NMR Structural Study of Two-Disulfide Variant of Hen Lysozyme: 2SS[6−127, 30−115]−A Disulfide Intermediate with a Partly Unfolded Structure

Yasuo Noda,[‡] Atsushi Yokota,[§] Daisuke Horii,[‡] Takeshi Tominaga,[‡] Yoshiaki Tanisaka,[‡] Hideki Tachibana,[||] and Shin-ichi Segawa*[,‡]

Department of Physics, School of Science, Kwansei Gakuin University, Sanda, 669-1337, Japan, National Institute of Advanced Industrial Science and Technology (AIST), Ikeda, Osaka, 563-8577, Japan, and Department of Biology, Faculty of Science, and Graduate School of Science and Technology, Kobe University, Kobe, 657-8501, Japan

Received June 26, 2001; Revised Manuscript Received October 17, 2001

ABSTRACT: The ^{15}N-labeled recombinant hen lysozyme and two species of two-disulfide variants, denoted as 2SS[6−127, 30−115] and 2SS[64−80, 76−94], were studied by means of NMR spectroscopy. The former variant contains two disulfide bridges in the α-domain, while the latter has one disulfide bridge in the β-domain and the other one at the interface between two domains. Resonance assignments were performed using 3D TOCSY−HSQC and NOESY−HSQC spectra. The ^{15}N−^{1}H−HSQC spectrum of 2SS[6−127, 30−115] was similar to that of recombinant lysozyme as a whole, although a number of cross-peaks disappeared. On the other hand, the HSQC spectrum of 2SS[64−80, 76−94] was characteristic of unfolded proteins. The structure of 2SS[6−127, 30−115] was thoroughly examined on the basis of NOE contacts determined by NMR spectroscopy. The structure of the α-domain was quite similar to that of authentic lysozyme, while the β-domain was largely unstructured. However, NMR data clearly demonstrated that some residual structures exist in the β-domain. The β1 and β2 strands were maintained stably as an antiparallel β-sheet. In addition, the residues 55 and 56 were located in the vicinity of the end of the B-helix. Further, the C-helix was properly set with side-chains of I88, V92, K96, and V99 facing toward the hydrophobic core in the α-domain. These residual structures inherent in the amino acid sequence were evaluated concerning the folding process of lysozyme. Our experiments imply that the establishment of the backbone conformation ranging from residues 76−99 plays a key role in attaining the cooperativity between two domains required for the folding transition.

It is important for the elucidation of the mechanism of protein folding to examine the structure of partially folded intermediates. But the folding process is a very cooperative phenomenon and is accomplished rapidly, and therefore, folding intermediates are not suitable for study at equilibrium. However, it is possible to trap disulfide intermediates during the disulfide regeneration process. Fully reduced protein attains native disulfide pairings through one-, two-, and three-disulfide intermediates, etc., during the folding process. Analogues of disulfide regeneration intermediates can be prepared as variants lacking a specific disulfide bridge by amino acid substitution for Cys residue. The disulfide regeneration intermediates of the bovine pancreatic trypsin inhibitor were studied thoroughly at equilibrium (*1*−*3*).

In general, the variants with increasing number of disulfide bridges have increasing amounts of ordered structure. Hen lysozyme contains four disulfide bridges; here we call them SS1 (Cys6−Cys127), SS2 (Cys30−Cys115), SS3 (Cys64−Cys80), and SS4 (Cys76−Cys94). SS1 and SS2 are located in the α-domain, SS3 is located in the β-domain, and SS4 is located at the interface between two domains. One-disulfide and three-disulfide variants of hen lysozyme have been studied mainly by means of CD spectroscopy (*4*−*7*). All species of three-disulfide variant had the nativelike structure, while all species of one-disulfide variant were largely unstructured. On the other hand, two species of two-disulfide variant, which were constructed and characterized recently, showed contrasting results. Judging from their CD spectra, one species, 2SS[6−127, 30−115], containing two disulfide bridges, Cys6−Cys127 and Cys30−Cys115, appeared to have a substantial amount of secondary and tertiary structures. By contrast, the other species, 2SS [64−80, 76−94], containing Cys64−Cys80 and Cys76−Cys94, showed lim-

* To whom correspondence should be addressed. Phone: +81-795 (65) 8482. Fax: +81-795 (65) 9077. E-mail: shsegawa@kwansei.ac.jp.

† The NMR facility in KGU was supported by a grant from the Ministry of Education, Culture, Sports, Science and Technology, Japan (the High-Tech Research Center Project).

‡ Kwansei Gakuin University.

§ AIST.

|| Kobe University.

[1] Abbreviations: CD, circular dichroism; DQF−COSY, double-quantum filtered correlation spectroscopy; HSQC, heteronuclear single-quantum coherence; metLYZ, recombinant hen lysozyme containing the extra N-terminal Met; NMR, nuclear magnetic resonance; NOE, nuclear Overhauser effect; NOESY, nuclear Overhauser effect spectroscopy; SS1, disulfide bridge between Cys6 and Cys127; SS2, Cys30−Cys115; SS3, Cys64−Cys80; SS4, Cys76−Cys94; TOCSY, total correlation spectroscopy; 2D and 3D, two- and three-dimensional; 2SS[6−127, 30−115], hen lysozyme with Cys64, Cys76, Cys80, and Cys94 replaced by Ala and two disulfide bridges between Cys6 and Cys127 and between Cys30 and Cys115; 2SS[64−80, 76−94], hen lysozyme with Cys6 replaced by Ser, Cys30, Cys115, and Cys127 replaced by Ala and two disulfide bridges between Cys64 and Cys80 and between Cys76 and Cys94; the one letter code for amino acid residues is also used to indicate the amino acid and the residue number.

