Side Chain Accessibility and Dynamics in the Molten Globule State of α-Lactalbumin: A ^{19}F-NMR Study

Ping Bai, Li Luo, and Zheng-yu Peng*

Department of Biochemistry, University of Connecticut Health Center, 263 Farmington Avenue, Farmington, Connecticut 06032

Received September 1, 1999; Revised Manuscript Received November 1, 1999

ABSTRACT: The molten globule state of α-lactalbumin (α-LA) has been considered a prototype of partially folded proteins. Despite the importance of molten globules in understanding the mechanisms of protein folding and its relevance to some biological phenomena, site-specific information on the structure and dynamics of a molten globule is limited, largely because of the high conformational flexibility and heterogeneity. Here, we use selective isotope labeling and ^{19}F NMR to investigate the solvent accessibility and side-chain dynamics of aromatic residues in the molten globule of α-LA. Comparison of these properties with those of the native and unfolded protein indicates that the α-LA molten globule is highly heterogeneous; each residue has its unique solvent accessibility and motional environment. Many aromatic residues normally buried in the interior of native α-LA remain significantly buried in the molten globule and the side-chain dynamics of these residues are highly restricted. Our results suggest that hydrophobic and van der Waals interactions mediated by the inaccessible surface area could be sufficient to account for all the stability of the α-LA molten globule, which is approximately 50% of the value for the native protein.

Molten globules have long served as model systems for studying protein folding (*1*−*3*). The high level of secondary structure and rudimentary, nativelike tertiary topology found in these equilibrium partially folded species, often in the absence of rigid, specific side chain packing, resembles the structural characteristics of kinetic folding intermediates (*4*−*8*). Recently, molten globule-like proteins have been observed under physiological conditions, serving as precursors for protein-protein, protein-ligand, and protein-DNA interactions (*9*−*13*). Point mutations can also convert native proteins into molten globule-like species, often resulting in loss of function (*14*−*16*). Some of these mutations correlate with a genetic predisposition to human diseases (*16, 17*). Therefore, understanding the structure and dynamics of molten globules and other partially folded states of proteins is an important area in structural biology.

Traditionally, it has been difficult to obtain site-specific information for a molten globule, mainly because at a microscopic level, a molten globule does not have a unique conformation. Instead, it corresponds to an ensemble of closely related conformations (for example, see ref *18* and references therein). This structural heterogeneity and interconversion between different conformations inhibit crystallization and lead to poorly resolved NMR spectra with significant line broadening, making direct assignment of individual residues extremely difficult. Although NMR hydrogen exchange studies have been used to detect hydrogen bond formation in secondary structure elements in molten globules, these studies provide no information on side-chain conformation and dynamics. In particular, the side-chain solvent accessibility is an important parameter for understanding the driving force for protein folding, the molecular interactions that stabilize the molten globule, and the fundamental differences between a partially folded intermediate and a native or unfolded protein.

α-Lactalbumin (α-LA)[1] is a small, two-domain protein (Figure 1). The α-LA molten globule has been studied under a variety of conditions, including as a transient intermediate during the kinetic folding reaction and as an equilibrium species at acidic pH, in the presence of low concentrations of denaturant, or upon partial reduction of its disulfide bonds (for reviews, see refs *1, 19*). Several recent studies suggest that in the molten globule state, the α-helical domain of α-LA retains a nativelike tertiary topology, whereas the *β*-sheet domain is largely disordered (*20*−*22*). The best characterized form of the α-LA molten globule is the one obtained at low pH conditions, which is often referred to as the A-state.

Here, we report a ^{19}F-NMR study of the A-state of α-LA. ^{19}F is a spin 1/2 nuclide that has a high NMR sensitivity and 100% natural abundance. ^{19}F-substituted aromatic amino acids can be readily incorporated into proteins during the process of biosynthesis (for reviews, see refs *23, 24*). Since only a small number of residues are labeled, the ^{19}F-NMR spectra can usually be assigned by site-directed mutagenesis.

† This work was supported by the National Institute of Health (GM54533) to Z.-y. P.

* Corresponding author. Mailing address: MC-3305, University of Connecticut Health Center, 263 Farmington Avenue, Farmington, CT 06032. Telephone: (860) 679-2885. Fax: (860) 679-3408. E-mail: peng@sun.uchc.edu.

[1] Abbreviations used: α-LA, α-lactalbumin; A-state, the acid state; CD, circular dichroism; NMR, nuclear magnetic resonance; SIIS, solvent-induced isotope shift; ATP, adenosine 5′-triphosphate; DMF, *N*, *N*-dimethylformamide; 4-F-Phe, 4-fluoro-D,L-phenylalanine; 5-F-Trp, 5-fluoro-D,L-tryptophan; GuHCl, guanidine hydrochloride; NADH, *β*-nicotinamide adenine dinucleotide reduced form.

10.1021/bi992056f CCC: $19.00
Published on Web 12/15/1999

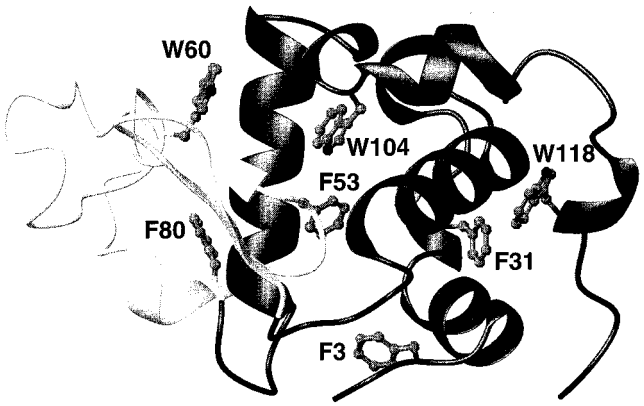

FIGURE 1: The backbone structure of human α-LA (PDB code 1HML) (*61*). The side chains of phenylalanine and tryptophan residues are shown as ball-and-stick models. The α-helical domain (dark) contains two phenylalanine (F3 and F31) and two tryptophan (W104 and W118) residues, and the β-sheet domain (light) contains two phenylalanine (F53 and F80) and one tryptophan (W60) residue. This figure was produced using the program MOLMOL (*62*).

Fluorine substitutions normally do not significantly perturb the structure and function of a protein because fluorine and hydrogen have similar atomic radii. The most important advantage of fluorine is that the chemical shift value of ^{19}F is extremely sensitive to its environment and solvent conditions, making it an excellent probe for studying protein folding and conformational changes (for examples, see refs *25*−*28*). In this study, we focus on the solvent accessibility and side-chain dynamics of aromatic residues in the molten globule state of α-LA and compare these properties with those in the native or unfolded state of α-LA.

