Conformational landscape of cytochrome *c* folding studied by microsecond-resolved small-angle x-ray scattering

Shuji Akiyama*, Satoshi Takahashi*†‡, Tetsunari Kimura*, Koichiro Ishimori*, Isao Morishima*, Yukihiro Nishikawa§, and Tetsuro Fujisawa‡§

*Department of Molecular Engineering, Graduate School of Engineering, Kyoto University, Sakyo, Kyoto 606-8501, Japan; †Precursory Research for Embryonic Science and Technology, Japan Science and Technology Corporation, Kawaguchi, Saitama 332-0012, Japan; and §RIKEN Harima Institute/SPring-8, Structural Biochemistry Laboratory, Sayo, Hyogo 679-5148, Japan

To investigate protein folding dynamics in terms of compactness, we developed a continuous-flow mixing device to make small-angle x-ray scattering measurements with the time resolution of 160 μs and characterized the radius of gyration (R_g) of two folding intermediates of cytochrome *c* (cyt *c*). The early intermediate possesses $\approx$20 Å of R_g, which is smaller by $\approx$4 Å than that of the acid-unfolded state. The R_g of the later intermediate is $\approx$18 Å, which is close to that of the molten globule state. Considering the α-helix content (f_H) of the intermediates, we clarified the folding pathway of cyt *c* on the conformational landscape defined by R_g and f_H. Cyt *c* folding proceeds with a collapse around a specific region of the protein followed by a cooperative acquisition of secondary structures and compactness.

Proteins are unique heteropolymers that possess a remarkable property to fold quickly to compact and specific conformations. Interiors of proteins are densely packed with minimum void volumes (1), indicating the specific interresidue contacts that determine secondary and tertiary structures. The compactness is therefore an essential property of the folded conformations of proteins; however, the dynamics of compaction in the process of protein folding from the extended random-coil structures are still poorly understood. Two classical models of protein folding, the hydrophobic collapse (2) and the framework models (3), suppose that proteins acquire compactness in a distinct dynamical process separated from the secondary structure formations. In the hydrophobic collapse model (2), a nonspecific collapse of protein main chain is hypothesized to trigger the tertiary and secondary structure formations. In contrast, the framework model assumes that the initial formation of secondary structures urges the subsequent folding into compact conformations (3). Recent theoretical investigations, however, suggest that main-chain collapse and secondary structure formation are mostly concerted (4). Although the different equilibrium conformations of a certain protein indicate a linear correlation between the secondary structure content and compactness (5, 6), the relationship has not been confirmed directly for kinetic folding intermediates of proteins. Experimental investigations on the protein folding dynamics in terms of compactness are urgently needed to differentiate these models, that is, to understand how unfolded proteins kinetically explore for the native states on the conformational landscape defined by compactness and secondary structure content.

Cytochrome *c* (cyt *c*) is a globular protein of 104 aa, whose folding dynamics has been the subject of extensive experimental investigations (7–21). A heme group is covalently connected to the main chain (22) and surrounded by the three major helices called N-terminal residues 6–14), C-terminal (residues 87–102), and 60's helices (residues 60–69). The time-resolved circular dichroism (CD) measurement on the folding process of cyt *c* clarified the stepwise formation of these helices through two

folding intermediates having $\approx$500-μs and 7-ms lifetimes (7). Fluorescence from Trp-59, located in the vicinity of the heme in the folded conformation, is significantly quenched by the heme within 100 μs after the initiation of the folding (the nascent phase), suggesting the early collapse of the main-chain structure (12, 15). Therefore, although the secondary structure contents of the intermediates are characterized, the lack of quantitative information on the size and shape of the short-lived intermediates hinders the characterization of the folding pathway of cyt *c*.

Small-angle x-ray scattering (SAXS) is a powerful method for obtaining information on the size and shape of proteins in solution. The time-resolved SAXS techniques using stopped-flow or continuous-flow mixers have been successfully applied to several protein folding processes in the millisecond time domain (23–26). However, it is a frequent observation that proteins collapse rapidly within the dead time of these devices (24, 26). Even with the diffusive mixing device that first achieved the submillisecond time resolution (15), the early dynamics of the main-chain collapse could not be followed, partly because of its limited mixing efficiency. We have therefore developed a system for SAXS measurements equipped with a newly designed rapid mixing apparatus, which has a perfect mixing efficiency and $\approx$2 orders higher time resolution ($\approx$160 μs) than that of the previous continuous-flow mixers (23). Combined use of the developed device and a brilliant synchrotron x-ray source at SPring-8 (Harima, Japan; ref. 27) enables us to observe the refolding process of cyt *c* in the microsecond time domain and to identify the main-chain contraction processes.