10.1021/bi0113418 CCC: $22.00
Published on Web 01/24/2002

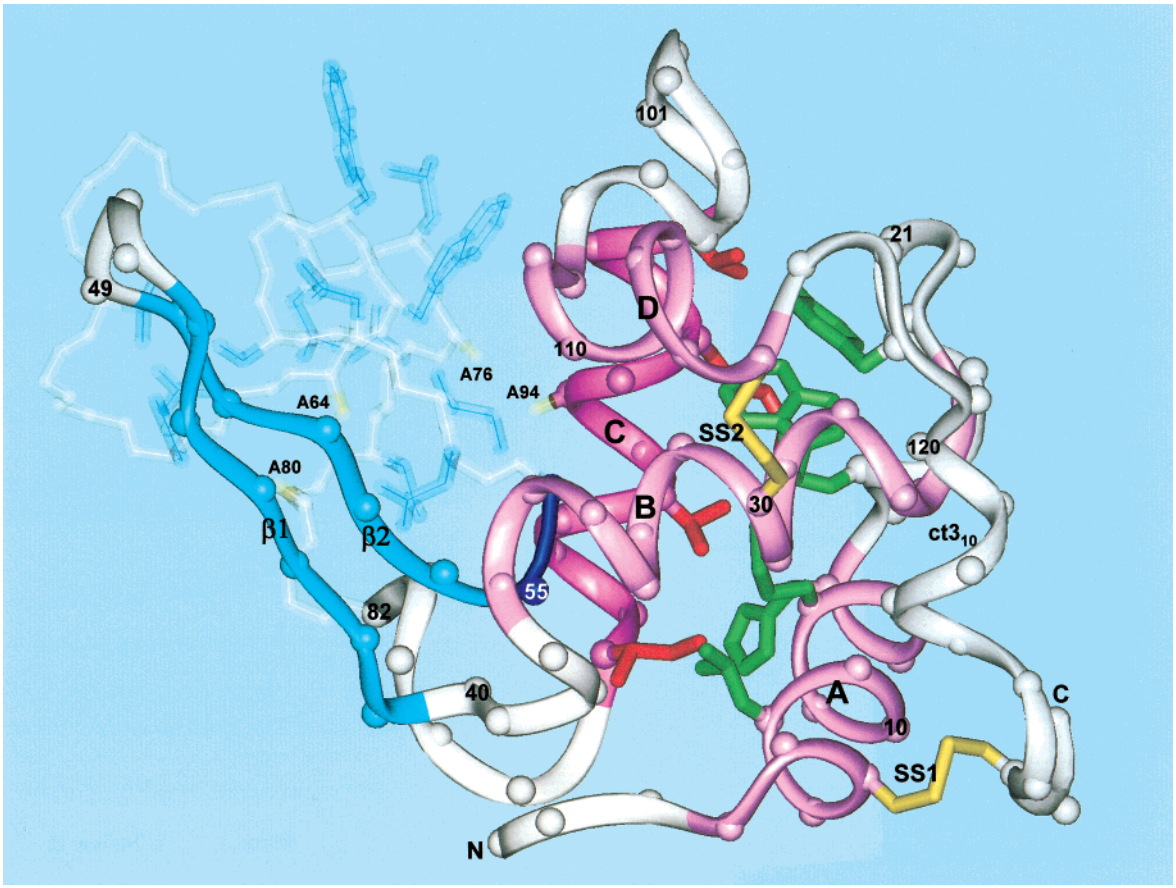

FIGURE 1: Ribbon diagram of 2SS variant of hen lysozyme: 2SS[6−127, 30−115], produced by Insight II on the basis of the X-ray crystallographic structure (PDB 6LYZ). A ribbon depicts the portion of polypeptide chain folded into the nativelike structure. Transparent sticks show the backbone and side-chains in the region where short- and long-range NOEs disappear; the backbone ranging from 57 to 81 is colored white and the side-chains blue. Balls on the ribbon indicate the position of C_α. Two disulfide bridges, C6−C127 and C30−C115, are colored yellow, four helices magenta, β1- and β2-strands blue, and I55 and L56 dark blue. The specific interactions exist between side-chains on C-helix (I88, V92, K96, and V99 colored red) and those within the N-terminal half of α-domain (L8, M12, H15, Y20, and W28 colored green). Yellow short sticks out of the backbone indicate the positions of Cys64, Cys76, Cys80, and Cys94 replaced by Ala.

ited amount of secondary and little tertiary structures (*8*). Thus, the two-disulfide variants correspond to the border between ordered and disordered states. Since we cannot acquire detailed information about structural changes in protein from their CD spectra, NMR spectra have been measured for these disulfide variants of hen lysozyme. Here we report the detailed results obtained for the variant: 2SS-[6−127, 30−115], compared with those for authentic lysozyme.

Since the recombinant hen lysozyme contains an extra Met residue at the N-terminus (here it is called metLYZ), at first, the ^{15}N-labeled metLYZ was studied as the standard reference. Resonance assignments were performed using 3D TOCSY−HSQC and NOESY−HSQC spectra, and then short-, medium-, and long-range NOE contacts were determined. The ^{15}N−^{1}H−HSQC spectrum of 2SS[6−127, 30−115] was similar to that of metLYZ as a whole, although a number of cross-peaks disappeared. On the other hand, the spectrum of 2SS[64−80, 76−94] was characteristic of unfolded proteins. It was found that the structure of the α-domain of 2SS[6−127, 30−115] is quite similar to that of metLYZ and can be maintained almost independently from the β-domain, only if both SS1 and SS2 exist. On the other hand, the interdomain cooperativity was indispensable to attaining the structure of β-domain even in the presence of both SS3 and SS4.

Figure 1 shows the ribbon diagram of 2SS[6−127, 30−115] produced from X-ray crystallographic structure (PDB 6LYZ), in which SS3 and SS4 are removed by alanine substitution. In 2SS[6−127, 30−115], the polypeptide segment ranging from 30 to 115 is quite free from the constraints due to disulfide bridges except around residues 30 and 115 so that β-domain was largely unstructured. However, NMR structural analysis clearly demonstrated that some residual structures exist in it. The β1 (residues 41−46) and β2 strands (50−54) were preserved stably as an antiparallel β-sheet, and the residues 55 and 56 were located in the vicinity of the end of the B-helix (residues 24−36). Further, the C-helix (residues 88−99) was properly set with side-chains of I88, V92, K96, and V99 embedded in the hydrophobic core in the α-domain. These facts indicate that the polypeptide segment (β1 and β2 strands or the C-helix) has a strong propensity to adopt such a residual structure inherent in the amino acid sequence. The role of these structures in the folding transition is discussed in conjuntion with the cooperativity between both domains.

MATERIALS AND METHODS

Two-Disulfide Variants of Hen Lysozyme. Details of the construction of the genes for the 2SS variants and identification of the disulfide linkage pattern in the reoxidized and purified 2SS variant proteins are described in the reference

(*8*). The 2SS variant gene which retains the Cys residues 6, 30, 115, and 127 was constructed by recombining a synthetic DNA fragment that coded for the region containing replacements, Cys64 → Ala, Cys76 → Ala, and Cys80 → Ala, with another DNA fragment that coded for a part of 3SS $\Delta 4_{Ala}$ gene (5) containing a replacement Cys94 → Ala. The other 2SS variant gene which retains Cys residues 64, 76, 80, and 94 was constructed in a similar way, in which the original Cys residues 6, 30, 115, and 127 were replaced by Ser, Ala, Ala, and Ala residues, respectively. Nucleotide sequences of the variant genes were confirmed with dideoxy sequencing. ^{15}N-labeled proteins were produced in the M9 medium that was supplemented with additives and contained 1.5 g/L [^{15}N]ammonium chloride (99.4% ^{15}N; Shoko Co. Ltd., Tokyo) as a sole nitrogen source. After the addition of IPTG to induce expression, the cultivation was continued for further 4 h at 40 °C before harvesting cells. The solubilization of inclusion bodies and purification of protein were carried out as described for unlabeled proteins (*4*, *5*, *7*).