The results of our studies strongly suggest that the aromatic residues buried in the interior of native α-LA remain significantly buried in the molten globule. Thus, both hydrophobic effect and van der Waals interactions could potentially make a significant contribution to the stability of the α-LA molten glonule. In addition, the side-chain dynamics in the α-LA molten globule are highly heterogeneous. Each residue has its own motional characteristic and the mobility is restricted by the compact structure and nativelike tertiary topology, resulting in slow conformational fluctuations that are unique to the molten globule. This information is useful for understanding how conformational space becomes progressively more restricted during the process of protein folding.

MATERIALS AND METHODS

Protein Expression and Labeling. Recombinant human α-LA and mutant proteins were expressed from a synthetic gene with an optimal *Escherichia coli* codon usage as described previously (*29*). The protein has an additional methionine at the N-terminus. Site-directed mutagenesis was performed using Kunkel's method (*30*). The mutations were verified by manual and/or automated DNA sequencing. Generally, the *E. coli* strain DL39 (genotype *aspC, ilvE, tyrB, avtA*; auxotrophic for Asp, Ile, Leu, Phe, and Tyr) was used to produce 4-F-Phe labeled protein and CT19 (genotype *aspC, ilvE, tyrB, avtA, trpB*; auxotrophic for Asp, Ile, Leu, Phe, Tyr, and Trp) was used to produce 5-F-Trp-labeled protein (*31, 32*). Both of these strains carry the λDE3 prophage, which contains an IPTG inducible T7 RNA polymerase gene. Two procedures were used to grow the cells. In the first procedure, cells were grown in a defined medium containing all twenty amino acids and four nucleosides to an $OD_{600} = 0.8$ (*31*). The cells were collected by centrifugation and resuspended in twice the original volume of the defined medium containing the labeled amino acid and other nineteen unlabeled amino acids. The cultures were grown for an additional 1−2 h before induction with 0.4 mM IPTG. In the second procedure, an overnight culture grown in a defined medium with unlabeled amino acids was diluted between 1:50 and 1:100 into a defined medium without phenylalanine or tryptophan. The cultures were grown to an $OD_{600} = 0.6$. By this time, most of the unlabeled phenylalanine or tryptophan carried over from the overnight medium would have been exhausted (this was indicated by slowing down in growth). The labeled amino acid was added to the growth medium and after 15−30 min, the cells were induced by 0.4 mM IPTG. Cells were harvested 2−3 h after induction. α-LA and its mutant proteins were purified from inclusion bodies and refolded according to protocols described previously (*20, 22*). The purified protein was lyophilized and stored at −80 °C.

One particular mutant protein, F31Y/F3Y, could not be expressed from auxotrophic strains. This protein was expressed using strain BL21 (DE3) grown in M9 minimum media with glyphosate (1 g/L) as an inhibitor of aromatic amino acid synthesis (*33*). The inhibitor was added at the same time with 4-F-Phe and unlabeled Tyr and Trp (1 mM each, final concentration) 30 min prior to IPTG induction as described earlier.

Circular Dichroism (CD). CD experiments were performed on a JASCO J-715 spectropolarimeter equipped with a thermoelectric temperature controller. The samples consisted of 10 or 20 μM protein, 10 mM Tris, 1 mM $CaCl_2$, at pH 8.5. Far-UV CD spectra were recorded using a 1 mm path length cuvette with 1 nm spectral resolution. Protein concentrations were determined by absorbance at 280 nm in 6 M guanidine hydrochloride (GuHCl) (*34*), assuming that the incorporation of 5-F-Trp does not alter the protein's extinction coefficient (*35*).

Enzymatic Assay. The lactose synthase activity was measured at room temperature using a coupled enzymatic reaction (*36*). The reaction mixture (1 mL) consisted of 0.4 mM uridine 5′-diphospho-galactose, 100 mM glucose, 0.1 mM ATP, 1 mM phospho(enol)pyruvate, and 0.2 mM NADH in 50 mM glycylglycine buffer, pH 8.0 with 5 mM $MnCl_2$. To this mixture, 0.02 unit of galactosyl-transferase (Sigma G5507, 1 Sigma unit equals to 1000 Ebner units), 10 μL of pyrurate kinase (Sigma type I, P1381, suspension in 3.2 M ammonium sulfate) and various amount of α-LA (diluted from a stock solution) was added. The reaction was followed by the absorbance at 340 nm for 15 min using a CARY 1E spectrophotometer and the initial part of the decay was fitted by linear regression. The specific activity was determined by plotting the initial rate as a function of the α-LA concentration.

NMR Spectroscopy. The ^{19}F-NMR spectra were collected using a Varian Unity Plus 500 MHz spectrometer with a triple resonance probe. The proton channel was tuned to fluorine frequency (471.2 MHz). The samples contained 0.5−2 mg of purified α-LA dissolved in 620 μL of buffer. The native buffer consisted of 10 mM Tris base, 1 mM

$CaCl_2$; the molten globule buffer consisted of 5 mM HCl; and the unfolded buffer consisted of 5 mM HCl and 6 M GuHCl. All buffers contained 10% D_2O, except for those used in solvent accessibility measurements. The pH of the sample was adjusted to 8.5 for the native state and to 2.0 for the molten globule and unfolded states with 0.1 and 1 M HCl (the pH value was not corrected for the presence of D_2O). The total amount of acid added was always less than 1% of the total sample volume. Free 5-F-Trp (20 μM) was used as an internal chemical shift standard for 4-F-Phe-labeled proteins, and free 4-F-Phe and/or 3-F-Tyr (20 μM each) was used as an internal chemical shift standard for 5-F-Trp-labeled proteins. Since the probe has only one high-frequency channel, no proton decoupling was used during the acquisition. The excitation pulse length was 8.8 μs (approximately 90°) and the acquisition time and relaxation delay were 0.68 and 1 s, respectively. Typically, 8192−32768 scans were averaged for each spectrum. The spectra were processed using the program FELIX with 5 Hz exponential line broadening.

Determination of Solvent Accessibility. For solvent accessibility measurements, the ^{19}F-NMR spectra were recorded in buffers containing 10, 30, 50, 70, and 90% D_2O and the chemical shift value of each peak was plotted as a function of the D_2O content in the sample. The chemical shift differences ($\Delta\delta$) between 0 and 100% D_2O were determined by linear regression. This value was divided by the $\Delta\delta$ of a solvent-induced isotope shift (SIIS) standard, which contains the same type of side chain completely exposed to solvent. The result is presented as a percentage of solvent accessibility. Since the accuracy of chemical shift determination was approximately ±0.01 ppm (~5 Hz), we expect that the experimental error in our solvent accessibility measurements would be ±5%.

To compare the solvent accessibility measured by ^{19}F NMR with the values derived from X-ray crystal structure, we calculated the solvent accessible surface area for each Cζ in phenylalanine and Cζ3 in tryptophan (these are the carbon atoms to which the fluorine labels were attached to) using the PDB data set of human α-LA (accession code 1HML) and a computer program based on the polyhedron integration algorithm (*37, 38*). The percentage of solvent accessibility was obtained by dividing the solvent accessible surface area in the intact protein with that of a free amino acid.