Materials and Methods

Rapid Solution Mixer. The developed mixing apparatus for transient SAXS measurements is schematically illustrated in Fig. 1. The apparatus utilizes a T-shaped mixing point (Fig. 1 *Inset*) (18) constructed on a stainless plate (Machida Toolex, Tokyo). The collimated x-ray is irradiated to a spot along the observation channel. Time-resolved measurements are achieved by changing the distance between the mixing and observation points. Because of the small dimension (200 μm width) of the observation channel, the high photon flux from synchrotron x-ray source and the sensitive detection system (27) are indispensable for successful measurements. We confirmed that the SAXS profiles for
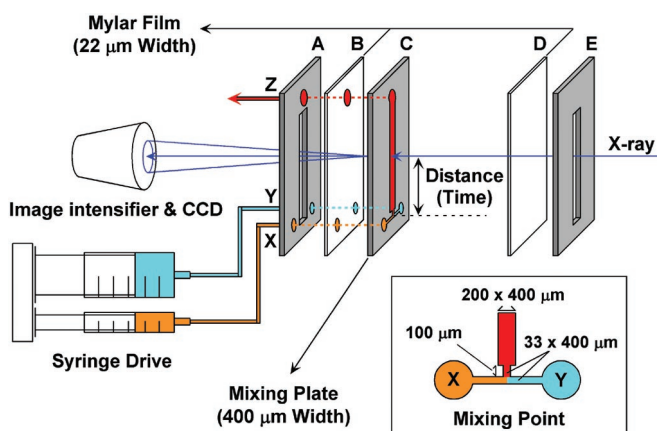
BIOPHYSICS

Fig. 1. Schematic drawing of the rapid solution mixer developed for the measurements of microsecond-resolved SAXS profiles. A stainless mixing plate (*C*) with 400-μm thickness is placed between two Mylar films (*B* and *D*) with 22-μm thickness, which are tightened between two metal plates (*A* and *E*). Two holes (*X* and *Y*) of the plate *A* and film *B* serve as the entrances of two solutions to be mixed. The denatured protein solution and refolding buffer are placed in the large (25 ml) and small syringes (5 ml), respectively, and are supplied continuously to a T-shaped mixing point of the plate *C* (*Inset*). The mixed solution flows inside the observation slit (200 μm width and 400 μm depth) and is wasted from an exit hole (*Z*). The incident x-ray (1.5 Å) was irradiated to a spot (400 $\times$ 400 μm) on the flow channel. The flow line speeds of the sample inside the channel are either 4.17 or 2.09 m·s^{-1}.

several proteins obtained by using this system were identical to that measured by using standard cells.

To evaluate the mixing efficiency of the mixer, we replaced the films B and D of the mixer (Fig. 1) with quartz windows and observed a pH jump discoloration reaction of a pH-sensitive dye (bromocresol purple, from pH 7.0 to 4.3) along the flow channel by using a microscope equipped with a digital camera. The completion point of the discoloration reaction ($\approx$300 μm from the center of the T-shaped mixing point) corresponds to $\approx$70 μs at the flow line speed of 4.17 m·s^{-1}, if we set the center of the mixing point to $t = 0$. To avoid the overlaps of the incident x-ray beam (400 $\times$ 400 μm) with the mixing region and the slit E (Fig. 1), we set the first observable point to 160 μs. Thus, the mixing time and the observation dead time of the mixer are 70 and 160 μs, respectively.

SAXS Measurements. All SAXS measurements were performed at beamline BL45XU of SPring-8 (27). The intensity of the incident x-ray was monitored by an ionization chamber installed before the mixing devices. Scattering profiles were collected by a charge-coupled device equipped with an x-ray image intensifier (XR-II; ref. 27), whose linearity was confirmed in a certain experimental condition (28). The refolding kinetics for the time domain from 160 μs to 6 ms was monitored by using the mixing device. The x-ray wavelength and the camera length were 1.5 Å and 585 mm, respectively. The data collection time was 2 s. The refolding kinetics for the time domain after 22 ms was monitored by using a stopped-flow mixer with a mixing volume ratio of 1:6 and a capillary cell with 2 mm of path length (Unisoku, Osaka). The x-ray wavelength and the camera length were 1.0 Å and 575 mm, respectively. The scattering signals were collected by the charge-coupled device intensified by using XR-II at 22 ms of time interval and averaged for at least 30 shots.

The static SAXS profiles of the acid-unfolded state (pH 2.0), the A state (0.5 M NaCl, pH 2.0), and the native state (pH 4.5) were measured by using a standard cell with 3-mm optical path length. The x-ray wavelength and the camera length were 1.0 Å and 575 mm, respectively. The R_g values for these states de-

scribed in the text are the extrapolated values to the infinite dilution.

Sample Conditions. Horse heart cyt *c* (type IV, Sigma), treated with ferricyanide and purified by using a cation exchange column, was repeatedly dialyzed against a large volume of 20 mM HCl at 4°C and concentrated to 28.9 mg·ml^{-1}. The solutions containing unfolded cyt *c* were equilibrated for at least 3 h before the refolding measurements and passed through a 0.22-μm filter. Refolding of acid-denatured cyt *c* at pH 2.0 was initiated by a 1.2-fold dilution with 500 mM acetate at pH 5.5 to the final pH of 4.5 and to the final protein concentration of 24.1 mg·ml^{-1}. The acid-unfolded state at the protein concentration of 13.1 mg·ml^{-1} in 16.4 mM HCl at pH 2.0 was prepared in the same manner, whose refolding was initiated by a 1.2-fold dilution with 500 mM acetate at pH 5.3 to the final pH of 4.5 and to the final protein concentration of 10.9 mg·ml^{-1}. All kinetic and static measurements were performed at 27°C.