For the oxidative refolding to produce 2SS[6−127, 30−115], in which the two disulfide bridges are formed in native pairings, the reduced form of the variant was oxidized in 100 mM Tris acetate, 1.0 mM EDTA, pH 7.85, 0.1 mM GSSG, 1mM GSH, at the protein concentration of 3.3 μM, in the presence of 30% glycerol and at 30 °C for 2 h. Likewise, 2SS[64−80, 76−94] was obtained by oxidation of its purified, reduced form in 15% glycerol and at 10 °C for 2 h, under otherwise the same conditions as used for 2SS[6−127, 30−15]. The reoxidized materials were purified by reversed-phase HPLC on a μ-Bondasphere C4 column (19 × 150 mm, Waters) with a linear gradient of acetonitrile from 32.5 to 37.5% in 50 min (100 mL). The peak fraction that corresponded to the 2SS variant of native disulfide linkage was recovered and freeze-dried. Only the fractions that showed a single peak without any skew in analytical scale rechromatography were used. Concentrations of the 2SS-variants of lysozyme were estimated by using A_{280} = 2.64 per mg/mL.

NMR Spectroscopy. The freeze-dried samples of 2SS variants of lysozyme were dissolved in 95% H_2O/5% D_2O. The pH was adjusted to 3.8 by adding HCl or NaOH. Concentrations of recombinant hen lysozyme (metLYZ) and 2SS[6−127, 30−115] were about 1.0 and 1.2 mM, respectively. The concentration of 2SS[64−80, 76−94] could not be increased more than 0.47 mM to avoid the protein aggregation. The two-dimensional (2D) ^{1}H−^{1}H NMR spectra were recorded at 600.13 MHz on a Bruker Avance 600 DRX spectrometer equipped with a Bruker/SGI workstation. DQF−COSY, NOESY (mixing time 150 ms), and HOHAHA (mixing time 50 ms) experiments were carried out using standard procedures (*9*, *10*) at different temperatures (298 K, 308 K). Solvent signal suppression was achieved by using the WATERGATE scheme (*11*). Typically, data were collected at 1024 × 2048 data points in t_1 and t_2 directions with 64 transients. 2D ^{15}N−^{1}H−HSQC spectra (*12*) were collected at 277 and 298 K with 1024 × 2048 points in t_1 and t_2 directions, using uniformly ^{15}N-labeled samples. 3D-NOESY−HSQC and 3D-TOCSY−HSQC (*13*) were collected with 128 × 64 × 2048 in the t_1, t_2, and t_3 directions at 298 K. Zero filling was applied prior to Fourier transformation, and data were processed with a shifted squared sine-bell window functions in both dimensions.

RESULTS

HSQC Spectra of metLYZ and 2SS Variants. ^{15}N−^{1}H−HSQC spectra for metLYZ, 2SS[6−127,30−115], and 2SS-[64−80,76−94] were measured at pH 3.8 and 25 °C. The HSQC spectrum of metLYZ was substantially the same as that reported already for ^{15}N-labeled wild-type hen lysozyme expressed in *Aspergillus niger* (*14*). 126 cross-peaks were identified in Figure 2(a-1, a-2) for the backbone amide ^{1}H−^{15}N correlations, except for residues T40, P70, and P79. Resonance assignments were carried out using 3DTOCSY−HSQC and 3D NOESY−HSQC spectra that will be described later. Assigned cross-peaks are labeled in Figure 2(a-1, a-2, b-1, b-2). Six cross-peaks were identified for tryptophan indole ^{1}H−^{15}N correlations, although they are out of range of Figure 2(a-1). In addition, 17 pairs of cross-peaks were obtained for $^{15}NH_2$ groups of the 14 Asn and 3 Gln residues around 112 ppm of ^{15}N in the HSQC spectrum. Assignments of these side-chain ^{1}H−^{15}N correlations were substantially the same as those in wtLYZ, except for N39 and Q41. In addition, 11 cross-peaks were identified for arginine ^{1}H−$^{15}N_\epsilon$ side-chain correlations, although they are also out of range of Figure 2(a-1). R5, R61, R73, R112, R114, and R125 were assigned to each cross-peaks, while the cross-peaks corresponding to R14, R21, R45, R68, and R128 were detected separately but could not assigned to each residue even using the 3D NOESY−HSQC spectrum because resonance frequencies of HC_δ or HN_ϵ are too close to each other. NMR spectra of metLYZ were also measured at 35 °C for the comparison with chemical shift values of wtLYZ reported previously (*14*, *15*).

^{15}N−^{1}H−HSQC spectra of 2SS varinats of lysozyme are shown in Figure 2(b-1, b-2, c). As shown in Figure 2(c), the HSQC-spectrum of 2SS[64−80, 76−94] is analogous to that observed for reduced S-methylated lysozyme in 8 M urea (*16*) and characteristic of unfolded proteins. In this paper, the spectrum of 2SS[64−80, 76−94] will not be mentioned further. On the other hand, the HSQC-spectrum of 2SS[6−127, 30−115] is quite similar to that of metLYZ as a whole. More than 10 cross-peaks disappeared in the HSQC spectrum of 2SS[6−127, 30−115] in comparison with that of metLYZ. Instead, many more cross-peaks were crowded in the small region (Figure 2(b-2)) around 8.2 and 120 ppm. Boxed residue numbers in Figure 2(a-1, a-2) mark the cross-peaks that remarkably moved or disappeared in Figure 2(b-1, b-2). Most of them are located in the β-domain (T51, Y53, I55, L56, I58, N59, S60, R61, W62, W63, C64, N65, R68, T69, G71, S72, R73, L75, C76, N77, C80, S81, A82, and S85), although it is a matter of course that C64, C76, and C80 replaced by Ala are involved in them. Several pairs of cross-peaks were found for $^{15}NH_2$ groups of Asn and Gln residues. Assignments to N19, N37, N39, N44, and N113 were confirmed in the 3D NOESY−HSQC spectrum by connectivities between HN_δ and Hβ of their own residues. Other cross-peaks were assigned to some residues with reference to the spectrum of metLYZ (N46, Q57, N93, and N106). For Asn or Gln residues contained in the β-domain (Q41 N59, N65, N74, and N77), side-chain ^{1}H−^{15}N correlations disappeared or moved into the crowded region around 112 ppm of ^{15}N. Further, several cross-peaks were detected for arginine ^{1}H−$^{15}N_\epsilon$ side-chain correlations. The cross-peaks of R5, R112, R114, and R125 were identified with reference

