Dehydration of main-chain amides in the final folding step of single-chain monellin revealed by time-resolved infrared spectroscopy

Tetsunari Kimura*[†‡], Akio Maeda*, Shingo Nishiguchi*, Koichiro Ishimori[†§], Isao Morishima[†], Takashi Konno[¶], Yuji Goto*, and Satoshi Takahashi*[‖]**

*Institute for Protein Research, Osaka University, Suita, Osaka 565-0871, Japan; [†]Department of Molecular Engineering, Graduate School of Engineering, Kyoto University, Nishikyo, Kyoto 615-8510, Japan; [¶]Department of Molecular Physiology and Biophysics, Faculty of Medical Science, University of Fukui, Matsuoka, Yoshida, Fukui 910-1193, Japan; and [‖]Core Research for Evolutional Science and Technology, Japan Science and Technology Agency, Kawaguchi 332-0012, Japan

Edited by Robert L. Baldwin, Stanford University Medical Center, Stanford, CA, and approved July 10, 2008 (received for review February 10, 2008)

Kinetic IR spectroscopy was used to reveal β-sheet formation and water expulsion in the folding of single-chain monellin (SMN) composed of a five-stranded β-sheet and an α-helix. The time-resolved IR spectra between 100 μs and 10 s were analyzed based on two consecutive intermediates, I_1 and I_2, appearing within 100 μs and with a time constant of ≈100 ms, respectively. The initial unfolded state showed broad amide I′ corresponded to a fluctuating conformation. In contrast, I_1 possessed a feature at 1,636 cm^{-1} for solvated helix and weak features assignable to turns, demonstrating the rapid formation of helix and turns. I_2 possessed a line for solvated helix at 1,637 cm^{-1} and major and minor lines for β-sheet at 1,625 and 1,680 cm^{-1}, respectively. The splitting of the major and minor lines is smaller than that of the native state, implying an incomplete formation of the β-sheet. Furthermore, both major and minor lines demonstrated a low-frequency shift compared to those of the native state, which was interpreted to be caused by hydration of the C═O group in the β-sheet. Together with the identification of solvated helix, the core domain of I_2 was interpreted as being hydrated. Finally, slow conversion of the water-penetrated core of I_2 to the dehydrated core of the native state was observed. We propose that both the expulsion of water, hydrogen-bonded to main-chain amides, and the completion of the secondary structure formation contribute to the energetic barrier of the rate-limiting step in SMN folding.

protein folding dynamics | β-sheet

Various dynamics of polypeptides observed in protein folding are related to the different aspects of interactions between protein and water (1–4). Hydrophobic side chains and water form a van der Waals interaction in unfolded polypeptides (5, 6). The water molecules around hydrophobic residues are considered to be mostly replaced with other hydrophobic residues in the collapse phase of protein folding (7, 8), which occurs within several hundreds of microseconds (9–13). In contrast, main-chain amides and hydrophilic side chains in the unfolded polypeptides possess much stronger electrostatic and hydrogen-bonding interactions with water (5). It has been suggested that the dehydration of main-chain amides, required for the burial of main chains in the native proteins, should burden proteins with a significant energetic barrier for folding (5, 14). Furthermore, a partial dehydration of amides by side chains is suggested to regulate the propensity for secondary structure formation (15, 16). These suggestions might be related to the search phase of protein folding, which occurs after the collapse in the time domain from milliseconds to seconds and involves the formation of correct secondary and tertiary structures. A recent kinetic IR study on the collapsed folding intermediates of apomyoglobin (apoMb) identified the solvated helices that persist until the final folding phase and demonstrated the difficulty of detaching water around main-chain amides in an α-helical protein (14). In this study, we observed the folding of a β-sheet protein, single-chain monellin (SMN), using kinetic IR spectroscopy to examine whether the slow dehydration of main-chain amides is still observable or not and to establish the roles of the amide hydration in protein folding.

IR spectroscopy has emerged as a powerful method for structural characterization of proteins during folding in real time (17, 18). The method can monitor amide I mode, mainly attributed to C═O stretching vibration of amides, whose frequency is firstly modulated by changes in the intrinsic force constants of C═O caused by environmental effects (19–21). The frequency is further modulated by the transient dipolar coupling between C═O groups (19–21). In the case of antiparallel β-sheets, coupling, mainly between the nearest C═O groups in the adjacent strands, splits amide I into major and minor peaks at ≈1,630 and ≈1,680 cm^{-1}, respectively (22, 23). Thus, the method can differentiate subtle changes in β-sheets by examining the splitting between the major and minor peaks and the shift in the intrinsic frequency. IR spectra are not subject to motional averaging for flexible states and are capable of characterizing transient structures by combining with various kinetic methods including solution mixing (18, 24). Furthermore, information on the hydration structures of polypeptides can be obtained because hydrogen bonds on C═O by water cause the low-frequency shift of amide I (25, 26). Thus, the method is suited for the investigation of secondary structure formation and the action of water in the folding of β-sheet proteins (27).

SMN is a single-domain protein with 94 residues (28) and is composed of a five-stranded β-sheet and an α-helix attached on one side of the sheet (29, 30). In our previous investigation, the folding dynamics of SMN initiated by a pH jump from 13.0 to 9.4 was characterized by small-angle x-ray scattering (SAXS), CD, and intrinsic and extrinsic fluorescence spectroscopies (11). A significant collapse was observed within 300 μs after the pH jump, leading to a conformation with a small amount of secondary structures but with an overall oblate shape. Subsequently, the stepwise formation of secondary and tertiary structures was detected in two phases with time constants of 14 ms and 1.2 s. We proposed that SMN folds along the sequential scheme with two on-pathway intermediates (I_1 and I_2) as in Scheme 1.

Author contributions: T. Kimura, K.I., I.M., Y.G., and S.T. designed research; T. Kimura, A.M., and K.I. performed research; S.N. and T. Konno contributed new reagents/analytic tools; T. Kimura and S.T. analyzed data; and T. Kimura and S.T. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[‡]Present address: Beckman Institute, California Institute of Technology, Pasadena, CA 91125.

[§]Present address: Division of Chemistry, Graduate School of Science, Hokkaido University, Sapporo 060-0810, Japan.