Data Analysis. The sample and buffer profiles were collected consecutively at the same flow speed to ensure the identity in the experimental conditions. After the subtraction of the buffer profile, each profile was scaled by the ion-chamber current and the protein concentration. The radius of gyration was determined from the normalized profiles by using the Guinier analysis based on the following approximation:

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

where $S = (2 \sin\theta)/\lambda$; 2θ is the scattering angle; λ is the wavelength of the incident x-ray; R_g is the radius of gyration; and $I(S)$ is the scattering intensity at S. The scattering profiles from 0.005 to 0.01 Å^{-1} in S were selected as the linear fitting region of the Guinier analysis.

Singular value decomposition (SVD) and global fitting analysis were carried out by using SPECFIT (version 2.12, Spectrum Software Associates, Marlborough, MA). Pair distribution functions were calculated by using the GNOM package (29). The S range used for the calculation was from 0.0056 to 0.0367 Å^{-1} following the procedure of Flanagan *et al.* (30).

Results and Discussion

Stepwise Chain Condensation During Cyt *c* Folding. We followed the refolding process of cyt *c* after the rapid pH jump from the acid-denatured state at pH 2.0 to the native state at pH 4.5, because little misligation of His-33 to the heme iron is considered to occur under this condition that simplifies the folding mechanism of cyt *c* (7, 14). As shown in Fig. 2, R_g (blue trace) determined from Guinier analysis of the time-resolved scattering profiles indicates a biphasic decrease with the rates of 2,000 $\pm$ 800 and 68 $\pm$ 7 s^{-1}. The observed rates were in agreement with those obtained from the kinetic fluorescence and CD measurements in the same solution condition, considering the differences in temperature and the detection method (7, 12). Extrapolating the double exponential fitting in Fig. 2 to $t = 0$ leads to R_g of 21 $\pm$ 1 Å that is smaller than R_g of the initial acid-unfolded state (24.3 Å) estimated separately (see Fig. 2 legend). The observed $\approx$3-Å contraction of R_g during the nascent phase is synchronous with the significant quenching (16,900 s^{-1}) of fluorescence from Trp-59 (12). We conclude that the chain condensation in cyt *c* folding proceeds in three kinetic phases: the $\approx$3-Å collapse within the observation dead time ($<$160 μs) and the two phases observed in the submillisecond and millisecond time domains.

The presence of two kinetic phases in the observed time domain ($>$160 μs) is further supported from the analysis of scattering profiles in the intermediate angle region, which is best enhanced by the Kratky plot as shown in Fig. 3*A*. At the first
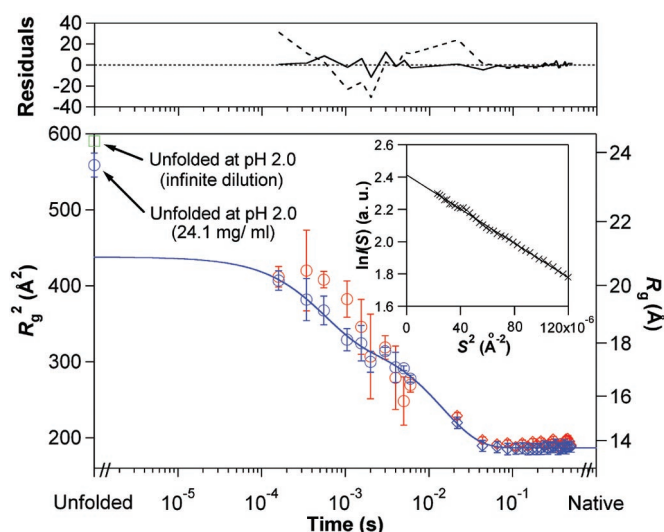
Fig. 2. Time-dependent changes in radius of gyration (R_g) during cyt *c* folding. The circles and diamonds represent R_g determined from the Guinier analyses of the scattering profiles acquired by the developed system and a stopped-flow apparatus, respectively. The blue and red points correspond to the results obtained for the protein concentrations at 24.1 and 10.9 mg·ml^{-1}, respectively. The bars denote the standard deviations from the independent measurements repeated several times for each condition. The blue line shows the double exponential fitting on the whole kinetic data obtained at 24.1 mg·ml^{-1}, which gives the rates of 2,000 and 68 s^{-1}. (*Upper*) The solid and dashed lines represent the residuals of double- and single-exponential fits to the overall data at 24.1 mg·ml^{-1}. Because of the significant interparticle scattering at the high concentration of cyt *c* (24.1 mg·ml^{-1}), it was impossible to determine R_g of the acid-unfolded state from the Guinier analysis (not shown). We therefore estimated R_g of the unfolded state by using a Debye function of a random coil chain (31). The estimated R_g was 23.6 ± 0.3 Å (blue circle), and was slightly smaller than the values determined by using the Guinier analysis of the acid-unfolded state by us (24.3 Å, green box) and by others (24.2 ± 0.3 Å) (32). (*Inset*) An example of the Guinier plot at 160 μs. The Guinier analysis of the datum in the fitting range from 0.005 to 0.01 Å^{-1} in *S* gives 20.1 Å of R_g.

observable point of 160 μs (Fig. 3*A*, red circle) after the initiation of folding, we detected a distinct peak around 0.017 Å^{-1} in *S*, which indicates the accumulation of compact species within the observation dead time. Subsequently, the peak intensity dramatically grows with a gradual shift to the high-angle region (0.02 Å^{-1} in *S*). The scattering intensity integrated from 0.017 to 0.020 Å^{-1} in *S* indicates two distinct phases (Fig. 3*A Inset*), which can be fitted with a double exponential function using the same rate constants estimated from the time dependency of R_g^2 (Fig. 2). These results suggest the stepwise globularization of cyt *c* through the two kinetic intermediates.