Line Width Analysis. The line width of ^{19}F resonances were determined using the "optimize" routine in FELIX to fit the spectra with a set of Lorentzian functions by simulated annealing. The values listed in Table 2 were obtained from a representative spectrum. In several cases, the line width measurements were repeated 2−5 times using spectra acquired independently and different initial parameters for fitting; the variations were found to be less than ±10%.

Synthesis of SIIS Standards. The extent of SIIS effect for free 4-F-Phe and 5-F-Trp are pH dependent (*39*). Therefore, they cannot be used as SIIS standards at low pH conditions. Consequently, we synthesized *N*-acetyl-4-fluoro-D,L-phenylalanine methyl ester and *N*-acetyl-5-fluoro-D,L-tryptophanamide as pH independent SIIS standards. To synthesize *N*-acetyl-4-fluoro-D,L-phenylalanine methyl ester, approximately 200 mg of 4-fluoro-D,L-phenylalanine methyl ester (Bachem) was dissolved in 10 mL of *N*,*N*-dimethylformamide (DMF) and 1 mL of acetic anhydride. To this solution, 1 mL of diisopropylethylamine was added, and the reaction mixture was stirred at room temperature for 1−2 h. After the solvent was removed, the residue was dissolved in 5% acetic acid and *N*-acetyl-4-fluoro-D,L-phenylalanine methyl ester was purified by reverse phase HPLC using a Vydac C_{18} column and water−acetonitrile buffer. To synthesize *N*-acetyl-5-fluoro-D,L-tryptophanamide, we first synthesized *N*-acetyl-5-fluoro-D,L-tryptophan using the same method as described above and purified the product by HPLC. Approximately 180 mg *N*-acetyl-5-fluoro-D,L-tryptophan was obtained from 1 mmol (222 mg) of starting material. For the second step of the reaction, 170 mg of *N*-acetyl-5-fluoro-D,L-tryptophan was dissolved in 10 mL of DMF. To this solution, 570 mg of 1,3-dicyclohexyl-carbodiimide was added, and the reaction mixture was stirred at room temperature for 30 min. Insoluble material started to form in about 10 min and was removed by filtration after 30 min; 10 mL of ammonia (0.5 M solution in 1,4-dioxane) was added to the clarified solution and the reaction was allowed to proceed at room temperature overnight. After the solvent was removed, the residue was dissolved in ethanol and diluted by an equal volume of 5% acetic acid. *N*-acetyl-5-fluoro-D,L-tryptophanamide was purified from this material by reverse phase HPLC. The identity of the purified, protected forms of the fluorine-labeled amino acids was confirmed by mass spectrometry.

RESULTS

Protein Labeling and Spectral Assignment. Human α-LA contains four phenylalanines and three tryptophans, which are evenly distributed in the three-dimensional structure (Figure 1). In this study, we produced 4-F-Phe and 5-F-Trp labeled protein in vivo by biosynthetic incorporation of fluorinated amino acids using auxotrophic bacterial strains and/or an inhibitor of aromatic amino acid synthesis. The labeling efficiency of the wild type α-LA and several mutant proteins were estimated by electrospray mass spectrometry, since each fluorine substitution increases the molecular mass of the protein by 18 dalton and molecules with different number of fluorine atoms can be separated. The labeling efficiency was found to be 60−100%, depending on the bacterial strain and labeling method (data not shown). Figure 2 shows the far-UV CD spectra of 4-F-Phe and 5-F-Trp-labeled α-LA (both are essentially 100% labeled). The 4-F-Phe-labeled α-LA exhibits a CD spectrum identical to that of the unlabeled protein, whereas the 5-F-Trp-labeled α-LA has a slightly higher mean residue ellipticity than that of the unlabeled protein in the far-UV region (this could be caused by a small difference in the extinction coefficient between the labeled and unlabeled protein). In addition, the stability of the 4-F-Phe- and 5-F-Trp-labeled α-LA is similar to that of the unlabeled material, as indicated by urea denaturation studies (data not shown). The 5-F-Trp-labeled α-LA has the same specific activity as the unlabeled protein in stimulating the galactose transferase reaction using an in vitro enzymatic assay, whereas the 4-F-Phe-labeled α-LA has approximately 40% of the activity of the unlabeled protein (data not shown). A moderate reduction in enzymatic activity has been observed in other fluorine-labeled proteins, which usually arises because a labeled residue is located near the active site, rather than because of a severe structural perturbation

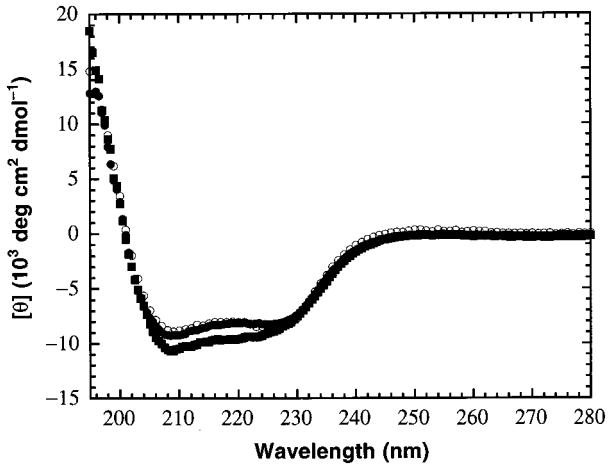

FIGURE 2: Far-UV CD spectra of 4-F-Phe- (●) and 5-F-Trp-labeled (■) α-LA at pH 8.5, 4 °C. The spectrum of the unlabeled protein (○) is included for comparison.

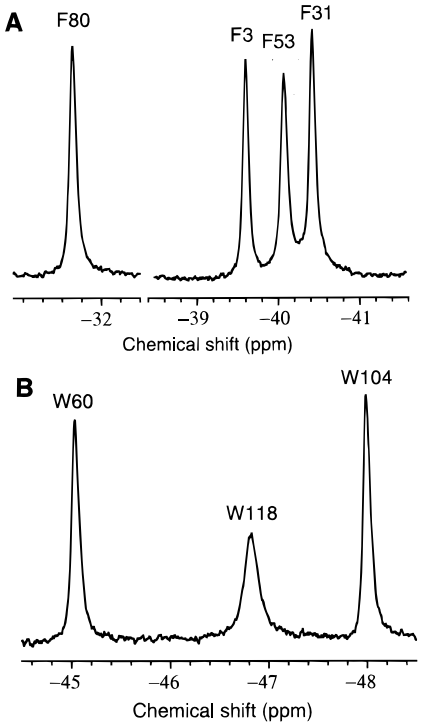

FIGURE 3: ^{19}F-NMR spectra of fluorine-labeled α-LA in the native state at pH 8.5, 25 °C. (A) 4-F-Phe-labeled α-LA (wild-type). (B) 5-F-Trp-labeled α-LA (wild-type). The residues were assigned by site-directed mutagenesis, replacing each phenylalanine or tryptophan by other type of residues. F3Y, F31Y, and F80L were used to assign the spectrum of phenylalanine; W60F, W104F, and W118F were used to assign the spectrum of tryptophan. F53 cannot be replaced by any other type of residues, therefore, the resonance of F53 was assigned by excluding other possibilities.