**To whom correspondence should be addressed. E-mail: st@protein.osaka-u.ac.jp.

This article contains supporting information online at www.pnas.org/cgi/content/full/0801316105/DCSupplemental.

BIOPHYSICS

$$U \rightleftarrows I_1 \rightleftarrows I_2 \rightleftarrows N$$

Scheme 1.

I_1 has a native-like oblate shape, while it is slightly expanded relative to the native state (N). The small changes in the CD spectrum at 300 μs compared with the alkaline unfolded state (U) suggested that stable secondary structures are not formed in I_1. In I_2, the protein possesses an overall dimension indistinguishable from N and a moderate amount of secondary structures. We proposed that the formation of β-turns and the nonspecific hydrophobic collapse are necessary to explain the rapid formation of the oblate shape within 300 μs.

In this study, we characterized the equilibrium unfolding transition of SMN at alkaline pH based on various spectroscopic methods and investigated the refolding dynamics of SMN by kinetic IR spectroscopy. We developed a time-resolved IR spectrometer equipped with IR microscope and continuous-flow rapid solution mixers (14, 31) and characterized the kinetic folding process of apoMb in the time domain from 100 μs to 5 ms (14). In the current investigation we improved the device and extended the observation range to 10 s to cover the entire process of SMN folding. We assigned the hydrated β-sheet in the late intermediate of SMN. Furthermore, the shift in the intrinsic frequency of β-sheet caused by solvation is smaller than that of α-helix. We propose a general action of water in the folding of proteins by comparing these results with those for apoMb folding.

Results

Equilibrium Unfolding Transition of SMN. We first characterized the equilibrium unfolding transition of SMN by CD and fluorescence spectroscopies to obtain a basis for the analysis of IR spectra. At pH 7 and 9.4, the CD spectra show a major negative peak at 216 nm characteristic of β-sheets and a weak positive feature at 233 nm corresponding to aromatic residues [in supporting information (SI) Fig. S1*a*] (32). The plot of the ellipticities as a function of pH (Fig. 1*A*) shows three transitions occurring at around pH 10, 11, and 11.6. In the transition at pH 10, small changes are observed in the regions shorter than 210 nm and at around 235 nm. In the transition at pH 11, a significant reduction in the negative ellipticity at 216 nm is observed. In the second transition at pH 11.6, a further reduction in the negative ellipticities was observed. The pK_a values, the number of protons involved, and the spectra for the intermediates were estimated by a global-fitting analysis based on the three-step model (Fig. 1*B*, *SI Text*, and Table S1). Fluorescence titration was carried out by exciting at 290 nm. Fluorescence at this excitation mainly originates from Trp-4 and exhibits a maximum at 338 nm at pH 9.4 (Fig. S1*b*), which is blue-shifted from that of *N*-acetyltryptophan amide (352 nm at pH 9.4). In the first transition at pH 10, a gradual decrease in intensity was observed. In the second transition at pH 11, the red shift in the peak wavelength was observed. In the third transition at pH 11.6, a small increase in intensity was observed. The results of the global fitting analysis based on the three-step model are shown in Fig. 1*C* and Table S1. Thus, the CD and fluorescence data showed a three-step unfolding in the alkaline transition. We term the equilibrium intermediates as I_{eq2} and I_{eq1} in the order of appearance in the alkaline transition.

To examine the amide I spectra of SMN under different conditions, we observed the alkaline transitions by IR spectroscopy (Fig. 2*A*). The spectrum for the native (N) state at pD 9.4 overlaps the spectrum at pD 7.0 (Fig. S2*c*) and possesses a major peak and a minor peak assignable to the β-sheet at 1,629 and 1,687 cm^{-1}, respectively. The spectrum for the unfolded (U) state at pD 13.0 is broad and centered at 1641 cm^{-1} assigned to a disordered structure. The IR spectra of this alkaline U state and the unfolded state induced by 7 M ^{13}C-urea overlap each other (data not shown),

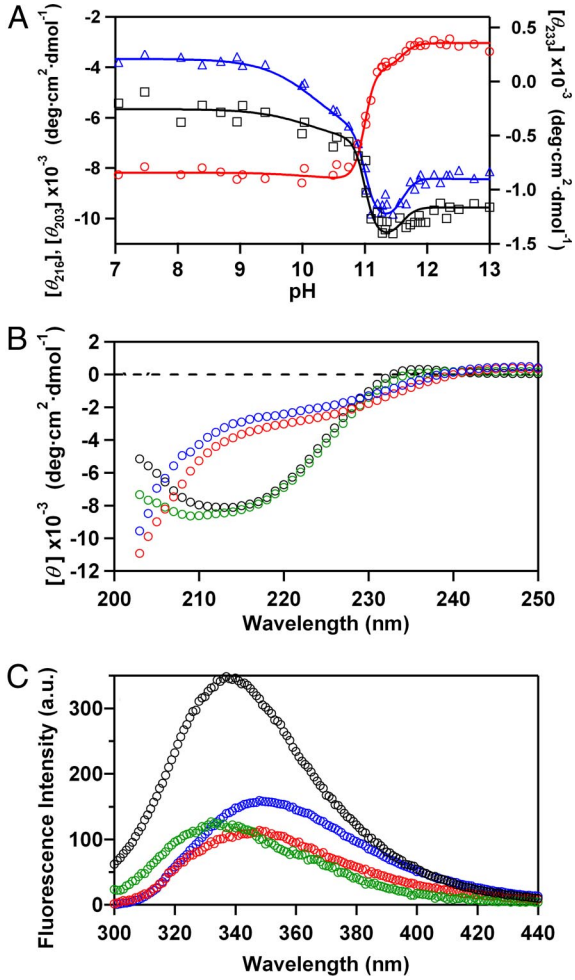

Fig. 1. The equilibrium unfolding transition of SMN at 20°C. (*A*) The unfolding transitions monitored by CD at 203 (black), 216 (red), and 233 nm (blue). Continuous lines are the fitted curves obtained by global fitting. (*B*) CD spectra of the N (black), I_{eq2} (green), I_{eq1} (red), and U (blue) states obtained by global fitting of the total of 37 spectra based on the three-step model. (*C*) Trp-fluorescence spectra of the N (black), I_{eq2} (green), I_{eq1} (red), and U (blue) states obtained by global fitting of the total of 25 spectra.