We previously interpreted that cyt *c* folding is described by the sequential kinetic scheme with two intermediates (U → I → II → N) from the time-resolved CD measurements (7). Although one could argue for other folding scenarios such as the parallel folding scheme and the involvement of a dimeric intermediate, these scenarios seem unlikely as described below. If we assume the parallel formations of the native state for both 2,000 s^{-1} and 68 s^{-1} phases (U → I$_1$ → N and U → I$_2$ → N), a significant amount of the native state should accumulate within 1 ms. However, our results from the hydrogen/deuterium (H/D) exchange study indicated the absence of the native conformation at 700 μs after the pH jump (unpublished data). Accumulation of the dimeric intermediate, which is observed during cyt *c* folding initiated from the guanidine-denatured state (9), is also unlikely. First, the zero angle scattering intensities of all of the kinetic profiles are nearly constant (not shown), indicating the
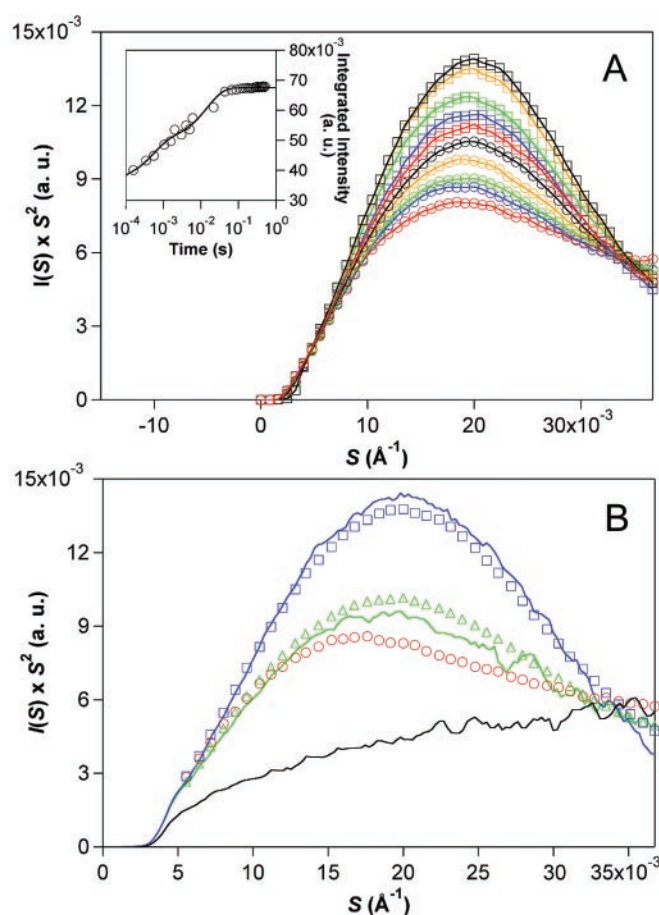
Fig. 3. Time-resolved changes in the scattering profiles during cyt *c* folding initiated by a pH jump from 2.0 to 4.5. (*A*) The scattering profiles presented as the Kratky plots. The protein concentration is 24.1 mg·ml^{-1}. For clarity, only 10 of the 33 observed profiles are displayed (red circle, 160 μs; blue circle, 340 μs; green circle, 550 μs; orange circle, 1 ms; black circle, 3 ms; red square, 4 ms; blue square, 6 ms; green square, 22 ms; orange square, 44 ms; black square, 500 ms). Note that the scattering profile at the earliest detectable point (160 μs) is distinct from that of the initial unfolded state indicated in *B*. (*Inset*) The circles illustrate the time-dependent changes in the intensity integrated from 0.017 to 0.020 Å^{-1} in *S*, and the solid line denotes double exponential fit to the data using the rate constants (2,000 and 68 s^{-1}) estimated from the time dependency of R_g^2 (Fig. 2). (*B*) Reconstructed scattering profiles for the three kinetic components observed during the folding of cyt *c*. The time-resolved Kratky profiles shown in *A* were analyzed by using the SVD method and globally fitted to the sequential folding scheme with four components, U ↔ I → II → N (see text for details). Red circles, green triangles, and blue squares represent scattering profiles for components I, II, and N, respectively. Scattering profiles for the samples in equilibrium conditions are also presented for comparison; black line, green line, and blue line represent the acid-denatured state (1.32 mg·ml^{-1} at pH 2.0), the MG state (the A state, 1.49 mg·ml^{-1} in 0.5 M of NaCl at pH 2.0), and the native state (1.73 mg·ml^{-1} at pH 4.5), respectively. The R_g of the respective species estimated from Guinier analysis are as follows: component I, 20.5 Å; component II, 17.7 Å; component N, 13.9 Å; the acid-denatured state at infinite dilution, 24.3 Å; the A state at infinite dilution, 18.3 Å; the native state at infinite dilution, 13.8 Å.

constant molecular weight of the observed intermediates (33). Second, the refolding experiments at the different protein concentrations (24.1 and 10.9 mg·ml^{-1}) gave essentially the same biphasic decrease in R_g, although the R_g values at the smaller protein concentration are larger because of the smaller interparticle scattering effect (Fig. 2). These observations are consistent with our proposal that the pH jump refolding process of cyt *c* proceeds through the sequential folding pathway with two monomeric intermediates (7).