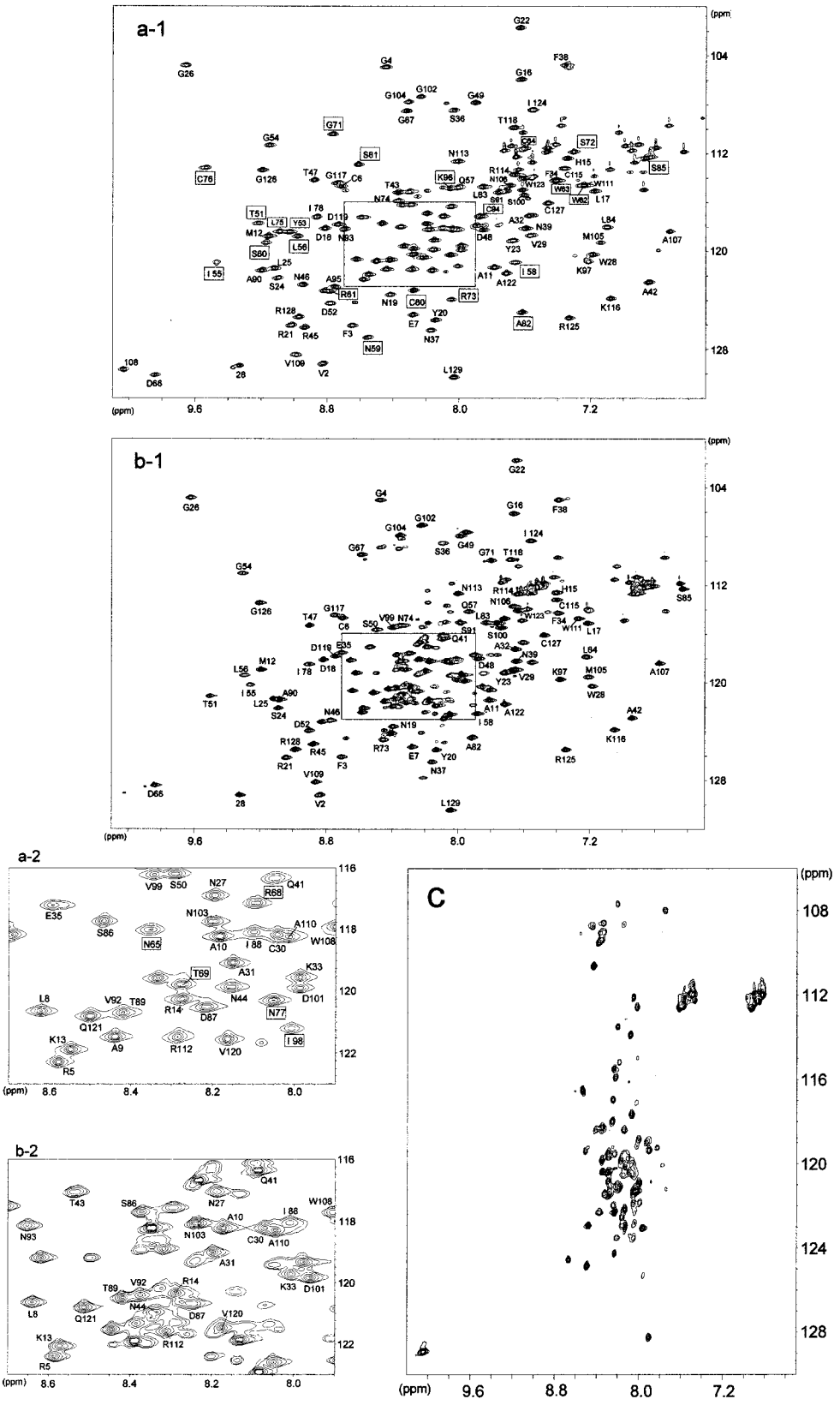

FIGURE 2: $^{15}N-^{1}H-$HSQC spectra of metLYZ and 2SS variants of lysozyme. All spectra were measured at 25 °C and pH 3.8. (a-1) The spectrum of metLYZ. Identified cross-peaks are labeled by the residue name for the backbone $^{1}H-^{15}N$ correlations. Seventeen pairs of cross-peaks were identified for $^{15}NH_2$ groups of Asn and Gln residues. They appear around 112 ppm of ^{15}N and 7.2 ppm of ^{1}H, but their peak labels are omitted. (a-2) The central region enclosed by a box is enlarged. (b-1) The spectrum of 2SS[6−127, 30−115]. Peak assignments were carried out using 3D TOCSY−HSQC and 3D NOESY−HSQC spectra. Cross-peaks of most residues contained in the α-domain appear at almost the same positions as those in metLYZ. On the contrary, those of residues contained in the β-domain remarkably move or disappear. Boxed residue numbers in panels a-1 and a-2 represent them. Ten pairs of cross-peaks are found for $^{15}NH_2$ groups of Asn and Gln residues, but their peak labels are omitted. (c) The spectrum of 2SS[64−80, 76−94]. This is characteristic of unfolded proteins.

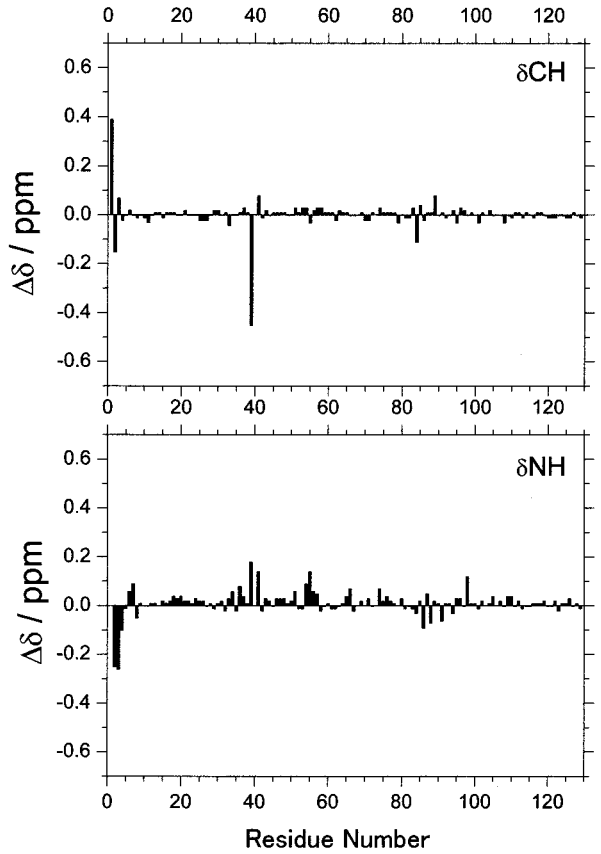

FIGURE 3: Differences in chemical shift values ($\Delta\delta$) between metLYZ and wtLYZ at 35 °C for Hα and HN. $\Delta\delta = \delta_{metLYZ} - \delta_{wtLYZ}$ is plotted vs amino acid residue number. The chemical shift values are strongly perturbed at the N-terminus and around N39. Further, they are perturbed around residues 55 and 86. These residues are located near the extra N-terminal Met.

to the spectrum of metLYZ, but those of R61 and R73 disappeared. These results appear to indicate that a part of the β-domain becomes flexible by the removal of two disulfide bridges in the β-domain. This tendency is more visible in the DQF−COSY spectrum as described below.

