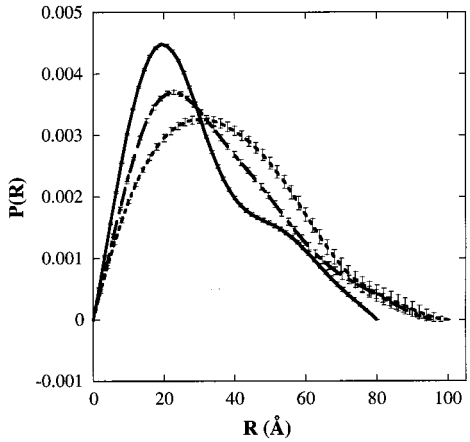

FIGURE 8: Pair distribution function, $P(R)$, for the dimer state as a function of denaturant. The dashed, broken, and solid lines are the $P(R)$ curves for the dimer at 2.4 M, 1.8 M, and low (look at Table 3 caption) GdnHCl. Notice that with increasing [GdnHCl] the hump at ~50 Å disappears and D_{max} (where $P(R = D_{max}) = 0$) increases. Also, the first peak at ~20 Å for low GdnHCl moves out to ~35 Å at 2.4 M GdnHCl. These trends signal a decrease in globularity and structure within the dimer.

D), is characterized by an apparent molecular mass of ~26 kDa. This value is somewhat larger than that expected for the dimer of native Cyt *c* ($2 \times 11.7 = 23.4$ kDa), reflecting the fact that the transient associated state is comprised of compact, but non-native molecules.

Figure 7 illustrates that imidazole decreases the amount of the Cyt c dimer. Thus, the results of the SEC experiments confirm the SAXS observations that dimers are formed during refolding, that there are no significant concentrations of higher oligomers, and that the presence of imidazole decreases the amount of dimer.

Characteristics of the Transient Dimer. The conformation of the dimer demonstrates a denaturant dependence, as observed by R_g (Table 3) and the pair distribution function, $P(R)$ (Figure 8). R_g of the dimer expands from 23.5 ± 0.2 Å at low GdnHCl concentrations to 26.7 ± 0.5 Å at 2.4 M GdnHCl. More structural information about the dimer can be gained by inspection of the $P(R)$ function. $P(R)$ measures the density of intramolecular pair distance distributions within the scatterer, that is, the protein molecule. $P(R)$ for the dimer at low denaturant concentration displays a bimodal feature, with a peak at ~20 Å and a hump at ~50 Å. The peak at ~20 Å represents electron−electron pairs within the protein monomers that are involved in the dimer. The hump at ~50 Å indicates the electron−electron pairs between the two protein monomers. The disappearance of this hump and the increasing maximum dimension, D_{max}, with increasing denaturant concentration illustrate decreasing globularization of the dimer and increased size (as also determined by R_g). This behavior of the dimer suggests that some parts of the protein molecule are distorted and became even more random and expanded with increasing denaturant concentration.

DISCUSSION

By using time-resolved SAXS and SVD, we have successfully identified and characterized a transient dimer formed during the refolding of Cyt *c* at millimolar concentrations. The dimer has properties similar to a molten-globule state. Kratky plots and $P(R)$ plots show the dimer to be globular. However, the increase in values of R_g and D_{max} in $P(R)$ plots with increasing denaturant concentration reflects the presence of some disordered segments in the dimer molecule. Both observations are consistent with the concept of partially folded intermediates in monomeric proteins in which parts of the molecule have nativelike structural organization, whereas other segments are considerably distorted (*42*, *43*). Denaturant-induced expansion of molten globule-like intermediates has been observed previously; for example, the GdnHCl-induced unfolding of the molecular chaperone DnaK is accompanied by the formation of an intermediate state that continuously expands with increasing GdnHCl concentration (*44*).

Two important questions arise about the transient dimer of Cyt *c*. (1) What interactions are responsible for its stabilization? (2) Is the dimer a kinetic trap or on a parallel and productive pathway? A transient associated state has been observed during the refolding of myoglobin with time-resolved SAXS (*1*). In that case, a 2-fold decrease in $I(0)$ suggested that the associated state was a dimer. We have established that a dimer is the only associated species formed during refolding of Cyt *c*. This suggests the existence of specific interactions stabilizing the dimers; otherwise larger oligomers would be anticipated. Very stable dimers of partially folded intermediates have been observed with staphylococcal nuclease (*45*). The most likely explanation for relatively stable dimers of partially folded intermediates is some type of domain-swapping interaction (*46*) in which part of one molecule is deeply embedded in the other, and vice versa.