Akiyama *et al.*

BIOPHYSICS
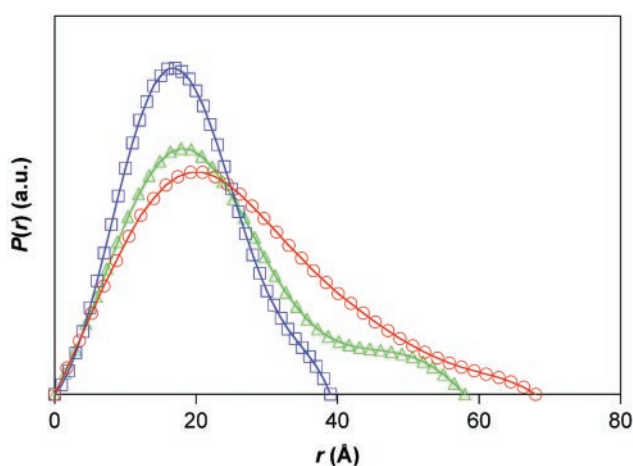
Fig. 4. Pair distribution functions, $P(r)$, of the observed states. Circles, triangles, and squares correspond to $P(r)$s for intermediate I (component I), intermediate II (component II), and the native state (component N), respectively.

Table 1. Structural parameters calculated for the folding intermediates and the native state of cyt c

| Scheme	Component I		Component II		Component N	
	R_g, Å	D_{max}, Å	R_g, Å	D_{max}, Å	R_g, Å	D_{max}, Å
U ↔ I → II → N*	20.5	66	17.7	58	13.9	39
I → II → N†	21.1	68	17.8	58	13.9	39

*The time-resolved scattering profiles presented in Fig. 3*A* were globally analyzed using the four-state sequential scheme. The folding rate from components U to I and the unfolding rate from components I to U were assumed to be 12,500 and 2,000 s^{-1}, respectively, according to Shastry *et al* (13). The folding rates from components I to II and from components II to N were calculated to be 2,400 and 68 s^{-1}, respectively, from the apparent rates estimated from the double exponential fitting on R_g^2 (Fig. 2). The scattering profile of component U was constrained to that of the static acid-denatured state measured separately.

†The same scattering profiles in Fig. 3*A* were globally fitted using the three-state sequential scheme. The folding rates from components I to II and from components II to N were adjusted to 2,000 and 68 s^{-1}, respectively, which were the apparent rates from the double exponential fitting on R_g^2 (Fig. 2). See the text for the details of the model calculations.

Geometric Shapes of the Two Intermediates. To extract information on the size and shape of the two intermediates, we analyzed the time-resolved Kratky plots using SVD and global fitting methods (34). The SVD analysis revealed the necessity of at least three independent components to reconstruct the overall data set adequately. Because the previous kinetic study indicates the involvement of a small fraction of the unfolded state at several hundreds of a microsecond after the initiation of folding (13), we used a sequential folding scheme with four states (U ↔ I → II → N), adopting the folding rate from component U to I (12,500 s^{-1}) and the unfolding rate from component I to U (2,000 s^{-1}) from the previous fluorescence measurement (13). The folding rates from component I to II and from component II to N were calculated to be 2,400 and 68 s^{-1}, respectively, from the apparent folding rates determined from the double exponential fitting on R_g^2 (Fig. 2). Furthermore, we assumed that the scattering profile of component U is identical to that of the acid-denatured state measured independently by using a standard static cell. The global SVD data were fitted under the constraint of the above kinetic model. The decomposed profiles of the respective components are displayed in Fig. 3*B* together with those of the native and unfolded states. To further extract information on the conformation of each component, we calculated a pair distribution function, $P(r)$, using the decomposed profiles as shown in Fig. 4. Because $P(r)$ indicates the distributions of linear distances between every pair of two atoms, characteristic lengths of each component such as the most frequent r value and the largest r value (D_{max}) can be determined from $P(r)$, and are compiled in Table 1.

It should be noted that the geometric parameters determined from the above analysis depend little on the choice of the kinetic models. We also fitted the SVD-filtered data by using the minimum kinetic model without the unfolded state (I → II → N), and listed the obtained parameters in Table 1. It is apparent that although the variables for intermediate I are slightly different in the two models, the differences are small and do not change our main conclusion described below. The invariance is caused by the small contribution of the unfolded component in our kinetic data. The estimated parameters from the three- and four-state analyses correspond to the upper and lower limits, respectively, for the actual values of intermediate I.

The conformational transition from the acid-unfolded state (24.3 Å) to intermediate I (20.5 Å) accompanies ≈4-Å reduction of R_g (≈40% of overall change). The Kratky plot of intermediate

I differs from those of typical random coils as evidenced by the weak peak (Fig. 3*B*), which is interpreted as the organization of a compact cluster associated with tertiary contacts between separated segments along the sequence (35, 36). For example, the α-helix formation of poly(L-lysine) from the random coil state accompanies little alternation in its SAXS profile (36), indicating that formation of the local secondary structure without tertiary contacts gives no peak in Kratky plot. The $P(r)$ of intermediate I shown in Fig. 4 also suggests a collapse, because the most frequent r value of the intermediate I (≈21 Å) is only ≈4 Å larger than that of the native state (≈17 Å). At the same time, the maximum r value (D_{max}) of intermediate I (≈66 Å) was still ≈27 Å larger relative to that of the native state (≈39 Å), indicating the presence of an extended region. These results suggest that the overall structure of intermediate I consists of the clustered region, stabilized by tertiary contacts, and the extended region.