DQF−COSY Spectra of metLYZ and 2SS[6−127, 30−115]. DQF−COSY spectra are available for resonance assignments of Hα, Hβ, and other side-chain protons after backbone HN−Hα correlations were identified. In the DQF−COSY spectrum of metLYZ, almost all cross-peaks were detected between HN and Hα except for L25, T40, P70, P79, C80, and G104. In addition, cross-peaks were identified between Hα and Hα, between Hα and Hβ, or between side-chain and side-chain protons, including NH_2 protons. Peak assignments in metLYZ are quite consistent with those reported previously for wtLYZ (*15*, *17*), except for remarkable changes in chemical shift value at some specific sites described below. All the chemical shift values identified at 25 °C are listed in the Supporting Information.

Although both NMR spectra of metLYZ and wtLYZ are substantially the same, the resonance frequencies of individual Hα and HN are slightly perturbed by the addition of the extra N-terminal Met. Chemical shift differences of metLYZ from wtLYZ at 35 °C are plotted in Figure 3. It is interesting that the perturbation in resonance frequencies is remarkable around the N-terminus, N39, I55, and S86 in both Hα and HN. The cross-peak of T40 could be found in neither $^{15}N-^{1}H-$HSQC nor $^{1}H-^{1}H$ DQF−COSY spectra. This indicates that the presence of an extra N-terminal Met perturbs the so-called hinge region. As described below, NMR structural analysis clearly demonstrates that the three-dimensional structure of metLYZ is quite the same as that of wtLYZ despite the chemical shift perturbations found in the hinge region.

On the other hand, the cross-peaks significantly weakened in the DQF−COSY spectrum of 2SS[6−127, 30−115] compared with that of metLYZ. Especially in the fingerprint region, some HN−Hα correlations disappeared in the spectrum of 2SS[6−127, 30−115], although they intensely appeared in the spectrum of metLYZ. They are residues mainly involved in the β-sheet and the loop region: the cross-peaks of S36, N37, F38, S50, T51, D52, Q53, S60, R61, W62, W63, A64, N65, D66, R68, T69, G71, R73, N74, A76, N77, S81, A82, V92, A94, and V99. Also, cross-peaks of L56, I58, I78, and L83 are greatly weakened, although they were detectable. Most residues mentioned above are involved in the β-domain. Besides them, some of cross-peaks with medium intensities in metLYZ disappeared in the fingerprint region of 2SS[6−127, 30−115]. They are residues mainly involved in the α-domain: F3, C6, L8, A9, M12, K13, N19, G22, S24, N27, A31, L75, A90, A95, K96, G102, V120, and G126.

Although several HN−Hα cross-peaks of the residues contained in the N-terminal half (residues 1−39) of α-domain were undetectable as mentioned above, most of Hα-Hα, Hα-Hβ, and Hβ-Hβ cross-peaks within their own residues were clearly obtained in the DQF−COSY spectrum. Therefore, resonance frequencies of almost all protons within residues 1−58 could be determined using HN−Hα correlations obtained from 3D TOCSY−HSQC spectra, despite the loss of HN−Hα cross-peaks in the DQF−COSY spectrum. Also in the region of residues 82−129, in general, HN−Hα, HN−Hβ, and Hα-Hβ cross-peaks were clearly detected in either the DQF−COSY or the 3D TOCSY−HSQC spectrum except for A94, A95, K96, and I98. Therefore, resonance frequencies of almost all protons contained in this region could be successfully determined.

For most of residues 50−81 of 2SS[6−127, 30−115], however, not only HN−Hα but also Hα-Hβ and Hβ-Hβ cross-peaks disappeared from the DQF−COSY spectrum, although they are definitely detectable in metLYZ. As a result, resonance assignments are quite incomplete for the residues contained in this region. Probably this reflects the fact that the region ranging from S50 to S81 is strongly perturbed by the removal of two disulfide bridges in the β-domain. For residues S50 to I58, G67, G71, R73, N74, and I78 involved in this region, HN−Hα correlations were determined from the 3D TOCSY−HSQC spectrum.

NOESY Spectra of metLYZ and 2SS[6−127, 30−115]. Sequential NOE connectivities (NN(i,i+1) and αN(i,i+1)) are important for the accomplishment of resonance assignments. The 3D NOESY−HSQC spectrum was measured for metLYZ and sliced at the respective ^{15}N frequency (F2) to obtain F1−F3 spectra. In such $^{1}H-^{1}H$ NOESY spectra, on the average, about seven NOE cross-peaks were obtained at a resonance frequency of a specific HN. NN(i,i+1) and αN(i,i+1) connectivities detected for metLYZ are summarized in Figure 4(a), where NOE cross-peaks are classified into weak, medium, and strong ones. A single open square represents a weak NOE, and single and double filled squares

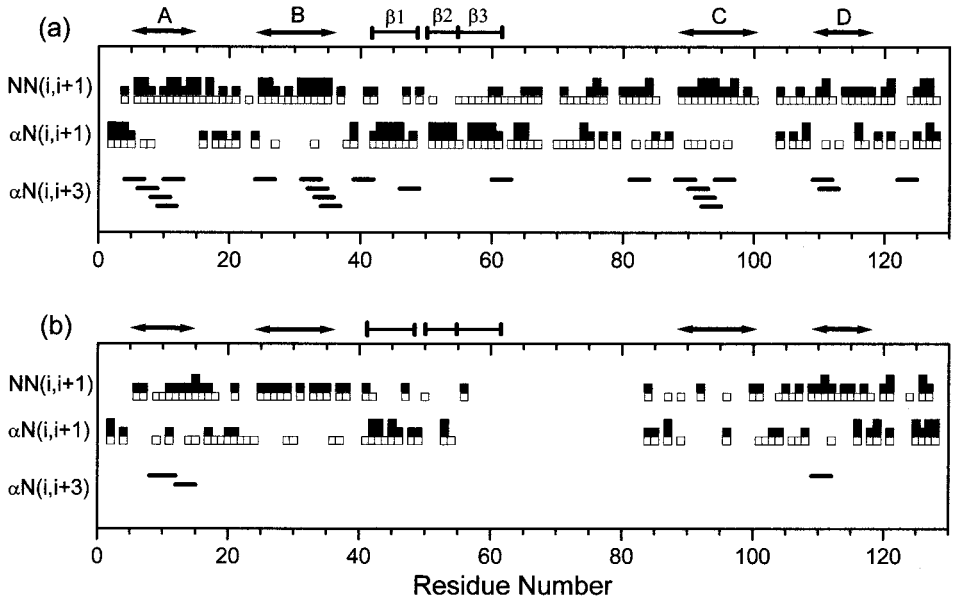

FIGURE 4: Summary of sequential and short-range NOE connectivities. NOE cross-peaks are classified into three groups according to their relative intensities; Open squares represent weak NOEs, and single and double filled squares on them denote medium and strong NOEs, respectively. The secondary structures observed in the wtLYZ are indicated along residue numbers. A, B, C, and D denote α-helices. Continuous αN(*i*,*i*+3) connectivities are indicative of α-helical structure. (a) metLYZ. (b) 2SS[6−127, 30−115]. Almost all sequential NOEs disappear in the region of residues 58−83 in 2SS[6−127, 30−115], although a lot of NOEs are observed in metLYZ.

on it denote medium and strong NOEs, respectively. First, this figure clearly shows that almost all backbone protons in metLYZ are connected by either NN(*i*,*i*+1) or αN(*i*,*i*+1) NOEs, indicating that resonance assignments are entirely consistent with those obtained from DQF−COSY spectra within own residues. Second, strong NN(*i*,*i*+1) connectivities are found in A-, B-, and C-helices and strong αN(*i*,*i*+1) connectivities in the N-terminal and the *β*-sheet regions. In Figure 4, short lines represent observed αN(*i*,*i*+3) connectivities. They concentrate in A-, B-, and C-helices. This is the strong indication that the polypeptide backbone adopts a regular helical structure.