In the absence of imidazole, cross-ligation could be considered for the stabilization of the dimer. As mentioned in the introduction, Met80 detaches from the heme iron in denaturing conditions (*3*). In cross-ligation, a histidine or methionine residue from one protein molecule would bind to the vacated position on the heme iron of another protein molecule. However, in the experiments presented in this paper, cross-ligation can be ruled out, as experiments were carried out in the presence of imidazole, and as the bound imidazole would prevent cross-ligation from occurring. The decreased amount of dimer in the presence of imidazole may reflect a small but significant change in the structure of the partially folded (collapsed) intermediate depending on the heme ligand, which affects the intermolecular interaction leading to dimer formation. Alternately, it may reflect the absence of such an intermediate when the heme ligand is imidazole.

Because of the simplified kinetics of Cyt *c* refolding in the presence of imidazole, only a few places along the folding pathway exist where dimerization can occur: in the unfolded, collapsed, or native states. The unfolded and native states are stable under equilibrium conditions (high and low GdnHCl concentrations, respectively), and their dimerization can be easily tested using small-angle X-ray scattering from the dependence of R_g on concentration. With both the unfolded and native states, no aggregation was observed in the protein concentration range where the refolding experiments took place. This means that the dimer must be formed from the collapsed state.

An interesting question is whether protein self-association can facilitate the folding of monomeric proteins. We have shown that formation of the dimer does not prevent the

eventual formation of the native state. However, the slow rate of dimer dissociation indicates that in this case the self-associated state is effectively a kinetic trap, regardless of the actual folding pathway.

The presence of a relatively long-lived aggregated intermediate during refolding emphasizes that caution is necessary in interpreting kinetic transients during protein folding, especially with higher protein concentrations. As has been noted before (*47*), transient aggregates in protein folding are easily mistaken for folding intermediates. The marked dependence of the stability and lifetime of the cytochrome *c* dimer as a function of denaturant concentration is very interesting. The fact that relatively low concentrations of GdnHCl, such as are often found in refolding conditions, led to very long-lived dimers is also cause for concern.

ACKNOWLEDGMENT

We thank Dr. Hirotsugu Tsuruta for help with the SAXS setup. We thank Jim Hofrichter for helpful discussions. Experiments were performed on Beam Line 4-2 at Stanford Synchrotron Radiation Laboratory (SSRL).