(*24*). Taken together, these data suggest that the fluorine-labeled α-LA maintains a nativelike structure similar to that of the unlabeled protein.

The ^{19}F-NMR spectra of 4-F-Phe- and 5-F-Trp-labeled α-LA under native conditions are shown in Figure 3. These spectra exhibit several well-resolved peaks with similar intensities, each corresponding to one phenylalanine or tryptophan in the protein. The identity of these peaks was assigned by site-directed mutagenesis, with substitutions of each phenylalanine by leucine or tyrosine and each tryptophan by phenylalanine. In the molten globule state, the

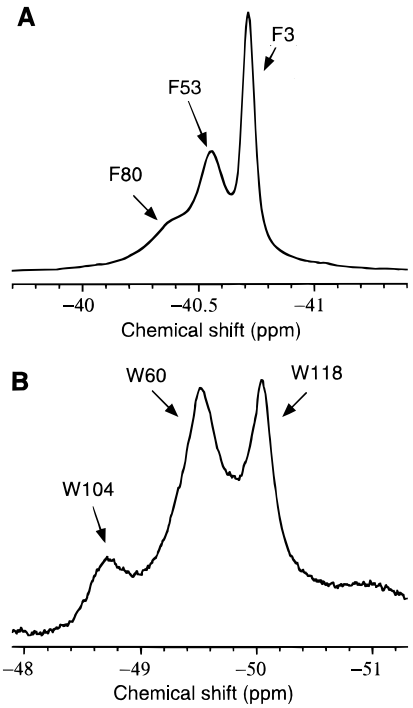

FIGURE 4: ^{19}F-NMR spectra of fluorine-labeled α-LA in the molten globule state at pH 2.0, 35 °C. (A) 4-F-Phe-labeled α-LA (mutant F31Y). (B) 5-F-Trp-labeled α-LA (wild-type). The following mutant proteins were used in the spectral assignment. F3Y/F31Y and F31Y/F80L were used to assign the spectrum of phenylalanine; W60F, W104F, and W118F were used to assign the spectrum of tryptophan.

chemical shift dispersion seen in the native state is largely collapsed, resulting in significant spectral overlap. To obtain the highest spectral resolution, we recorded the ^{19}F-NMR spectra at several different temperatures. The best resolution was obtained at 35 °C. Even at this temperature, the four phenylalanine resonances in the molten globule state of α-LA still cannot be completely resolved. After the spectrum of the wild-type α-LA was compared with that of mutant proteins used for spectral assignment, we decided to adopt the mutant F31Y as a pseudo wild-type for solvent accessibility and side-chain dynamics studies in the A-state. F31 is 100% exposed to solvent in the native α-LA (see below), therefore, we expected that it would also be exposed in the molten globule. The remaining three phenylalanine resonances in F31Y can be assigned unambiguously by site-directed mutagenesis (Figure 4A). The assignment of tryptophan residues were carried out using a similar method, by comparing the spectrum of the wild-type α-LA with that of mutant proteins with single tryptophan to phenylalanine substitutions (Figure 4B).

In the unfolded state, the chemical shift dispersion is even less than that in the molten globule state; however, since the lines are also sharper, the relative spectral resolution remains approximately the same. We observed three peaks in 4-F-Phe- labeled α-LA and two peaks in 5-F-Trp-labeled α-LA (Figure 5). In each case, one of the peaks corresponds to the overlap of two spectral resonances, which were assigned by examining the spectra of mutant proteins.

Solvent Accessibility. The solvent accessibility of fluorines located on aromatic side chains was measured by the solvent induced isotope shift (SIIS) effect (*40*, *41*). This effect can be described as the following. If a fluorine atom is solvent exposed, its chemical shift value will depend linearly on the

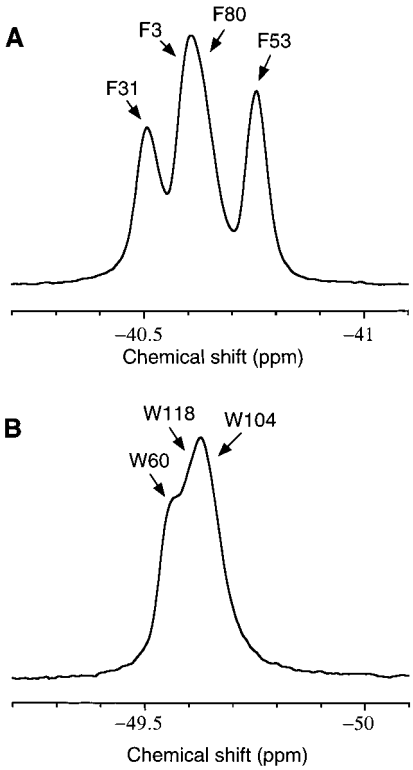

FIGURE 5: ^{19}F-NMR spectra of fluorine-labeled α-LA in the unfolded state at pH 2.0, 25 °C in 6 M GuHCl. (A) 4-F-Phe-labeled α-LA (wild-type). (B) 5-F-Trp-labeled α-LA (wild type). The assignment was obtained using the same set of mutant proteins as used for the native and molten globule state.

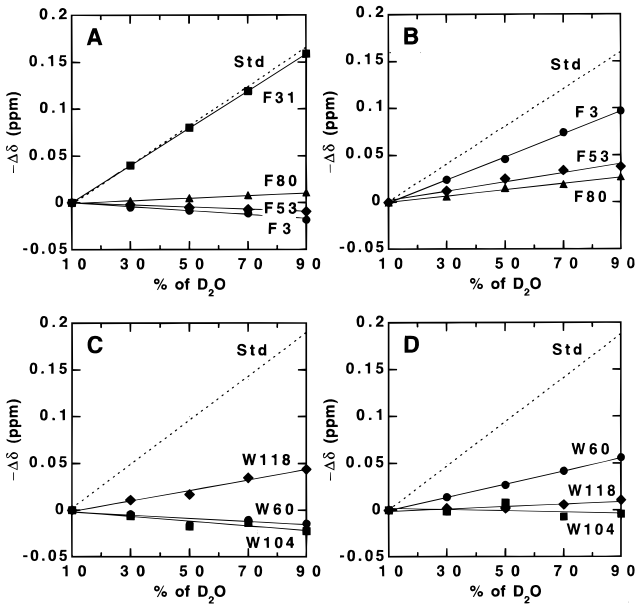

FIGURE 6: The chemical shift change of fluorine (Δδ) as a function of the percentage of D_2O in solvent, relative to the chemical shift value measured in a buffer containing 10% D_2O. (A) 4-F-Phe-labeled α-LA in the native state. (B) 4-F-Phe-labeled α-LA (mutant F31Y) in the molten globule state. (C) 5-F-Trp-labeled α-LA in the native state. (D) 5-F-Trp-labeled α-LA in the molten globule state. The lines represent the best linear least-squares fitting. The slopes of these lines were reported in Table 1. The dotted lines represent the chemical shift changes observed for the SIIS standards.

percentage of D_2O in the solvent. This effect will not be observed if the fluorine atom is buried in the interior of a protein. To calculate the percentage of solvent accessibility, we divide the chemical shift difference (Δδ) between 0 and 100% D_2O observed in a labeled protein by that of an SIIS standard, for which the side chain is completely exposed to solvent. We synthesized two protected forms of fluorine-labeled amino acid, *N*-acetyl-4-fluorophenylalanine methylester or *N*-acetyl-5-fluorotryptophanamide, which have the same side chain as 4-F-Phe and 5-F-Trp. The chemical shift values of these compounds are independent of pH (data not shown), therefore, they can be used as SIIS standards at both neutral and acidic conditions.