An apparent peak in the Kratky plot of intermediate II (Fig. 3*B*) suggests its highly globular and compact conformation, which is consistent with the significantly reduced R_g (17.7 Å). The well-structured feature of intermediate II is further supported by its $P(r)$ function that exhibits the native-like peak position with increased number of the pair length in 18–22 Å (Fig. 4). These observations show that intermediate II has a densely packed core. In the previous kinetic CD experiments (7), we suggested that intermediate II might be a molten globule (MG) state that possesses extensive α-helical structures (≈70% of native level). Interestingly, the scattering profile of intermediate II illustrated in Fig. 3*B* is almost identical to that of the static MG state of cyt c called the A state, which is stable at low pH with high salt concentration (10, 32). The H/D exchange study of the A state indicated that the three major helices are formed, and the remaining parts are unfolded (11). A small fraction of the extended structures in intermediate II, suggested from the low distribution of the pair length in 40–60 Å (Fig. 4), is consistent with the presence of the unfolded region in the A state (35). We suggest that intermediate II is the kinetic MG state, which possesses the well-structured core and a small fraction of the extended structure.

Conformational Landscape of Cyt c Folding. The current and previous observations allow us to characterize the collapsing dynamics of cyt c folding. The acid-unfolded protein converts to intermediate I within the observation dead time (<160 μs), resulting in ≈4-Å reduction of R_g. Intermediate I contains a

Akiyama *et al.*

www.manaraa.com
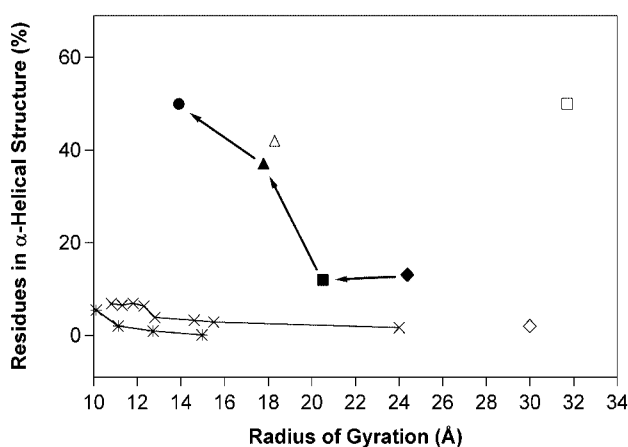
Fig. 5. Conformational landscape of cyt c defined by α-helical content (f_H) and radius of gyration (R_g). Filled symbols correspond to the kinetic intermediates of cyt c, whose R_g and f_H were identified in the current SAXS experiments and previous CD measurements (7), respectively. The acid-unfolded protein (◆) first indicates a moderate collapse to intermediate I (■), and subsequently develops both helical structures and compactness to form intermediate II (▲) and the native state (●). Note that the estimated secondary structure contents from CD spectra may involve a certain error. As stated previously (7), f_H estimated for the native state of cyt c from the CD spectrum (≈50%) does not coincide with that estimated from the crystal structure (≈45%) (22). The open symbols show the other static states of cyt c; □, the H state in 60% methanol at pH 1.7 (37); △, the molten globule state (the A state) in 0.5 M NaCl at pH 2.0; ◇, the unfolded state in 4.4 M guanidine hydrochloride (8). Crosses and asterisks denote the dependence of the secondary structure contents on the radius of gyration calculated for 82-residue proteins (38) and 100-residue polyalanine (39), respectively, which are adopted from figures in refs. 38 and 39. We plotted the calculated values for the secondary structure contents including both α-helix and β-sheet as the f_H values.

small hydrophobic cluster and an extended region as evidenced by its $P(r)$ function, and ≈20% native helical content (7). The significant quenching of Trp-59 fluorescence in intermediate I (12, 16, 20) indicates that the heme and Trp-59 are involved in the cluster. The presence of the collapsed and extended regions in intermediate I suggests that the initial transition corresponds to the collapse around the specific region of the main chain and the heme, not to the nonspecific and uniform collapse. The second collapse to intermediate II occurs with a time constant of ≈500 μs, whose helix content (≈70% native) (7) and R_g (17.7 Å) are very similar to those of the static MG state. The similarity suggests that intermediate II has the well-structured core including the three major helices. The native structure (13.8 Å) is formed from intermediate II upon the final collapse that takes ≈15 ms. Thus, the folding dynamics of cyt c comprises three discrete steps, each of which likely involves the collapse around the specific region of cyt c.