The 3D NOESY−HSQC spectrum of 2SS[6−127, 30−115] was recorded in a similar manner. In F1−F3 spectra sliced at a fixed F2 frequency, on the average, five or six NOE cross-peaks were detected at the resonance frequencies of a specific HN. As shown in Figure 4(b), sequential NOE connectivities, NN(*i*,*i*+1) and αN(*i*,*i*+1), were obtained between neighboring residues in the regions of residues 1−57 and 84−129. They certified resonance assignments for these residues.

By contrast, any proper sequential NOE connectivity could not be obtained for residues 59−81, although some NOE cross-peaks were detected at unassigned HN resonance frequencies ranging from 7.8 to 8.6 ppm in F1−F3 spectra. Peak assignments for $^{1}H-^{15}N$ correlations of D66, G67, G71, R73, N74, and I78 in HSQC spectra are incomplete for lack of sequential NOE connectivities. For these residues, one of cross-peaks with almost the same ^{15}N and the nearest HN resonance frequencies was picked out from the HSQC spectrum of 2SS[6−127, 30−115] with reference to that of metLYZ and tentatively assigned to an unidentified residue. Then some evidence rationalized it. Since HN resonance frequencies of G67 and G71 shifted markedly, their resonance assignments were rationalized by the facts that their TOCSY cross-peaks are characteristic of Gly, and their Hα resonance frequencies in 2SS[6−127, 30−115] are close to those in metLYZ. The resonance assignment of R73 was confirmed by the presence of intraresidue HN−Hα and HN−H*β* correlations obtained in 3D TOCSY− HSQC spectra. Those of N74 and I78 were only rationalized by the slight shifts of ^{15}N, HN, and Hα resonance frequencies from corresponding ones in metLYZ. Since the cross-peak of D66 in the 2D HSQC spectrum was far away from others, it was assumed correct.

In sequential NOE connectivities of 2SS[6−127, 30−115] shown in Figure 4(b), NN(*i*,*i*+1) connectivities concentrate in A- and B-helices, although almost all connectivities are medium or weak. Most of NN(*i*,*i*+1) are lost in the C-helix region, while they are rather preserved well in the D-helix region. Further, αN(*i*,*i*+3) connectivities are hardly found even in the A- and B-helix regions. Compared with the α-helical structure existing in metLYZ, these data indicate that A-, B-, and D-helices are formed but considerably perturbed and that the C-helix is significantly deformed in 2SS[6−127, 30−115]. On the other hand, αN(*i*,*i*+1) connectivities are found in the region of residues 41−46 (*β*1-strand), but they are lost in the *β*2- and *β*3-strands. Probably this indicates that the *β*-sheet structure is remarkably distorted by the removal of SS3 and SS4 bridges. Although a large number of sequential NOE connectivities were obtained in the loop region (residues 62−80) of metLYZ, they were entirely lost in 2SS[6−127, 30−115].

As a result of resonance assignments of 2SS[6−127, 30−115], chemical shift values for HN, Hα, and side-chain protons are summarized in the Supporting Information. Resonance frequencies of HN and Hα in 2SS[6−127, 30−115] were compared with those in metLYZ at 25 °C. Figure 5 shows chemical shift differences of 2SS[6−127, 30−115] from metLYZ. It is an interesting fact that the chemical shift perturbation is very small in the region of residues 1−39 and 101−129. By contrast, it is considerably large in the region of residues 41−58 and 82−100. Since resonance frequencies of HN and Hα could not be deter-

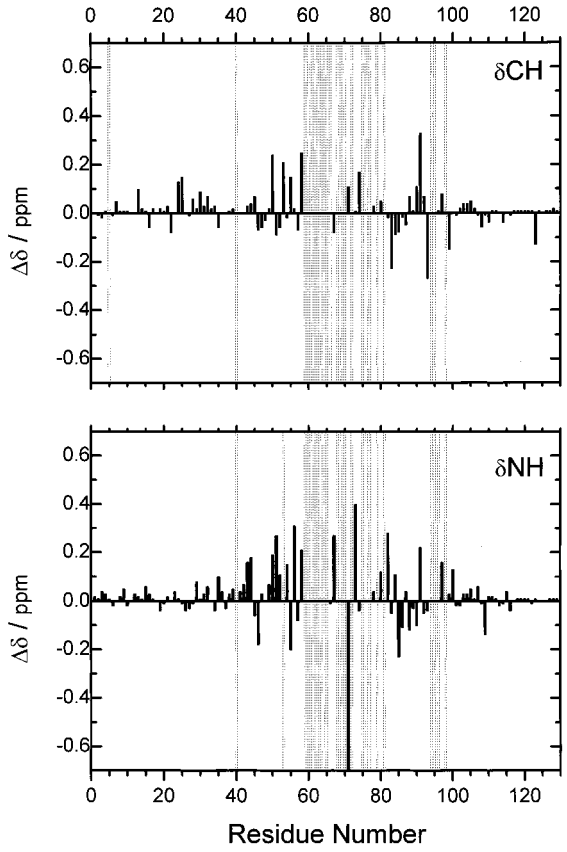

FIGURE 5: Differences in chemical shift values between 2SS[6−127, 30−115] and metLYZ. $\Delta\delta = \delta_{2SS} - \delta_{metLYZ}$. Chemical shift perturbation is very small in the region of residues 1−39 and 101−129. Since resonance frequencies of HN and Hα could not be determined in the region of residues 59−81, long broken lines are drawn to imply considerable perturbation in resonance frequencies.

mined in the region of residues 59−81 of 2SS[6−127, 30−115] except for some residues, long broken lines are drawn in Figure 5 to imply considerable perturbation in resonance frequencies at such sites. These results indicate again that the folded structure of α-domain is well preserved in 2SS[6−127, 30−115] except around C-helix, while the β3-strand (residues 57−61) and the loop region are unstructured.