Figure 6 shows the results of out solvent accessibility measurements with the numerical values summarized in Table 1. First, we determined the percentage of solvent accessibility for each fluorine in labeled α-LA under native conditions. This serves as a control and calibration of our method. The solvent accessibility measured by the SIIS effect generally agrees well with those calculated from the X-ray crystal structure. In several cases, for residues that are completely buried, we observed a small chemical shift change toward an opposite direction, which would translate into a negative solvent accessibility. This could be due to a secondary effect as the magnitude of this effect usually does not exceed 10%. Both the NMR and X-ray data indicate that the fluorine atoms in F3, F53, F80, W60, and W104 are completely buried, while that in W118 is partially exposed to solvent and that in F31 is completely exposed to solvent. This agreement allows us to use the SIIS effect to measure

Table 1: Solvent Accessibility of Fluorine in 4-F-Phe- and 5-F-Trp-labeled α-LA[a]

	native state (pH 8.5, 1 mM Ca^{2+})			molten globule state (pH 2.0)	
residue	chemical shift difference ($\delta^{0\%D_2O}-\delta^{100\%D_2O}$) (ppm)	solvent accessibility of fluorine (%)	solvent accessibility of Cζ in Phe or Cζ3 in Trp as derived from the crystal structure (%)	chemical shift difference ($\delta^{0\%D_2O}-\delta^{100\%D_2O}$) (ppm)	solvent accessibility of fluorine (%)
F3	0.021	−10.5[b]	2.0	−0.122	61.0
F31	−0.199	99.3	100	n.a.[c]	n.a.
F53	0.011	−5.5	0	−0.049	24.5
F80	−0.014	7.0	0	−0.034	17.0
W60	0.017	−7.4	0	−0.07	30.4
W104	0.025	−10.8	0	0.007	−3.0
W118	−0.056	24.3	19	−0.013	5.7

[a] The percentage solvent accessibility was calculated by dividing the chemical shift difference ($\delta^{0\%D_2O}-\delta^{100\%D_2O}$) measured in a protein with that of a small molecule SIIS standard for which the side chain is completely exposed to solvent. The $\delta^{0\%D_2O}-\delta^{100\%D_2O}$ for N-acetyl-4-fluorophenylalanine methylester (SIIS standard for 4F-Phe) and N-acetyl-5-fluorotryptophanamide (SIIS standard for 5F-Trp) were −0.20 and −0.23 ppm respectively. [b] The small negative accessibility is probably due to a secondary effect, which usually does not exceed 10%. [c] The solvent accessibility cannot be determined for F31 due to spectral overlap.

Table 2: Linewidth of ^{19}F resonances in 4-F-Phe-labeled α-LA under Different Conditions

	native state (pH 8.5, 1 mM Ca^{2+})		molten globule state (pH 2.0)		unfolded state (pH 2.0, 6 M GuHCl)	
residue	25 °C (Hz)	35 °C (Hz)	25 °C (Hz)	35 °C (Hz)	25 °C (Hz)	35 °C (Hz)
F3	47	50	44	29	25[c]	25[c]
F31	39	55	n.a.[a]	n.a.[a]	23	23
F53	38	38	119	59	22	22
F80	38	37	177	110	27[c]	27[c]
W60	41	43	293	201	29[d]	30[d]
W104	38	42	112[b]	114[b]	40[d]	33[d]
W118	77	83	172	109	43[d]	36[d]

[a] The line width of 4-F-Phe-labeled α-LA in the molten globule state was obtained using the mutant F31Y. Thus, the value for F31 is not available. [b] These values are less reliable due to a weak signal. [c] These values were obtained using the mutant F31Y/F3Y and F31Y/F80L in order to resolve overlapped peaks. [d] These values were obtained using the mutant W104F and W118F in order to resolve overlapped peaks.

the side-chain solvent accessibility in the molten globule state for which a high-resolution structure is not available.

In the molten globule state, many of the fluorine atoms that were buried in the native state of α-LA are now partially accessible to solvent, but the solvent accessibility is generally less than 50%. The fluorine atoms in W104 and W118, located near the C-terminus of α-LA, remain significantly buried in the molten globule. The solvent accessibility of W118 in the molten globule state is actually less than that in the native state. The N-terminal region of the α-LA molten globule apparently is more exposed to solvent than the C-terminal region, because the fluorine atom in F3 has the highest solvent accessibility. The three aromatic residues located in the β-sheet domain have intermediate values of solvent accessiblity, suggesting that the β-sheet domain is not completely unfolded even though it does not prefer a nativelike tertiary topology as indicated by disulfide exchange studies (*21*).

In the unfolded state, all fluorine atoms have the same value of $\Delta\delta$ between 0 and 100% D_2O, which suggests that they are equally accessible to solvent, as expected for a polypeptide chain with an extended structure (data not shown). It is difficult to measure the solvent accessibility quantitatively in 6 M GuHCl because a significant portion of the solvent volume is occupied by the denaturant. Nevertheless, the value of $\Delta\delta$ for the unfolded α-LA is essentially the same as that of the SIIS standard measured under the same condition, suggesting that the side chains are fully accessible to solvent.

Side-Chain Dynamics. The spectral line width of fluorine is inversely proportional to the apparent transverse relaxation time, T_2^*, which is extremely sensitive to slow conformational fluctuations in the protein. Table 2 lists the observed line widths for 4-F-Phe- and 5-F-Trp-labeled α-LA under different conditions. A relatively narrow line width was observed in both the native and unfolded state. In the native state at 25 °C, three of the four phenylalanine resonances (F31, F53, and F80) and two of the three tryptophan resonances (W60 and W104) have the same line width, suggesting that the majority of aromatic side chains are fixed by the tertiary structure of the protein and the molecule tumbles as a rigid body in solution. Compared to these residues, the resonance of F3 exhibits a slight line broadening and the resonance of W118 exhibits a significant line broadening. This could be caused by ring flipping motions or other types of local fluctuations that are prevalent in the native state. At 35 °C, the line width of all residues, except for F31 remains essentially the same. The line width of F31 increases significantly between 25 and 35 °C, which occurs prior to the thermal unfolding of α-LA.