suggesting that the alkaline U state is highly disordered. A careful examination of the region around 1,680~1,700 cm^{-1} of the second derivative spectra shows three pD-induced transitions (Fig. S2*b*). First, a low-frequency shift of the negative peak is observed at around pD 10. The negative peak grows in amplitude at pD 11. Finally, the amplitude of the negative peak decreases at pD higher than 12. The pD values of the changes are similar to those obtained in the analysis of the Trp-4 fluorescence and CD results (Fig. 1). The transition curves observed at representative IR frequencies support three-step unfolding model (Fig. S2*c*) based on which we performed global fitting with fixed *n* values and obtained pK_a values at 10.3, 11.1, and 11.9 (Table S1). The calculated IR spectra for the intermediates and their second derivatives are given in Fig. 2 *B* and *C*, respectively. The peak frequencies assigned in the second derivatives are listed in Table 1 with the proposed assignments. Interestingly, both the major and minor peaks for β-sheet show low-frequency shifts for I_{eq2} and I_{eq1}. Furthermore, a feature for solvated helices at ≈1,637 cm^{-1} appears for the two intermediates. Thus, I_{eq2} and I_{eq1} are characterized by a low-frequency shift in the amide I′ lines for β-sheet and by the appearance of solvated helix.

Kinetic Folding Dynamics of SMN. To characterize the dynamics of SMN folding, we followed the folding after a pD jump from 13.0 to

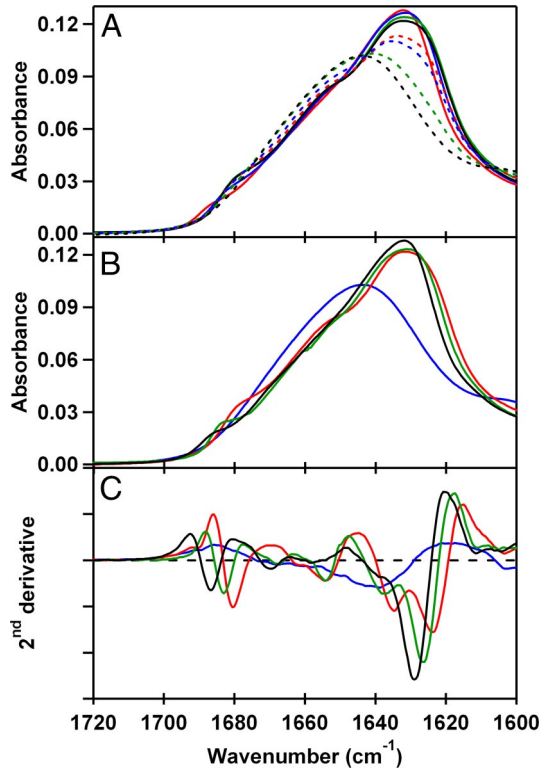

Fig. 2. The equilibrium unfolding transition of SMN at 20°C monitored by FTIR. (*A*) The alkaline-induced changes of amide I′ line. The presented spectra were selected from the total of 14 measured spectra and were obtained at pD 9.4 (red), 10.4 (blue), 11.2 (green), 11.5 (black), 11.8 (dotted red), 11.9 (dotted blue), 12.1 (dotted green), and 13.0 (dotted black). (*B*) The IR spectra of the N (black), I_{eq2} (green), I_{eq1} (red), and U (blue) states calculated by global fitting of the 14 spectra based on the three-step model. (*C*) The second derivative spectra of the N (black), I_{eq2} (green), I_{eq1} (red), and U (blue) states.

9.4 by time-resolved IR spectroscopy. By using three continuous-flow mixers, we obtained kinetic IR spectra for the folding of SMN in the time domain from 100 μs to 10 s (Fig. 3*A*). The absence of isosbestic points in the kinetic spectra indicates the presence of kinetic intermediates. The spectral changes were further visualized by means of second derivative spectra (Fig. S3*b*). A distinct change occurs within the dead time of 150 μs. Furthermore, major and minor peaks assignable to the β-sheet at 1,627 and 1,680 cm^{-1}, respectively, emerge in the time domain around 100 ms. Finally, both peaks show a high-frequency shift to the values corresponding to N in the time domain around 1 s. The changes in kinetic absorbance monitored at representative frequencies also demonstrated the changes occurring within 150 μs and in the two time domains around 100 ms and 1 s (Fig. S3*d*). The data suggest the formation of the initial and second intermediates within the dead time and with a time constant of ≈100 ms, respectively, and the formation of the native state with a time constant of ≈1 s.

To quantitatively analyze the kinetic data, we performed global fitting of the kinetic IR spectra based on a sequential folding model, in which the components are formed in the order of I, II, and III. We obtained two rate constants of 10.0 ± 0.7 s^{-1} and 0.91 ± 0.15 s^{-1} for the first and second kinetic transitions, respectively. The rate constants were slower than the reported rate constants (11). The difference is likely attributed to the solvent isotope effect and to the difference in the detection method because the separate folding experiment detected by fluorescence spectroscopy similarly showed the decelerated rate constants of the two phases in D_2O (32.4 and 0.99 s^{-1}) compared with those in H_2O (84.6 and 1.14 s^{-1}) (Fig. S4). The IR spectra of the three components (I–III) in the sequential model are given in Fig. 3*B* together with the spectrum of the U state. The spectrum of the N state is identical to that of component III. In contrast, the spectrum of component I that accumulates from 100 μs to 10 ms is distinct from that of U. This difference corresponds to the burst phase, which is similarly observed in the kinetic SAXS and CD measurements (11). These results indicate that at least two intermediates are required to explain the folding dynamics of SMN, which proceeds in three sequential phases: the conformational change within the observation dead time (150 μs) and the other two phases observed in the time domains around ≈100 ms and ≈1 s. Based on the kinetic scheme proposed previously, we assign components I and II to the kinetic intermediates I_1 and I_2, respectively.