To visualize the folding pathway of cyt c, we present the observed R_g values of the intermediates together with the α-helical contents (f_H) in a two-dimensional conformational landscape defined by R_g and f_H (Fig. 5). Longitudinal and lateral movements in this plot indicate changes in the helix contents and compactness, respectively. The acid-unfolded protein (Fig. 5, filled diamond) possesses enlarged R_g and low f_H. Its conversion to the native state (Fig. 5, filled circle) proceeds in the R_g–f_H plot through the two intermediates (Fig. 5, filled square and triangle for intermediate I and II, respectively) that are distinct from the following extreme intermediates expected for the hydrophobic collapse and framework models. The maximally collapsed conformations of model polypeptides (38, 39) with almost the same chain length as cyt c will be treated as the extreme intermediates for the hydrophobic collapse model, and appear in the lower left

corner of the R_g–f_H plot (Fig. 5, crosses and asterisks). In contrast, a methanol-denatured state of cyt c (37) called the H-state (Fig. 5, open square) possesses an expanded conformation with a pronounced amount of α-helix (more than 50%) that can be considered as the extreme case for the framework intermediate. The points for the kinetic intermediates significantly differ from these examples and indicate that neither the extreme hydrophobic collapse nor framework models describe the folding of cyt c.

We previously concluded that the initial step of cyt c folding is described by the hydrophobic collapse model (7), because the nascent phase involves the significant quenching of Trp-59 fluorescence (12) without the marked development of the far-UV CD signal (7). The quantitative estimation from the current SAXS results shows the intermediate collapse (≈40% of total change) and demonstrates that the folding mechanism of cyt c does not follow the extreme hydrophobic collapse model. Because intermediate I likely possesses the collapsed cluster around the specific region of the protein, we conclude that cyt c folding proceeds with the specific collapse followed by the cooperative acquisition of secondary structures and compactness.

The determined pathway of cyt c folding in the R_g–f_H plot is consistent with the calculation of the folding landscape for an α-helical protein by Guo *et al.* (4). They demonstrated that the folding pathway, defined by a valley in the calculated energy landscape, extends almost diagonally from the unfolded state to the folded state in the R_g–f_H plot. The current results are therefore experimental evidence for the concerted acquisition of the compactness and the secondary structures during the actual folding dynamics of helical proteins. It is interesting to note that the initial collapse causes a slight bend in the overall folding pathway on the R_g–f_H plot. The slightly bent pathway may be essential for the topology of cyt c fold, because it is predicted that the folding pathway in the R_g–f_H plot should depend on the folded conformation of a protein (4).

The stepwise and specific collapses observed in cyt c folding deserve further discussion in terms of the fastness of protein folding. Fast-folding proteins, including cyt c, are considered to possess a smooth energy landscape and to satisfy the principle of minimum frustration (40, 41). We suggest that the smoothness might be partly explained by avoidance of the nonspecifically collapsed conformation during cyt c folding, because the frustration tends to become larger in the more compact conformations that can increase the nonnative as well as the native contacts (42, 43). Interestingly, Thirumalai *et al.* (44, 45) reported the time scale for the specific collapse of small proteins (≈3 ms), which is in the same order as those of the second and third collapses of cyt c folding.

In summary, the submillisecond time-resolved SAXS measurements reported here provided detailed information on the time evolution of the shape and compactness of the main chain during cyt c folding. The conformational landscape of protein folding was experimentally characterized in the two-dimensional space defined by R_g and f_H. Cyt c folding proceeds with the specific collapse followed by the concurrent acquisition of secondary structures and compactness, which is different from the extreme hydrophobic collapse or framework models. Methodical time-resolved SAXS experiments on other proteins with different topologies and secondary structures are required to examine the relationship between folding mechanism and overall topology.