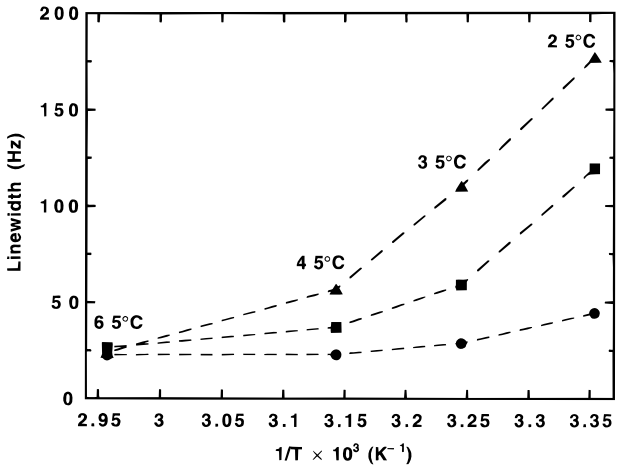

FIGURE 7: The spectral line width of fluorine in F3 (●), F53 (■), and F80 (▲) as a function of temperature in the molten globule state of α-LA. The line width was measured from the spectra of F31Y, using a least-squares fitting program implemented in FELIX.

In the unfolded state, all phenylalanine residues and one tryptophan residue (W60) have the same line width, which is independent of temperature, suggesting that they are located in the unstructured regions. Interestingly, the resonance of the other two tryptophan residues, W104 and W118, exhibit a slight line broadening. The line width of these residues become narrower at a higher temperature, suggesting that the C-terminal region of α-LA has a residual structure even in 6 M GuHCl.

The most striking result was observed in the molten globule state. First, the ^{19}F resonances are significantly broader in the molten globule state than in the native or unfolded states, suggesting that the α-LA molten globule is particularly rich in low-frequency fluctuations. Second, in the molten globule state, different ^{19}F resonances exhibit significantly different line widths, indicating that each individual residue has an unique motional environment and that the side-chain dynamics are highly heterogeneous. The residue F3 has the narrowest line width among all phenylalanine residues, and the residues W104 and W118 have relatively narrow line widths among all tryptophan residues. These residues are located near the N- and C-terminus of the protein, suggesting that these regions may have a higher degree of motional freedom than the rest of the molecule. The temperature dependence of line width in the molten globule state is also different from that in the native or unfolded states. With the exception of W104, for which the line width cannot be determined accurately due to a weak signal intensity, the line width of all other residues exhibits a strong temperature dependence. In Figure 7, the line widths of three phenylalanine residues were plotted as a function of 1/*T*. As temperature increases, the line width decreases, indicating that additional motional processes take place at higher temperatures. At 65 °C, all 4-F-Phe resonances have essentially the same line width, which is identical to that of the unfolded protein.

DISCUSSION

The first goal of this study was to provide a residue-specific, quantitative measurement of solvent accessibility in the α-LA molten globule. In the absence of a high resolution structure, the solvent accessibility of residues that are normally buried in the interior of a protein provides a sensitive assay for structural integrity. A number of studies have been carried out to address the solvent accessibility of residues in the α-LA molten globule. These studies yielded somewhat conflicting conclusions, perhaps because different methods were used in different studies and perhaps also because most of these methods do not provide residue-specific information. For example, measurements of the high-affinity interactions between the α-LA molten globule and the fluorescent dye, 1-anilino-naphthalene-8-sulfonic acid (ANS), suggest that the molten globule has a considerable amount of solvent exposed hydrophobic surface area (*42*). Thermodynamic compressibility and sound absorption studies also indicate that a large number of water molecules penetrate into the interior of the α-LA molten globule (*43*). On the other hand, more recent NMR relaxation dispersion studies show that the properties of water molecules associated with the α-LA molten globule are similar to those associated with the native protein (*44*). Fluorescence quenching studies performed at low pH suggest that the tryptophan residues in the A-state of α-LA are partially buried (*45*, *46*), whereas the same studies performed on the kinetic folding intermediate of α-LA suggest that the tryptophan residues are fully exposed to solvent (*7*). One method that does give residue-specific information is the free radical solvent perturbation method, which has been applied to the heat denatured state of α-LA (*47*). However, in this experiment, only a few residues from the denatured protein can be assigned. Moreover, the heat denatured state of α-LA has been shown to be essentially unfolded, and not equivalent to the molten globule state by recent thermodynamic analysis (*48*−*50*).

Our solvent accessibility measurement is based on chemical shift changes of fluorine caused by an isotope effect induced by deuterated water. Therefore, no exogenous compound is added to the protein solution. The chemical shift values can usually be measured more accurately than signal intensity. In our case, we estimated the error in our chemical shift measurements is less than 5%. However, when calibrated using the native state of α-LA, the difference between the value of solvent accessibility measured by NMR and those calculated from the X-ray crystal structure is approximately 10%. This could happen for a number of reasons. First, in the X-ray crystal structure, the solvent accessibility was not calculated for the fluorine atom, instead, it was calculated for the nearest carbon atom in the PDB coordinate. Second, substitution of hydrogen by fluorine could slightly perturb the local structure around that atom. Third, although the chemical shift changes are dominated by the SIIS effect, there may be a small secondary effect, which could explain why sometimes a small, but negative accessibility is observed for a completely buried residue. A possible mechanism for this effect is that D_2O often stabilizes proteins. As a result, the molecule may become slightly more compact and have a slightly reduced solvent accessible surface area.

We find that most of the aromatic residues in the α-LA molten globule have a greater solvent accessibility than in the native state, as expected for a partially folded structure with an increasing amount of solvent exposed hydrophobic surface area. One particularly interesting observation is that the solvent accessibility of W118 in the molten globule state is actually less than that in the native state. In an earlier study, W118 was found to be one of the most important residues for stabilizing the nativelike tertiary topology in α-LA molten globule (*51*, *52*). Another residue that has low solvent accessibility is W104. The importance of this residue in terms of aromatic packing interactions has been discussed previously (*46*) and the side chain amide of W108, a corresponding residue in equine lysozyme, was found to exchange slowly in the lysozyme molten globule (*53*). Our data suggest that the β-sheet domain of α-LA is not completely unfolded in the molten globule state because the aromatic residues in this domain (F53, W60, and F80) are only partially accessible to solvent. The β-sheet domain likely retains a nonspecific compact structure as first indicated by solution X-ray scattering studies (*54*).

For those residues that have a fractional solvent accessibility, it is possible that the side-chain samples multiple conformations, spending a fraction of its time in a solvent accessible environment, while spending the rest of its time buried inside the protein. It is also possible that for some of the molecules, water penetrates into the interior of the protein, while for other molecules, water remains outside. Since we have never observed two different populations, the solvent accessible and inaccessible states must be in rapid equilibrium. The difference between these two scenarios is somewhat semantic and cannot be distinguished by the current experiment.