To obtain detailed structural information for the kinetic intermediates, we calculated second derivative spectrum for each component (Fig. 3*C*). The frequencies and assignments of the peaks identified in the second derivative spectra were listed in Table 1 together with the rough estimates of secondary structure contents based on Gaussian fitting on the resolution-enhanced spectra (*SI Text* and Table S2). The transition from U to I_1 accompanies the shift of the main negative peak in second derivative IR spectra from 1,641 to 1,637 cm^{-1}. The second derivative spectrum of I_1 further reveals negative peaks at 1,675 and 1,666 cm^{-1}. The weak spectral features of these peaks suggest that the structures of I_1 are fluctuating. In contrast, the spectrum for I_2 possesses a significant peak at 1,627 cm^{-1}. Furthermore, the second derivative spectrum for I_2 reveals a shoulder at 1,637 cm^{-1} and negative peaks at 1,654, 1,666,

Table 1. Assignments of amide I′ components to secondary structures of SMN and their contents

	Wavenumber, cm^{-1} (Contents, %)						
	Static components				Kinetic components		
Secondary structure	U	I_{eq1}	I_{eq2}	N	I_1	I_2	N
Turn			1,673.5 (8)	1,670.0 (26)	1,675.3 (12)		1,669.5 (25)
Turn	1,663.7 (35)	1,661.8 (21)	1,667.6 (16)		1,665.6 (23)	1,665.6 (27)	
α-Helix		1,654.1 (6)	1,654.1 (9)	1,657.0 (10)		1,654.1 (5)	1,656.0 (11)
Unordered structures	1,641.0 (65)	(20)	(18)	(16)	1,646.3 (54)	(19)	(15)
Solvated helices		1,634.8 (13)	1,637.2 (8)	1,638.6 (6)	1,636.7 (11)	1,636.7 (12)	1,638.6 (6)
β-Sheet							
Major line		1,623.7 (35)	1,626.1 (37)	1,629.0 (40)		1,627.1 (34)	1,629.0 (41)
Minor line		1,680.1 (5)	1,683.0 (4)	1,686.9 (2)		1,679.6 (3)	1,686.8 (2)
Intrinsic frequency		1,651.9	1,654.6	1,658.0		1,653.3	1,657.9
Splitting		56.4	56.9	57.9		52.5	57.8

The negative peaks observed in the second derivatives are shown. The secondary structure contents were evaluated as relative areas based on Gaussian fitting of the resolution-enhanced spectra and should be considered rough estimates. The details of the calculation are explained in *SI Text* and Table S2.

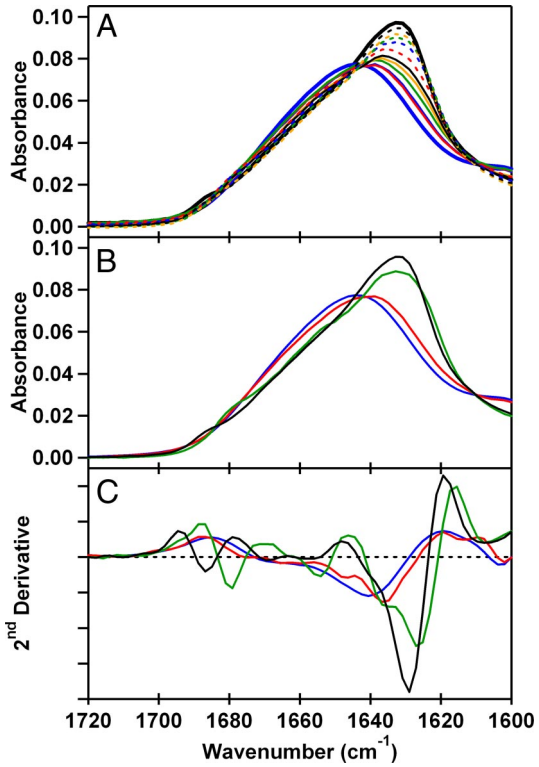

Fig. 3. The time courses of FTIR spectra during the folding of SMN. The kinetic folding of SMN was triggered by pD jump from 13.0 to 9.4 at 25°C. The final concentration of SMN was 1.4 mg·ml^{-1}. (*A*) The time-resolved amide I′ lines of FTIR spectra. Only 12 of 34 spectra are shown for clarity, corresponding to the initial unfolded state at pD 13.0 (bold blue), the kinetic data at 150 μs (red), 12 ms (blue), 50 ms (green), 90 ms (orange), 100 ms (black), 200 ms (dotted red), 600 ms (dotted blue), 800 ms (dotted green), 2 s (dotted orange), and 10 s (dotted black) after the pD jump, and the native state at pD 9.4 (bold black). (*B*) The IR spectra for the kinetic components observed during the folding of SMN. All of the time-resolved spectra were analyzed by using the global fitting procedure based on a sequential folding scheme with three components in addition to the initial unfolded state. The extracted spectra for the components I (red), II (green), and III (black) were assigned to I_1, I_2, and N, respectively. The FTIR spectrum of the unfolded state at pD 13.0 (blue) is shown for comparison. (*C*) The second derivative spectra of the U (blue), I_1 (red), I_2 (green), and N (black) states.

and 1,680 cm^{-1}. Finally, the N state possesses a fully developed peak at 1,629 cm^{-1}. However, the feature observed at 1,637 cm^{-1} in I_1 and I_2 becomes weaker, and the negative band at 1,680 cm^{-1} in I_2 shifts to 1,687 cm^{-1}. Thus, the IR spectrum for the initial kinetic intermediate, I_1, is characterized by the appearance of a solvated helix (11%) and by the absence of features assignable to the β-sheet. The second intermediate, I_2, possesses the features of a β-sheet (37%) as well as a solvated helix (12%). The final N state possesses the increased α-helix (11%), the decreased solvated helix (6%), and the major and minor lines for β-sheet (43%), both showing a high-frequency shift compared with those of I_2.