Long-Range NOE Contacts in metLYZ. The 3D NOESY−HSQC spectrum of metLYZ provided about 900 cross-peaks between HN and protons in its vicinity, and that of 2SS[6−127, 30−115] provided about 600 cross-peaks. However, intraresidue or sequential ($|i - j| = 1$) or short-range ($|i - j| \leq 4$) NOE contacts were major components among them, while long-range ($|i - j| \geq 5$) NOE contacts were rather minor because either proton must be HN on the backbone. On the other hand, in 2D ^{1}H−^{1}H NOESY spectrum, since too many cross-peaks were crowded into the center of the spectrum, they were not easy to be distinguished from each other. However in the spectral region below 1.0 ppm, between 5.5 and 6.8 ppm, or above 9.0 ppm along the F2 axis, cross-peaks were well separated from each other so that they could be identified as NOE contacts between a specific pair of protons. In these spectral regions, a large number of long-range NOEs were obtained between side-chain and side-chain protons. As a result, 193 long-range NOEs were obtained between the *i*th and *j*th residues of $|i - j| \geq 5$ in the case of metLYZ, while 76 long-range NOEs in the case of 2SS[6−127, 30−115]. Figure 6 shows the 2D NOESY spectrum of 2SS[6−127, 30−115] in the −1.0 to 2.4 ppm range. Many long-range NOEs are found in this Figure: for example, V92Hγ−L17Hδ, K96Hϵ−L17Hδ, A31Hβ−M105Hϵ, A11Hβ−I88Hδ, T43Hγ−T51Hγ, A32Hβ−L56Hδ, etc. Long-range NOEs provide definitive evidence for the presence of three-dimensional structure. Figure 7 represents the map of such long-range NOE contacts. Above the diagonal line, NOE contacts in 2SS[6−127, 30−115] are plotted, and those in metLYZ are plotted below the diagonal.

The contact map is divided into three regions: residues 1−39, 40−87, and 88−129. The region of 1−39 is the N-terminal half of the α-domain, 88−129 the C-terminal half of the α-domain, and 40−87 the β-domain. First, we describe NOE contacts obtained in metLYZ. In Figure 7, J_{AB} represents NOE contacts found in the joint region connecting A-helix (residues 4−15) and B-helix (24−36): M12-L17, L17-L25, D18-L25, N19-L25, L17-W28, Y20−W28, etc. A−B represents NOE contacts obtained at the interface between A- and B-helices: L8-V29, L8-A32, A9-V29, etc. Further, long-range NOEs are obtained at the sites where both ends of the N-terminal half of α-domain are close to each other: V2−N39, F3−N39, F3−F38, R5−F38, L8-F38, etc. Residues 38−40 correspond to the so-called hinge (hinge1) connecting the α-domain to the β-domain, while the joint J_{CD} connecting C-helix (89−99) and D-helix (108−115) is involved in the C-terminal half of the α-domain, containing a number of NOE contacts: I98-A107, I98−W108, V99-M105, V99−W108, etc.

The N- and C-terminal halves of α-domain are in contact with each other. Long-range NOEs are found at the interface between B- and D-helices, between the C-terminal 3_{10}-helix (ct3_{10}, 120−124) and the B-helix, or between the C-helix and the joint connecting A- and B-helices (J_{AB}), which are referred to as B−D, ct3_{10}−B, and C−J_{AB} in Figure 7, respectively. Long-range NOEs, N27−W111, A31-W111, E35-A110, W28-M105, Y23−W111, Y23-M105, W28−W108, etc., are involved in B−D region. In particular, NOE contacts involved in C−J_{AB} region are important. Residues on C-helix facing toward the center of the α-domain are in close proximity to residues located in J_{AB} region, with both side-chains being in contact with each other. The highlights of them are V92−(M12, H15, L17), K96−(L17, Y20), V99−(Y20, W28), etc. The contacts between I88 and the A-helix (I88−A) may be added to them (see Figure 1). These NOE contacts found at the interface between N- and C-terminal halves of α-domain rather predominate in the α-domain.

Also, within the β-domain of metLYZ, there exist a lot of long-range NOE contacts. In Figure 7, β1−2 represents NOE contacts between β1 (residues 41−46) and β2 strands (50−54), and β2−3 between β2- and β3-strands (57−61). In addition, many long-range NOEs are contained in the so-called loop region (residues 62−79) referred to as loop in Figure 7. Further NOE contacts of (Q41, A42, T43)−L84 or Y53−(A80, L83, L84) are found in the region referred to as $\beta 3_{10}$−β, where $\beta 3_{10}$ means the 3_{10}-helix (residues 80−84) existing in the β-domain.

On the other hand, interdomain contacts are limited between the β-domain and the N-terminal half of α-domain. In this region, only the turn of I55 and L56 (referred to as