REFERENCES

1. Eliezer, D., Chiba, K., Tsuruta, H., Doniach, S., Hodgson, K. O., and Kihara, H. (1993) *Biophys. J. 65*, 912−917.
2. Bushnell, G. W., Louis, G. V., and Brayer, G. D. (1990) *J. Mol. Biol. 214*, 585−596.
3. Babul, J., and Stellwagen, E. (1971) *Biopolymers 11*, 2359−2361.
4. Babul, J., and Stellwagen, E. (1972) *Biochemistry 11*, 1195−1200.
5. Tsong, T. Y. (1975) *Biochemistry 14*, 1542−1547.
6. Roder, H., Elove, G. A., and Englander, S. W. (1988) *Nature 335*, 700−704.
7. Elove, G. A., Bhuyan, A. K., and Roder, H. (1994) *Biochemistry 33*, 6925−6935.
8. Sosnick, T. R., Mayne, L., Hiller, R., and Englander, S. W. (1994) *Nat. Struct. Biol. 1*, 149−156.
9. Colon, W., Wakem, L. P., Sherman, F., and Roder, H. (1997) *Biochemistry 36*, 12535−12541.
10. Takahashi, S., Yeh, S.-R., Das, T. K., Chan, C.-K., and Gottfried, D. S. (1997) *Nat. Struct. Biol. 4*, 44−50.
11. Yeh, S.-R., Takahashi, S., Fan, B., and Rousseau, D. L. (1997) *Nat. Struct. Biol. 4*, 51−56.
12. Brems, D. N., and Stellwagen, E. (1983) *J. Biol. Chem. 258*, 3655−3660.
13. 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*, 1179−1784.
14. Shastry, M. C. R., and Roder, H. (1998) *Nat. Struct. Biol. 5*, 385−392.
15. Sosnick, T. R., Shtilerman, M. D., Mayne, L., and Englander, S. W. (1997) *Proc. Natl. Acad. Sci. U.S.A. 94*, 8545−8550.
16. Shastry, M. C. R., Luck, S. D., and Roder, H. (1998) *Biophys. J. 74*, 2714−2721.
17. Colon, W., Elove, G. A., Wakem, L. P., Sherman, F., and Roder, H. (1996) *Biochemistry 35*, 5538−5549.
18. Elove, G. A., Chaffote, A. F., Roder, H., and Goldberg, M. E. (1992) *Biochemistry 31*, 6876−6883.
19. Tsong, T. Y. (1976) *Biochemistry 15*, 5467−5473.
20. Chen, L., Hodgson, K. O., and Doniach, S. (1996) *J. Mol. Biol. 261*, 658−671.
21. Segel, D. J., Fink, A. L., Hodgson, K. O., and Doniach, S. (1998) *Biochemistry 37*, 12443−12541.
22. Chen, L., Wildegger, G., Kiefhaber, T., Hodgson, K. O., and Doniach, S. (1998) *J. Mol. Biol. 276*, 225−237.
23. Segel, D. J., Bachmann, A., Hofrichter, J., Hodgson, K. O., Doniach, S., and Kiefhaber, T. (1999) *J. Mol. Biol. 288*, 489−499.
24. Wakatsuki, S., Hodgson, K. O., Eliezer, D., Rice, M., Hubbard, S., Gillis, N., and Doniach, S. (1992) *Rev. Sci. Instrum. 63*, 1736−1740.
25. Tsuruta, H., Brennan, S., Rek, Z. U., Irving, T. C., Tompkins, W. H., and Hodgson, K. O. (1998) *J. Appl. Crystallogr. 31*, 672−682.
26. Tsuruta, H., Nagamura, T., Kimura, K., Igarashi, Y., Kajita, A., Wang, Z.-X., Wakabayashi, K., Amemiya, Y., and Kihara, H. (1989) *Rev. Sci. Instrum. 60*, 2356−2358.
27. Glatter, O., and Kratky, O. (1982) *Small-Angle X-ray Scattering*; Academic Press, New York.
28. Pace, C. N. (1986) in *Methods in Enzymology*; (Colowick, S. P., Kaplan, N. O., Eds.) pp 266−280, Academic Press, Orlando, FL.
29. Semenyuk, A. V., and Svergun, D. I. (1991) *J. Appl. Crystallogr. 24*, 537−540.
30. Ackers, G. L. (1967) *J. Biol. Chem. 242*, 3237−3238.
31. Ackers, G. L. (1970) *Adv. Protein Chem. 24*, 343−446.
32. Liu, G., Shao, W., Huang, X., Wu, H., and Tang, W. (1996) *Biochim. Biophys. Acta 1277*, 61−82.
33. Alonso, D. O. V., and Dill, K. A. (1991) *Biochemistry 30*, 5974−5985.
34. Schellman, J. A. (1978) *Biopolymers 17*, 1305−1322.
35. Henry, E. R., and Hofrichter, J. (1992) *Methods Enzymol. 210*, 129−192.
36. Shalongo, W., Ledger, R., Jagannadham, M. V., and Stellwagen, E. (1987) *Biochemistry 26*, 3135−3141.
37. Shalongo, W., Jagannadham, M. V., Heid, P., and Stellwagen, E. (1992) *Biochemistry 31*, 11390−11396.
38. Shalongo, W., Jagannadham, M., and Stellwagen, E. (1993) *Biopolymers 33*, 135−145.
39. Palleros, D. R., Shi, L., Reid, K. L., and Fink, A. L. (1993) *Biochemistry 32*, 4314−4321.
40. Zerovnik, E., Jerala, R., Poklar, N., Zitko, L. K., and Turk, V. (1994) *Biochim. Biophys. Acta 1209*, 140−143.
41. Batas, D., Jones, H. R., and Chaudhuri, J. B. (1997) *J. Chromatogr. A 766*, 109−119.
42. Fink, A. L., Oberg, K. A., and Seshadri, S. (1998) *Folding Des. 3*, 19−25.
43. Fink, A. L. (1995) *Annu. Rev. Biophys. Biomol. Struct. 24*, 495−522.
44. Palleros, D. R., Reid, K. L., McCarty, J. S., Walker, G. C., and Fink, A. L. (1992) *J. Biol. Chem. 267*, 5279−5285.
45. Uversky, V. N., Segel, D. J., Doniach, S., and Fink, A. L. (1998) *Proc. Natl. Acad. Sci. U.S.A. 95*, 5480−5483.
46. Bennett, M. J., Sclunegger, M. P., and Eisenberg, D. (1997) *Protein Sci. 4*, 2455−2468.
47. Silow, and Oliveberg, M. (1997) *Proc. Natl. Acad. Sci. U.S.A. 94*, 6084−6086.

BI991337K