Discussion

We observed the kinetic folding of SMN upon a pD jump from 13.0 to 9.4 by time-resolved IR spectroscopy based on continuous-flow mixers. Observation of the entire folding process of proteins by kinetic IR spectroscopy is difficult because of the facile aggregation of proteins at the concentrations required for IR measurements. In our previous investigation we established that aggregate formation of SMN is completely suppressed by selecting the initial and final pH values of the kinetic experiments at 13 and 9.4, respectively, at a protein concentration of 4 mg/ml (11). We adopted that condition in the current investigation and successfully observed the kinetic IR spectra for the folding of SMN (*SI Text*). We identified a component assignable to solvated helix for the two kinetic intermediates, I_1 and I_2. Furthermore, we identified major and minor amide I′ lines characteristic of β-sheet for I_2. Interestingly, the two lines for β-sheet in I_2 had both shifted to lower frequencies than those of the native protein. The low-frequency shift of the lines for β-sheet was similarly observed in the equilibrium intermediates in the alkaline unfolding transition of SMN. As we will explain below, the detailed spectroscopic characterization of the equilibrium states enabled us to interpret the low-frequency shift of the amide I′ lines to solvated β-sheet. The identification of the lines in the kinetic folding process has a significant implication on the folding mechanism of SMN.

Structural Property of Equilibrium Conformations. The first equilibrium intermediate in the alkaline unfolding transition of SMN, I_{eq2}, is formed from N with a pK_a of ≈10.2. The CD spectrum for I_{eq2} is similar to that of N except for the region around 235 nm assignable to aromatic residues. In addition, we observed a significant quenching of Trp-4 fluorescence for I_{eq2} compared with that of N. These observations suggest structural changes around Trp-4 without a drastic destruction of secondary structures. In the x-ray structure of NMEI (29), Trp-4 is sandwiched by Met-42 and Lys-44. The association of sulfur-containing residues with Trp is known to quench Trp fluorescence moderately (33). Furthermore, the contact between Lys and Trp should be stabilized by a cation–aromatic interaction (34). Thus, we suggest that the sandwich structure by Met-42, Trp-4, and Lys-44 is present in I_{eq2} but is modulated in the transition from N to I_{eq2}.

The examination of the IR spectrum for I_{eq2} suggests the presence of water in the interface between β-sheet and α-helix. The second derivative IR spectrum for I_{eq2} possesses a negative peak at 1,637 cm^{-1}, which is assigned to solvated α-helix, in which the amide C═O group accepts hydrogen bond from water in addition to the hydrogen bond from amide proton (25, 26, 35, 36). Although the line is not sometimes apparent in helical proteins at room temperature because of thermal fluctuations of the hydrogen bonds (25, 37), helical polypeptides that are fully exposed to water show the line at room temperature (35, 36). The second derivative spectrum for I_{eq2} also possesses features at 1,626 and 1,683 cm^{-1} assigned to the major and minor amide I′, respectively, of the β-sheet split by transient dipolar coupling (19–21). To obtain detailed structural information of the β-sheet, the average and difference of the major and minor lines were estimated (Table 1), which roughly correspond to the intrinsic frequency and twice the coupling constant between the nearest C═O groups in the adjacent strands, respectively (19–21). Although the splitting of I_{eq2} (56.9 cm^{-1}) is smaller than that of N (57.9 cm^{-1}), the similarity in the CD spectra for N and I_{eq2} indicates that the β-sheet of I_{eq2} is still packed. The intrinsic frequency shows a small low-frequency shift (3.4 cm^{-1}). Because the hydrogen bonding interaction on amide C═O by water causes a low-frequency shift of amide I (25, 26), an additional interaction between amide C═O in the β-sheet and water likely occurs in the transition from N to I_{eq2}. The appearance of a solvated helix and small changes in the CD spectrum for I_{eq2} support our interpretation. The amide I′ lines for β-sheet are known to demonstrate a low-frequency shift under high pressure, which is generally explained by the elastic compression of hydrogen bonds (38). The explanation is less likely to apply to the current observation conducted at ambient pressure. Interestingly, penetration of water in the core cavity of proteins is demonstrated in the high-pressure crystallographic analysis (39). We propose that water penetrates inside the interface of the helix and sheet in I_{eq2} without causing a drastic destruction of the β-sheet.

The next intermediate, I_{eq1}, appears in the pH region between 11.0 and 11.9. It possesses reduced CD values at 216 nm, showing a reduction in β-sheet content. I_{eq1} has a weak fluorescence spectrum, whose maximum (348 nm) implies that Trp-4 is still

partially buried from solvent. We suggest that Trp-4 maintains contact with Met-42. The IR spectrum of I_{eq1} displays features for solvated helix (1,635 cm^{-1}) and for β-sheet (1,624 and 1,680 cm^{-1}). The splitting between the major and minor lines for β-sheet is 56.4 cm^{-1} and is smaller than those for I_{eq2} and N (Table 1). The smaller splitting suggests the reduced number of strands and/or an increase in the twist angle of the β-sheet (19, 40). The drastic reduction in the CD intensity for I_{eq2} supports the former possibility. Furthermore, the low-frequency shift in the intrinsic frequency (6.1 cm^{-1}) and the greater content of the solvated helix at 1,635 cm^{-1} imply that a larger amount of water penetrates the interface between helix and sheet. We conclude that the low-frequency shifts in the amide I′ lines of the β-sheet are caused by the hydration of main-chain amides. Thus, detailed structural information on the β-sheet can be obtained by examining the transient dipolar coupling and the intrinsic frequency of amide I′.

Dynamic Conformational Transitions in the Folding of SMN. The kinetic IR investigation on the process after the pD jump from 13 to 9.4 revealed detailed information on the folding of SMN. The kinetic spectrum obtained at 150 μs after the pD jump resembles that of the unfolded conformation and possesses no features of a β-sheet. However, the second derivative spectrum for I_1 clearly shows a component at 1,637 cm^{-1}, which is assigned to solvated helix with a content of ≈11%. Thus, the IR results show that there exists a significant amount of solvated helices in I_1. In addition, several weak features are seen at 1,666 and 1,675 cm^{-1}, which are assigned to turns. These observations can be compared with the previous structural data on I_1 (11). At 300 μs after the initiation of folding, we observed a very small change in the CD spectrum in the far-UV region, which is consistent with the small amount of β-sheet. Interestingly, the CD spectrum shows a small change around 233 nm, which can be ascribed to the aromatic residues. In addition, the Trp-4 fluorescence in I_1 is quenched. Together with the time-resolved SAXS results showing the collapsed conformation within the mixing dead time, we suggest that the specific structures around Trp-4 are formed in this time scale. The N state of SMN possesses a major hydrophobic core in the interface between β-sheet and helix. To explain the collapsed conformation of I_1 and the current IR results, we propose that the major hydrophobic core is loosely assembled and that the N-terminal of SMN forms some specific structures. These structures should be fluctuating, however, because the hydrophobic extrinsic dye cannot bind to I_1 (11).