In this study, the NMR line width of fluorine was used as a probe of the conformational flexibility and side-chain dynamics in α-LA under different conditions. Several relaxation mechanisms may contribute to the observed line width. Fluorine participates in dipolar interactions and has a large chemical shift anisotropy. The relatively narrow line width observed in both the native and unfolded states are likely due to the relaxation effects originating from these interactions. The line width in the native state is generally slightly broader than that in the unfolded state, because in the native state the molecule tumbles as a rigid body in solution, whereas in the unfolded state different parts of the molecule can move independently, resulting in a shorter motional correlation time. The temperature dependence of the line width in the native and unfolded states are also expected to go toward different directions. In the native state, most of the side chains are fixed by the tertiary structure of the protein. A local fluctuation that allows a side chain to sample different conformations with different isotropic chemical shift values will result in a line broadening. Our studies suggest that F3 and W118 are involved in such fluctuations. These residues are near the 6−120 disulfide bond, which is strained and may promote local fluctuations. Since the amplitude of such fluctuation typically increases with temperature, we expect that the line width of certain residues in the native state may also increase with temperature. Such an effect was observed for F31, and to a lesser extent, for F3 and all tryptophans. In the unfolded state, a narrow line width was observed because the residues are in

a relatively free motional environment. A residual structure, which may restrict the motional averaging, will lead to a line broadening. At a higher temperature, many of the local structures will dissolve and we expect that the line width will decrease with increase of temperature. Such an effect was observed for residues W104 and W118.

The most interesting result of side-chain dynamics was observed in the molten globule state. The line widths of F53, F80, W60, W104, and W118 in the α-LA molten globule are all 2−5 times broader than that in the native state. Under our experimental conditions, the α-LA molten globule is a monomer and has a radius of gyration only slightly larger than that of the native protein (*55*). Therefore, we expect that the α-LA molten globule and native α-LA have similar tumbling correlation times. This argument suggests that the relaxation contributions from dipolar interactions and chemical shift anisotropy for the α-LA molten globule should also be similar to that for the native protein. The line broadening observed in the α-LA molten globule likely reflects the balance of two opposing effects. First, the entropic contribution will drive the side chains to sample multiple conformations. Second, the motions in the α-LA molten globule must be highly restricted by its compact structure. Thus, from a dynamic point of view, the α-LA molten globule has properties intermediate of the native and unfolded states. This description is consistent with the energy landscape theory in which the conformational space covered by the molten globule is significantly smaller than that of the unfolded state but is significantly larger than that of the native state. Since only a single resonance is observed for each residue and the line width decreases with an increase of temperature, the rate of interconversion between multiple conformations in the α-LA molten globule must be greater, although not much greater, than the chemical shift differences between these conformations. Using the chemical shift difference between the native and unfolded states as an upper limit, we can estimate that time scale of the conformational dynamics in the α-LA molten globule is in the range of 10−100 μs. This type of slow fluctuations can be considered as a hallmark of the molten globule state.

Another unique feature of the α-LA molten globule is that different residues exhibit significantly different line widths, suggesting that they have different motional properties and environment. Thus, the side-chain dynamics in the α-LA molten globule are highly heterogeneous. Residues located at the N- and C-terminal regions have relatively narrow line widths, suggesting that these residues may experience a more complete motional averaging than residues located in the middle of the protein sequence. It is intriguing that the N- and C-terminal regions of α-LA are located close to each other in the native structure and that some of the residues in these regions, e.g., F3 and W118, were found to be mobile in the native state. Thus, the side-chain dynamics in the molten globule and native states seem to be correlated.

Finally, one important question in the protein-folding field is what molecular interactions stabilize the partially folded intermediate. For native proteins, the thermodynamic stability is mostly originated from hydrophobic and var der Waals interactions, with ionic and hydrogen bonding interactions contribute more toward structural specificity than stability (*56*, *57*). In case of α-LA, our data suggest that greater than 50% of the buried surface area of aromatic residues in the native protein remains buried in the molten globule, and the stability of the molten globule is approximately half of the value for the native protein (*58*, *59*), thus, hydrophobic and van der Waals interactions mediated by the contact of inaccessible surface area can potentially account for all the stability of the α-LA molten globule. This interpretation is consistent with previous alanine scanning mutagenesis studies, showing that the removal of hydrophobic side chains in α-LA molten globule has the largest effect on formation of the nativelike tertiary fold (*52*). In addition to these direct interactions, a solvent-separated hydrophobic effect may also contribute to the stability of the α-LA molten globule (*7*, *60*).

ACKNOWLEDGMENT

We thank Jackie Wilce and Kurt Hoffacker for help on organic synthesis, Lawrence McIntosh and David Waugh for providing bacterial strains, Gregory Mullen for using the NMR instrument, Walfrido Antuch, Mark Maciejewski, and Borlan Pan for help with the NMR experiments and computer programs, Karen LeCuyer for using the CARY 1E spectrophotometer, Suroopa Charkaborty for performing α-LA activity assays, John Glynn and Ming Li for DNA sequencing, Dan Minor and Peter Kim for discussion and encouragement, and Dan Minor, Lawren Wu, and Peter Setlow for critical reading of the manuscript. Mass spectrometry analysis was provided by the Washington University Mass Spectrometry Resource with support from the NIH National Center for Research Resources (Grant No. P41RR0954) and by Igor Kaltashov at the Mass Spectrometry Center of University of Massachusetts at Amherst.

REFERENCES

1. Kuwajima, K. (1989) *Proteins: Struct., Funct., Genet. 6*, 87−103.
2. Dobson, C. M. (1994) *Curr. Biol. 4*, 636−640.
3. Ptitsyn, O. B. (1996) *Adv. Prot. Chem. 47*, 83−229.
4. Ikeguchi, M., Kuwajima, K., Mitani, M., and Sugai, S. (1986) *Biochemistry 25*, 6965−6972.
5. Jennings, P. A., and Wright, P. E. (1993) *Science 262*, 892−896.
6. Balbach, J., Forge, V., van Nuland, N. A. J., Winder, S. L., Hore, P. J., and Dobson, D. M. (1995) *Nat. Struct. Biol. 2*, 865−870.
7. Arai, M., and Kuwajima, K. (1996) *Fold. Des. 1*, 275−287.
8. Raschke, T. M., and Marqusee, S. (1997) *Nat. Struct. Biol. 4*, 298−304.
9. Seeley, S. K., Weis, R. M., and Thompson, L. K. (1996) *Biochemistry 35*, 5199−5206.
10. Gursky, O., and Atkinson, D. (1996) *Proc. Natl. Acad. Sci. U.S.A. 93*, 2991−2995.
11. Carroll, A. S., Gilbert, D. E., Liu, X., Cheung, J. W., Michnowicz, J. E., Wagner, G., Ellenberger, T. E., and Blackwell, T. K. (1997) *Genes Dev. 11*, 2227−2238.
12. Lo, M. C., Ha, S., Pelczer, I., Pal, S., and Walker, S. (1998) *Proc. Natl. Acad. Sci. U.S.A. 95*, 8455−8460.
13. Zhang, J., and Matthews, C. R. (1998) *Biochemistry 37*, 14881−14890.
14. Craig, S., Hollecker, M., Creighton, T. E., and Pain, R. H. (1985) *J. Mol. Biol. 185*, 681−687.
15. Lim, W. A., Farruggio, D. C., and Sauer, R. T. (1992) *Biochemistry 31*, 4324−4333.
16. Zhang, B., and Peng, Z.-y. (1996) *J. Biol. Chem. 271*, 28734−28737.
17. Booth, D. R., Sunde, M., Bellotti, V., Robinson, C. V., Hutchinson, W. L., Fraser, P. E., Howkins, P. N., Dobson, C.