Akiyama *et al.*

1. Chothia, C. (1975) *Nature (London)* **254,** 304–308.
2. Dill, K. A. (1985) *Biochemistry* **24,** 1501–1509.
3. Kim, P. S. & Baldwin, R. L. (1982) *Annu. Rev. Biochem.* **51,** 459–489.
4. Guo, Z., Brooks, C. L., III & Boczko, E. M. (1997) *Proc. Natl. Acad. Sci. USA* **94,** 10161–10166.
5. Kataoka, M., Nishii, I., Fujisawa, T., Ueki, T., Tokunaga, F. & Goto, Y. (1995) *J. Mol. Biol.* **249,** 215–228.
6. Nishii, I., Kataoka, M., Tokunaga, F. & Goto, Y. (1994) *Biochemistry* **33,** 4903–4909.
7. Akiyama, S., Takahashi, S., Ishimori, K. & Morishima, I. (2000) *Nat. Struct. Biol.* **7,** 514–520.
8. Segel, D. J., Fink, A. L., Hodgson, K. O. & Doniach, S. (1998) *Biochemistry* **37,** 12443–12451.
9. Segel, D. J., Eliezer, D., Uversky, V., Fink, A. L., Hodgson, K. O. & Doniach, S. (1999) *Biochemistry* **38,** 15352–15359.
10. Ohgushi, M. & Wada, A. (1983) *FEBS Lett.* **164,** 21–24.
11. Jeng, M. F., Englander, S. W., Elöve, G. A., Wand, A. J. & Roder, H. (1990) *Biochemistry* **29,** 10433–10437.
12. Shastry, M. C. R. & Roder, H. (1998) *Nat. Struct. Biol.* **5,** 385–392.
13. Shastry, M. C. R., Sauder, J. M. & Roder, H. (1998) *Acc. Chem. Res.* **31,** 717–725.
14. Yeh, S. R., Han, S. & Rousseau, D. L. (1998) *Acc. Chem. Res.* **31,** 727–736.
15. Pollack, L., Tate, M. W., Darnton, N. C., Knight, J. B., Gruner, S. M., Eaton, W. A. & Austin, R. H. (1999) *Proc. Natl. Acad. Sci. USA* **96,** 10115–10117.
16. Hagen, S. J. & Eaton, W. A. (2000) *J. Mol. Biol.* **301,** 1019–1027.
17. Sosnick, T. R., Mayne, L., Hiller, R. & Englander, S. W. (1994) *Nat. Struct. Biol.* **1,** 149–156.
18. Takahashi, S., Yeh, S. R., Das, T. K., Chan, C. K., Gottfried, D. S. & Rousseau, D. L. (1997) *Nat. Struct. Biol.* **4,** 44–50.
19. Bai, Y., Sosnick, T. R., Mayne, L. & Englander, S. W. (1995) *Science* **269,** 192–197.
20. Chan, C. K., Hu, Y., Takahashi, S., Rousseau, D. L., Eaton, W. A. & Hofrichter, J. (1997) *Proc. Natl. Acad. Sci. USA* **94,** 1779–1784.
21. Jones, C. M., Henry, E. R., Hu, Y., Chan, C. K., Luck, S. D., Bhuyan, A., Roder, H., Hofrichter, J. & Eaton, W. A. (1993) *Proc. Natl. Acad. Sci. USA* **90,** 11860–11864.
22. Bushnell, G. W., Louie, G. V. & Brayer, G. D. (1990) *J. Mol. Biol.* **214,** 585–595.
23. Segel, D. J., Bachmann, A., Hofrichter, J., Hodgson, K. O., Doniach, S. & Kiefhaber, T. (1999) *J. Mol. Biol.* **288,** 489–499.
24. Arai, M., Ikura, T., Semisotnov, G. V., Kihara, H., Amemiya, Y. & Kuwajima, K. (1998) *J. Mol. Biol.* **275,** 149–162.
25. Plaxco, K. W., Millett, I. S., Segel, D. J., Doniach, S. & Baker, D. (1999) *Nat. Struct. Biol.* **6,** 554–556.
26. Eliezer, D., Jennings, P. A., Wright, P. E., Doniach, S., Hodgson, K. O. & Tsuruta, H. (1995) *Science* **270,** 487–488.
27. Fujisawa, T., Inoue, K., Oka, T., Iwamoto, H., Uruga, T., Kumasaka, T., Inoko, Y., Yagi, N., Yamamoto, M. & Ueki, T. (2000) *J. Appl. Crystallogr.* **33,** 797–800.
28. Fujisawa, T., Kostyukova, A. & Maeda, Y. (2001) *FEBS Lett.* **498,** 67–71.
29. Semenyuk, A. V. & Svergun, D. I. (1991) *J. Appl. Crystallogr.* **24,** 537–540.
30. Flanagan, J. M., Kataoka, M., Fujisawa, T. & Engelman, D. M. (1993) *Biochemistry* **32,** 10359–10370.
31. Calmettes, P., Durand, D., Desmadril, M., Minard, P., Receveur, V. & Smith, J. C. (1994) *Biophys. Chem.* **53,** 105–114.
32. Kataoka, M., Hagihara, Y., Mihara, K. & Goto, Y. (1993) *J. Mol. Biol.* **229,** 591–596.
33. Glatter, O. & Kratky, O. (1982) *Small-Angle X-Ray Scattering* (Academic, New York).
34. Goldbeck, R. A. & Kliger, D. S. (1993) *Methods Enzymol.* **226,** 147–177.
35. Doniach, S., Bascle, J., Garel, T. & Orland, H. (1995) *J. Mol. Biol.* **254,** 960–967.
36. Semisotnov, G. V., Kihara, H., Kotova, N. V., Kimura, K., Amemiya, Y., Wakabayashi, K., Serdyuk, I. N., Timchenko, A. A., Chiba, K., Nikaido, K., *et al*. (1996) *J. Mol. Biol.* **262,** 559–574.
37. Kamatari, Y. O., Konno, T., Kataoka, M. & Akasaka, K. (1996) *J. Mol. Biol.* **259,** 512–523.
38. Hunt, N. G., Gregoret, L. M. & Cohen, F. E. (1994) *J. Mol. Biol.* **241,** 214–225.
39. Yee, D. P., Chan, H. S., Havel, T. F. & Dill, K. A. (1994) *J. Mol. Biol.* **241,** 557–573.
40. Go, N. (1983) *Annu. Rev. Biophys. Bioeng.* **12,** 183–210.
41. Bryngelson, J. D., Onuchic, J. N., Socci, N. D. & Wolynes, P. G. (1995) *Proteins Struct. Funct. Genet.* **21,** 167–195.
42. Gutin, A. M., Abkevich, V. I. & Shakhnovich, E. I. (1995) *Biochemistry* **34,** 3066–3076.
43. Hardin, C., Luthey-Schulten, Z. & Wolynes, P. G. (1999) *Proteins Struct. Funct. Genet.* **34,** 281–294.
44. Thirumalai, D. (1995) *J. Phys. I (France)* **5,** 1457–1467.
45. Thirumalai, D. & Klimov, D. K. (1999) *Curr. Opin. Struct. Biol.* **9,** 197–207.

Akiyama *et al.*