The next intermediate, I_2, accumulates at 250 ms after the initiation of SMN folding. The kinetic IR spectrum for I_2 shows features at 1,637 cm^{-1} for the solvated helix (12%) and at 1,627 cm^{-1} (34%) and 1,680 cm^{-1} (3%) for the β-sheet. Whereas the content for the total of solvated helix and α-helix for I_2 (17%) is comparable to that of N (17%), the total content for β-sheet (37%) for I_2 is slightly smaller than that of N (43%), indicating that an additional stabilization of secondary structures is likely necessary to form N. The splitting of major and minor peaks for the β-sheet (52.5 cm^{-1}) is smaller than that of N (57.8 cm^{-1}), implying that the formation of the main-chain hydrogen bonds in β-sheet is not completed in I_2. The frequencies for the major and minor lines of the β-sheet are both lower than those for N, corresponding to the low-frequency shift of the intrinsic frequency by 4.6 cm^{-1} (Table 1). Thus, we conclude that the main-chain C═O in the β-sheet is hydrated. Together with the significant amount of solvated helix (12%), we interpret that the core domain in I_2 is hydrated. The interpretation is consistent with the fluctuating core domain suggested from the binding of the hydrophobic dye in I_2 (11). The time-resolved SAXS measurements showed that I_2 possesses small R_g and an overall oblate shape that are indistinguishable from N. The kinetic CD data at 216 nm for I_2 is enhanced more than that of I_1. These observations are in agreement with the formation of the hydrophobic core that is composed of the solvated helix and sheet.

The features of I_2 obtained from the kinetic IR, CD, and fluorescence data are similar to those of I_{eq2} except for the larger splitting of amide I′ for β-sheet (56.9 cm^{-1}) than that for I_{eq2} (52.5 cm^{-1}). The other equilibrium intermediate (I_{eq1}), which is most stabilized at pD ≈11.5, was not detected during the kinetic measurements at pD 9.4 likely because of the destabilization at lower pD.

Coincidence of the Dehydration with the Rate-Limiting Step of Protein Folding. The final process of SMN folding detected by IR spectroscopy is the conversion of the water-penetrated core to the dehydrated core and the additional stabilization of secondary structures. The corresponding process was observed in the folding of apoMb, in which the dehydration of the solvated helix occurs in the rate-limiting step (14). Considering that SMN and apoMb possess distinct folds and secondary structures, the dehydration dynamics, coincident with the rate-limiting step, is likely the general property of protein folding. The observations do not support the dewetting mechanism of polypeptide collapse, which explains the collapse as a contraction process of polypeptides inside a preformed microbubble and assumes the drying of water in a single step (41). The current observation is rather consistent with the water expulsion mechanism, in which water is expelled from the collapsed domain in a stepwise manner as demonstrated in several model calculations and molecular dynamics simulations (4, 7, 42). The origin of the stepwise expulsion was frequently suggested to be the stabilization of the direct contact between water and hydrophobic molecules (43, 44). We further suggest that hydrogen bonding interaction between main-chain amides and water might contribute to the slow dehydration. Although the current IR results do not directly provide information on the energetic importance of the hydration effects, it has been proposed that the expulsion of water molecules, which occurs in passing from the intermediate state to the transition state (45, 46), makes the rate-limiting transition state enthalpically unfavorable. The solvation enthalpy for amides is known to be significantly larger than that for hydrophobic molecules (5, 6). Therefore, we propose that the dissociation of hydrogen bonds between water and main-chain amides generally contributes to the enthalpic barrier in the rate-determining step of the folding of proteins.

Several additional features of the amide I′ spectra for the kinetic and equilibrium intermediates have an important implication on the folding dynamics. First, the observed amid I′ frequencies of the β-sheet in the equilibrium and kinetic intermediates appear at distinct locations without significant broadening or splitting. Considering that the transient dipolar coupling of amide I causes a β-sheet as a single vibrating unit by working mainly between the adjacent C═O (19), the observation suggests that the intermediates are not composed of fragments of several β-sheets but possess a single β-sheet. A recent observation on the importance of this cooperative effect in the hydrogen bonding network in the secondary structure might also explain the apparent homogeneity of amide I in the β-sheet (47). Second, we found that the low-frequency shift in the intrinsic frequency for the β-sheet caused by hydration (≈5 cm^{-1}) is smaller than that for the α-helix (≈15 cm^{-1}). The difference shows that the hydration strengths for the different secondary structures in the collapsed intermediates are distinct. It was proposed that a partial dehydration of amides by side chains regulates the propensity for secondary structure formation (15, 16). The results suggest that the hydration dynamics might regulate the formation of secondary structures in the collapsed conformation. Further examination of the folding dynamics of proteins based on kinetic IR spectroscopy will reveal rich interplay between water and the structural dynamics of proteins.

Materials and Methods

Preparation of SMN. All chemicals were of reagent grade (Wako Chemicals). SMN used in this study is the single-chain variant of monellin and was prepared as

described by Konno (28). We previously confirmed the monomeric conformation of SMN throughout its kinetic folding process induced by the pH jump from 13.0 to 9.4 at the protein concentration required for the kinetic IR investigation (11). Lyophilized SMN was dissolved in D_2O and centrifuged to remove insoluble components. The dissolved SMN was thermally unfolded to fully deuterize the protein and then lyophilized. The lyophilized SMN was again dissolved in D_2O and centrifuged. The protein concentrations were confirmed by measuring the absorption at 277 nm ($\varepsilon_{277} = 1.46 \times 10^4$ $M^{-1}\cdot cm^{-1}$) (28). The amount of SMN consumed for the entire measurements was ≈2 g. All pD measurements were made with the pH meter (P260; Beckman) equipped with a glass electrode at room temperature. To correct solvent isotope effect on the glass electrode, the direct pH meter reading was corrected by adding 0.40.