M., Radford, S. E., Blake, C. C. F., and Pepys, M. B. (1997) *Nature 385*, 787−793.
18. Smith, L. J., Dobson, C. M., and van Gunsteren, W. F. (1999) *J. Mol. Biol. 286*, 1567−1580.
19. Kuwajima, K. (1996) *FASEB J. 10*, 102−109.
20. Peng, Z.-y., and Kim, P. S. (1994) *Biochemistry 33*, 2136−2141.
21. Wu, L. C., Peng, Z.-y., and Kim, P. S. (1995) *Nat. Struct. Biol. 2*, 281−286.
22. Schulman, B. A., Redfield, C., Peng, Z.-y., Dobson, C. M., and Kim, P. S. (1995) *J. Mol. Biol. 253*, 651−657.
23. Sykes, B. D., and Hull, W. E. (1978) *Methods Enzymol. 49*, 270−295.
24. Gerig, J. T. (1994) *Prog. Nucl. Magn. Reson. Spectrosc. 26*, 293−370.
25. Frieden, C., Hoeltzli, S. D., and Ropson, I. J. (1993) *Protein Sci. 2*, 2007−2014.
26. Hoeltzli, S. D., and Frieden, C. (1996) *Biochemistry 35*, 16843−16851.
27. Sun, Z. Y., Pratt, E. A., Simplaceanu, V., and Ho, C. (1996) *Biochemistry 35*, 16502−16509.
28. Danielson, M. A., and Falke, J. J. (1996) *Annu. Rev. Biophys. Biomol. Struct. 25*, 163−195.
29. Peng, Z.-y., Wu, L. C., and Kim, P. S. (1995) *Biochemistry 34*, 3248−3252.
30. Kunkel, T. A., Robert, J. D., and Zakour, R. A. (1987) *Methods Enzymol. 154*, 367−382.
31. Muchmore, D. C., McIntosh, L. P., Russell, C. B., Anderson, D. E., and Dahlquist, F. W. (1989) *Methods Enzymol. 177*, 44−73.
32. Waugh, D. S. (1996) *J. Biomol. NMR 8*, 184−192.
33. Kim, H.-W., Perez, J. A., Ferguson, S. J., and Campbell, I. D. (1990) *FEBS Lett. 272*, 34−36.
34. Edelhoch, H. (1967) *Biochemistry 6*, 1948−1954.
35. Wong, C. Y., and Eftink, M. R. (1998) *Biochemistry 37*, 8938−8946.
36. Fitzgerald, D. K., Colvin, B., Mawal, R., and Ebner, K. E. (1970) *Anal. Biochem. 36*, 43−61.
37. Kabsch, W., and Sander, C. (1983) *Biopolymers 22*, 2577−2637.
38. Lee, B., and Richards, F. M. (1971) *J. Mol. Biol. 55*, 379−400.
39. Dettman, D. E. (1984), Ph.D. thesis, The University of Alberta.
40. Hansen, P. E., Dettman, H. D., and Sykes, B. D. (1985) *J. Magn. Reson. 62*, 487−496.
41. Rule, G. S., Pratt, E. A., Simplaceanu, V., and Ho, C. (1987) *Biochemistry 26*, 549−556.
42. Semisotnov, G. V., Rodionova, N. A., Razgulyaev, O. I., Uversky, V. N., Gripas, A. F., and Gilmanshin, R. I. (1991) *Biopolymers 31*, 119−128.
43. Kharakoz, D. P., and Bychkova, V. (1997) *Biochemistry 36*, 1882−1890.
44. Denisov, V. P., Jonsson, B.-H., and Halle, B. (1998) *Nat. Struct. Biol. 6*, 253−260.
45. Lala, A. K., and Kaul, P. (1992) *J. Biol. Chem. 267*, 19914−19918.
46. Chyan, C.-L., Wormald, C., Dobson, C. M., Evans, P. A., and Baum, J. (1993) *Biochemistry 32*, 5681−5691.
47. Improta, S., Molinari, H., Pastore, A., Consonni, R., and Zetta, L. (1995) *Eur. J. Biochem. 227*, 87−96.
48. Pfeil, W. (1987) *Biochim. Biophys. Acta 911*, 114−116.
49. Xie, D., Bhakuni, V., and Freire, E. (1991) *Biochemistry 30*, 10673−10678.
50. Griko, Y. V., Freire, E., and Privalov, P. L. (1994) *Biochemistry 33*, 1889−1899.
51. Wu, L. C., and Kim, P. S. (1998) *J. Mol. Biol. 280*, 175−182.
52. Song, J., Bai, P., Luo, L., and Peng, Z.-y. (1998) *J. Mol. Biol. 280*, 167−174.
53. Morozova-Roche, L. A., Arico-Muendel, C. C., Haynie, D. T., Emelyanenko, V. I., Van Dael, H., and Dobson, C. M. (1997) *J. Mol. Biol. 268*, 903−921.
54. Kataoka, M., Kuwajima, K., Tokunaga, F., and Goto, Y. (1996) *Protein Sci. 6*, 422−430.
55. Gast, K., Zirwer, D., Welfle, H., Bychkova, V. E., and Ptitsyn, O. B. (1986) *Int. J. Biol. Macromol. 8*, 231−236.
56. Dill, K. A. (1990) *Biochemistry 29*, 7133−7155.
57. Voet, D., and Voet, J. G. (1995) *Biochemistry*, pp174−179, John Wiley & Sons, New York.
58. Kuwajima, K., Nitta, K., Yoneyama, M., and Sugai, S. (1976) *J. Mol. Biol. 106*, 359−373.
59. Kuwajima, K. (1977) *J. Mol. Biol. 114*, 241−258.
60. Uchiyama, H., Perez-Prat, E. M., Watanabe, K., Kumagai, I., and Kuwajima, K. (1995) *Protein Eng. 8*, 1153−1161.
61. Acharya, K. R., Ren, J., Stuart, D. I., and Phillips, D. C. (1991) *J. Mol. Biol. 221*, 571−581.
62. Koradi, R., Billeter, M., and Wuthrich, K. (1996) *J. Mol. Graph. 14*, 51−55.

BI992056F