The Equilibrium Measurements. SMN samples at various pH (pD for IR measurements) were prepared by mixing two solutions with appropriate ratio: the native state in 50 mM glycine buffer at pH 7.0 and the alkaline unfolded state in 50 mM glycine at pH 13.0. The final protein concentrations for CD, Trp fluorescence, and IR measurements were 15, 5, and 180 μM, respectively. We used a FTIR spectrometer equipped with DTGS detector (FTS-575C; Bio-Rad) that is continuously purged with dry air gas. Interferograms of the samples were acquired at 2-cm^{-1} resolution 512 times and Fourier-transformed with a four-point triangular apodization function. Full deuteration of the amide groups was confirmed by the absence of NH and CNH vibration lines at 1,550 and 3,300 cm^{-1}, respectively, in the FTIR spectra. The analysis methods of the equilibrium data are described in *SI Text*.

Time-Resolved IR Measurements. The kinetic IR spectra were obtained by using an IR spectrometer (FTS-575C) and a microscope (UMA-500). These were continuously purged with dry N_2 gas to remove the vapor in the system. The IR light was focused to a spot (100 μm) at the observation channel by a casegrane of the microscope. Progress in the folding reaction was detected by moving the mixer-flow cell assembly to the focusing point of the IR light. Interferograms were acquired at 4-cm^{-1} nominal resolution 64 times and Fourier-transformed to the single-beam spectra with a four-point triangular apodization function by Win-IR (Bio-Rad). The temperature of the solution was monitored by using a thermocouple placed at the exit of the flow cell. Three kinds of continuous-flow mixers were used to initiate the folding reactions by mixing the solution of alkaline-unfolded SMN and 100 mM glycine buffer at a volume ratio of 1:1 to a final pD of 9.4 at 25°C. The final concentration of SMN was 1.4 $mg\cdot ml^{-1}$. The IR spectra in the time domain from 150 μs to 5.0 ms were obtained by using the T-shaped mixer whose observation flow channel is 100 μm wide and 100 μm deep (31). The flow speeds of the solution in the observation channel of the mixer were 16.8 and 8.4 $m\cdot s^{-1}$ for the observation of the time domains between 150 μs and 1.1 ms and between 1.0 ms and 5.0 ms, respectively. The different continuous-flow mixer developed by Unisoku was used to monitor the IR spectra in the time windows between 10 and 23 ms (14). The kinetic spectra after 50 ms were obtained by introducing solutions mixed by Nanomixer (Upchurch) (48) to the flow channel with 100-μm width and 100-μm depth.

ACKNOWLEDGMENTS. We thank Prof. Hajime Torii (Shizuoka University, Shizuoka, Japan) for numerous discussions. This study was supported by Grants-in-Aid for Scientific Research (to S.T., K.I., and I.M.) from the Ministry of Education, Culture, Sports, Science and Technology. T. Kimura was supported by a fellowship of the Japan Society for the Promotion of Science to Young Scientists.

1. Fersht AR (1999) *Structure and Mechanism in Protein Science* (Freeman, New York).
2. Paulaitis ME, Pratt LR (2002) Hydration theory for molecular biophysics. *Adv Protein Chem* 62:283–310.
3. Baldwin RL (2005) in *Protein Folding Handbook, Part I*, eds Buchner J, Kiefhaber T (Wiley, Weinheim, Germany), pp 127–162.
4. Levy Y, Onuchic JN (2006) Water mediation in protein folding and molecular recognition. *Annu Rev Biophys Biomol Struct* 35:389–415.
5. Baldwin RL (2007) Energetics of protein folding. *J Mol Biol* 371:283–301.
6. Makhatadze GI, Privalov PL (1995) Energetics of protein structure. *Adv Protein Chem* 47:307–425.
7. Cheung MS, Garcia AE, Onuchic JN (2002) Protein folding mediated by solvation: Water expulsion and formation of the hydrophobic core occur after the structural collapse. *Proc Natl Acad Sci USA* 99:685–690.
8. Rhee YM, Sorin EJ, Jayachandran G, Lindahl E, Pande VS (2004) Simulations of the role of water in the protein-folding mechanism. *Proc Natl Acad Sci USA* 101:6456–6461.
9. Akiyama S, *et al.* (2002) Conformational landscape of cytochrome *c* folding studied by microsecond-resolved small-angle x-ray scattering. *Proc Natl Acad Sci USA* 99:1329–1334.
10. Uzawa T, *et al.* (2004) Collapse and search dynamics of apomyoglobin folding revealed by submillisecond observations of α-helical content and compactness. *Proc Natl Acad Sci USA* 101:1171–1176.
11. Kimura T, *et al.* (2005) Specific collapse followed by slow hydrogen-bond formation of β-sheet in the folding of single-chain monellin. *Proc Natl Acad Sci USA* 102:2748–2753.
12. Uzawa T, *et al.* (2006) Time-resolved small-angle X-ray scattering investigation of the folding dynamics of heme oxygenase: Implication of the scaling relationship for the submillisecond intermediates of protein folding. *J Mol Biol* 357:997–1008.
13. Arai M, *et al.* (2007) Microsecond hydrophobic collapse in the folding of *Escherichia coli* dihydrofolate reductase, an α/β-type protein. *J Mol Biol* 368:219–229.
14. Nishiguchi S, Goto Y, Takahashi S (2007) Solvation and desolvation dynamics in apomyoglobin folding monitored by time-resolved infrared spectroscopy. *J Mol Biol* 373:491–502.
15. Avbelj F, Baldwin RL (2002) Role of backbone solvation in determining thermodynamic beta propensities of the amino acids. *Proc Natl Acad Sci USA* 99:1309–1313.
16. Luo P, Baldwin RL (1999) Interaction between water and polar groups of the helix backbone: An important determinant of helix propensities. *Proc Natl Acad Sci USA* 96:4930–4935.
17. Phillips CM, Mizutani Y, Hochstrasser RM (1995) Ultrafast thermally induced unfolding of RNase A. *Proc Natl Acad Sci USA* 92:7292–7296.
18. Troullier A, Reinstadler D, Dupont Y, Naumann D, Forge V (2000) Transient non-native secondary structures during the refolding of α-lactalbumin detected by infrared spectroscopy. *Nat Struct Biol* 7:78–86.
19. Barth A, Zscherp C (2002) What vibrations tell us about proteins. *Q Rev Biophys* 35:369–430.
20. Krimm S, Bandekar J (1986) Vibrational spectroscopy and conformation of peptides, polypeptides, and proteins. *Adv Protein Chem* 38:181–364.
21. Torii H, Tasumi M (1996) in *Infrared Spectroscopy of Biomolecules*, eds Mantsch HH, Chapman D (Wiley-Liss, New York), pp 1–18.
22. Krimm S, Abe Y (1972) Intermolecular interaction effects in the amide I vibrations of polypeptides. *Proc Natl Acad Sci USA* 69:2788–2792.
23. Miyazawa T (1960) Perturbation treatment of the characteristic vibrations of polypeptide chains in various configurations. *J Chem Phys* 32:1647–1652.
24. Kauffmann E, Darnton NC, Austin RH, Batt C, Gerwert K (2001) Lifetimes of intermediates in the β-sheet to α-helix transition of β-lactoglobulin by using a diffusional IR mixer. *Proc Natl Acad Sci USA* 98:6646–6649.
25. Manas ES, Getahun Z, Wright WW, DeGrado WF, Vanderkooi JM (2000) Infrared spectra of amide groups in α-helical proteins: Evidence for hydrogen bonding between helices and water. *J Am Chem Soc* 122:9883–9890.
26. Walsh ST, *et al.* (2003) The hydration of amides in helices; a comprehensive picture from molecular dynamics, IR, and NMR. *Protein Sci* 12:520–531.
27. Schultz CP (2000) Illuminating folding intermediates. *Nat Struct Biol* 7:7–10.
28. Konno T (2001) Multistep nucleus formation and a separate subunit contribution of the amyloidgenesis of heat-denatured monellin. *Protein Sci* 10:2093–2101.
29. Somoza JR, *et al.* (1993) Two crystal structures of a potently sweet protein. Natural monellin at 2.75 Å resolution and single-chain monellin at 1.7 Å resolution. *J Mol Biol* 234:390–404.
30. Lee SY, *et al.* (1999) Solution structure of a sweet protein single-chain monellin determined by nuclear magnetic resonance and dynamical simulated annealing calculations. *Biochemistry* 38:2340–2346.
31. Kimura T, *et al.* (2002) Direct observation of the multistep helix formation of poly-L-glutamic acids. *J Am Chem Soc* 124:11596–11597.
32. Vuilleumier S, Sancho J, Loewenthal R, Fersht AR (1993) Circular dichroism studies of barnase and its mutants: Characterization of the contribution of aromatic side chains. *Biochemistry* 32:10303–10313.
33. Yuan T, Weljie AM, Vogel HJ (1998) Tryptophan fluorescence quenching by methionine and selenomethionine residues of calmodulin: Orientation of peptide and protein binding. *Biochemistry* 37:3187–3195.
34. Shi Z, Olson CA, Kallenbach NR (2002) Cation-π interaction in model α-helical peptides. *J Am Chem Soc* 124:3284–3291.
35. Martinez G, Millhauser G (1995) FTIR spectroscopy of alanine-based peptides—Assignment of the amide I′ modes for random coil and helix. *J Struct Biol* 114:23–27.
36. Reisdorf WC, Krimm S (1996) Infrared amide I′ band of the coiled coil. *Biochemistry* 35:1383–1386.
37. Turner DR, Kubelka J (2007) Infrared and vibrational CD spectra of partially solvated α-helices: DFT-based simulations with explicit solvent. *J Phys Chem B* 111:1834–1845.
38. Herberhold H, Winter R (2002) Temperature- and pressure-induced unfolding and refolding of ubiquitin: A static and kinetic Fourier transform infrared spectroscopy study. *Biochemistry* 41:2396–2401.
39. Collins MD, Quillin ML, Hummer G, Matthews BW, Gruner SM (2007) Structural rigidity of a large cavity-containing protein revealed by high-pressure crystallography. *J Mol Biol* 367:752–763.
40. Kubelka J, Keiderling TA (2001) The anomalous infrared amide I intensity distribution in ^{13}C isotopically labeled peptide β-sheets comes from extended, multiple-stranded structures: An *ab initio* study. *J Am Chem Soc* 123:6142–6150.
41. Chandler D (2005) Interfaces and the driving force of hydrophobic assembly. *Nature* 437:640–647.
42. Athawale MV, Goel G, Ghosh T, Truskett TM, Garde S (2007) Effects of lengthscales and attractions on the collapse of hydrophobic polymers in water. *Proc Natl Acad Sci USA* 104:733–738.
43. Hummer G, Garde S, Garcia AE, Paulaitis ME, Pratt LR (1998) The pressure dependence of hydrophobic interactions is consistent with the observed pressure denaturation of proteins. *Proc Natl Acad Sci USA* 95:1552–1555.
44. Li JL, Car R, Tang C, Wingreen NS (2007) Hydrophobic interaction and hydrogen-bond network for a methane pair in liquid water. *Proc Natl Acad Sci USA* 104:2626–2630.
45. Kimura T, Sakamoto K, Morishima I, Ishimori K (2006) Dehydration in the folding of reduced cytochrome *c* revealed by the electron-transfer-triggered folding under high pressure. *J Am Chem Soc* 128:670–671.
46. Staniforth RA, *et al.* (2000) The major transition state in folding need not involve the immobilization of side chains. *Proc Natl Acad Sci USA* 97:5790–5795.
47. Zandomeneghi G, Krebs MR, McCammon MG, Fandrich M (2004) FTIR reveals structural differences between native β-sheet proteins and amyloid fibrils. *Protein Sci* 13:3314–3321.
48. Kakuta M, Hinsmann P, Manz A, Lendl B (2003) Time-resolved Fourier transform infrared spectrometry using a microfabricated continuous flow mixer: Application to protein conformation study using the example of ubiquitin. *Lab Chip* 3:82–85.