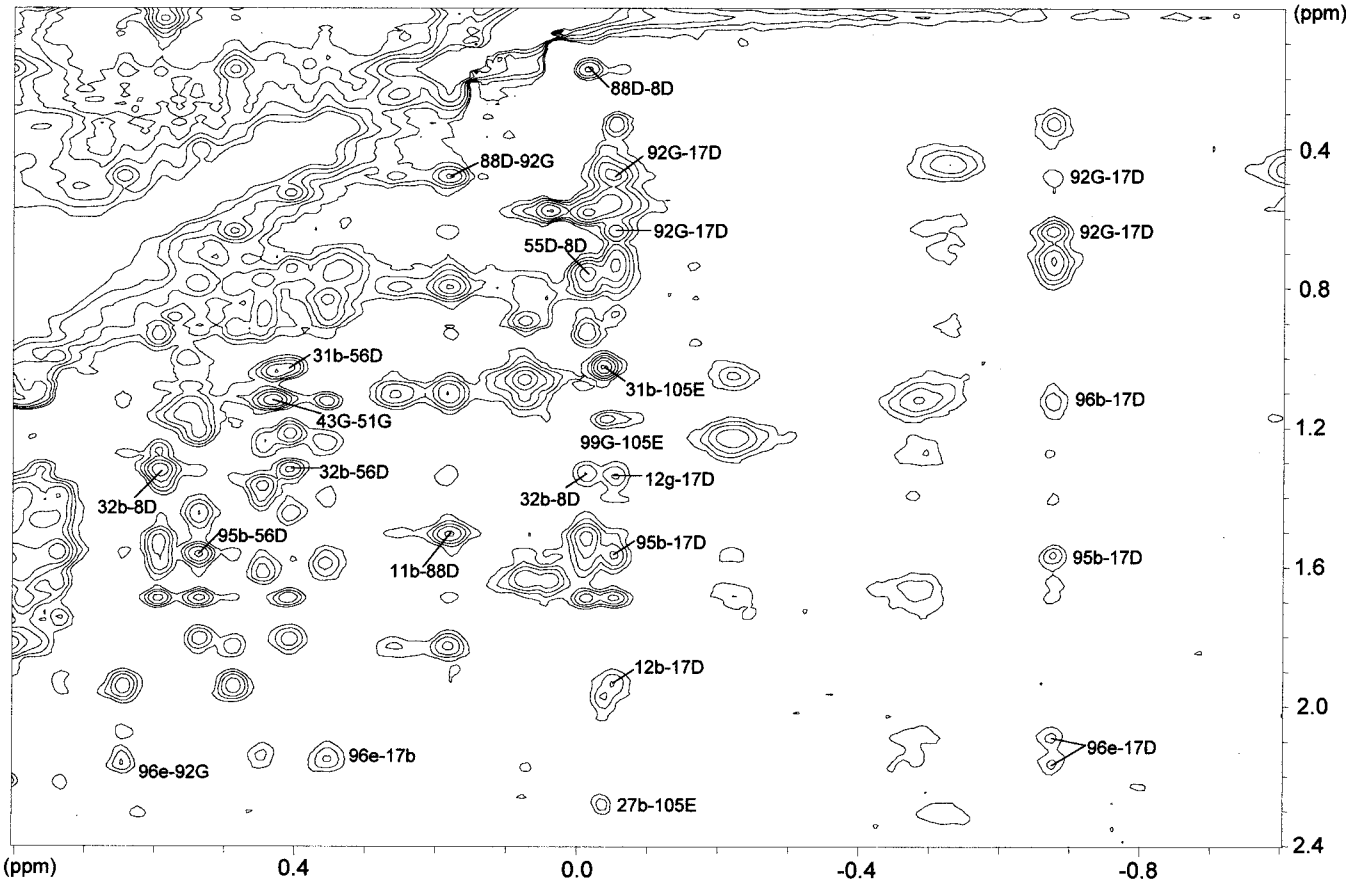

FIGURE 6: 2D $^1H-^1H$ NOESY spectrum of 2SS[6−127, 30−115] in the −1.0 to 2.4 ppm range. Many long-range NOE contacts appear between side-chain and side-chain protons. Only long-range NOEs are labeled in the spectrum, where b, g, d, and e mean $C_\beta H$, $C_\gamma H$, $C_\delta H$, and $C_\epsilon H$, respectively. Capital letters denote CH_3 protons.

tn55) participates in the long-range NOEs; in Figure 7, tn55-A and tn55-B represent the contacts of I55 or L56 with A- and B-helices, respectively. Indeed, side-chain protons of I55 or L56 have long-range NOEs with those of hydrophobic residues of F3, L8, M12, W28, A32, and F38. Long-range NOEs exist also between the β-domain and the C-terminal half of the α-domain. They are mainly I58−C94, I58−I98, I58−W108, W63−I98, and W63−A107, constructing the active site with a cleft. The turn of I55 and L56 also participates in the contact with the C-helix (tn55-C): I55−(I88, T89, S91) and L56−(A95, W108). Moreover, I78-A90 and I78−C94 are found between the C-helix and $\beta 3_{10}$. It appears important for folding of lysozyme that the turn of I55 and L56 is embedded in the hydrophobic pocket surrounded by A-, B-, and C-helices because the basement of the β-sheet is anchored to the α-domain, and it facilitates to attain the cooperativity between two domains.

Long-Range NOE Contacts in 2SS[6−127, 30−115]. Next we describe long-range NOE contacts found in 2SS[6−127,-30−115]. For instance, some of them are shown in Figure 6. The contact map between residues is plotted above the diagonal line in Figure 7. Also in this 2SS variant, NOE contacts are firmly maintained within the N- and C-terminal halves of α-domain despite the decrease in the total number of NOE contacts. In fact, A−B, J_{AB}, hinge1, and J_{CD} are symmetrically distributed in the contact map (Figure 7). Also in the interface region (I88−A, C−J_{AB}, B−D, $ct3_{10}$−B), the contact map is almost symmetrical with respect to the diagonal line; that is, NOE contacts are well preserved at

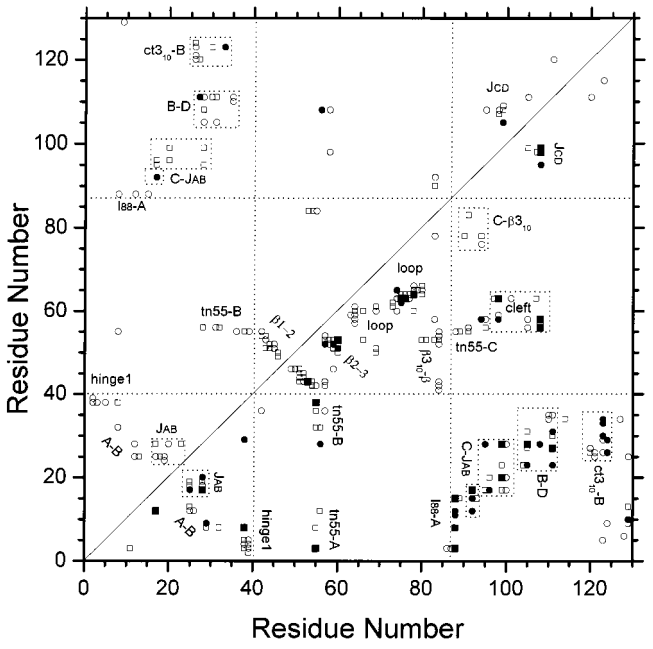

FIGURE 7: Long-range NOE contact map. A map below the diagonal represents the NOE contacts in metLYZ, a map above the diagonal in 2SS[6−127, 30−115]. In each map, a symbol indicates a long-range NOE connectivity observed between backbone and side-chain protons of the *i*th residue and those of the *j*th residue, where $|i - j| \geq 5$. Symbols vary according to the total number of observed NOE connectivities. 1, open circle; 2−3, open square; 4−5, filled circle; over 6, filled square. Long-range NOE contacts are classified into some groups and labeled (see text).

the interface between N- and C-terminal halves of the α-domain. In particular, it should be noted that NOE contacts are well preserved in C−J_{AB} and I88−A region of 2SS[6−127, 30−115]. Although the map of sequential NOEs indicated that the backbone of the C-helix is significantly deformed, the side-chains of I88, V92, K96, and V99 are properly embedded into the interior of the α-domain. As a result, side-chain packing between the N-terminal half of the α-domain (L8, M12, L17, Y20 and W28) and the C-helix is stably maintained to complete the hydrophobic core in the α-domain. The complementary interactions among these side-chains are illustrated in Figure 1. This seems to be the strong propensity inherent in C-helix because this polypeptide segment is quite free from the disulfide constraints. Long-range NOEs found in the regions of C−J_{AB}, B−D, ct3_{10}−B, and I88−A are definitive evidence that the α-domain of 2SS[6−127, 30−115] is quite similar to the native one.

On the other hand, only NOE contacts of β1−2 are preserved within the β-domain of 2SS[6−127, 30−115]. Any long-range NOE contact does not exist in the loop region. β2−3 contacts are also lost. Further, NOE contacts found around the cleft of metLYZ disappear in 2SS[6−127, 30−115]. These facts indicate that not only the loop but also the active site is unstructured. However, it is interesting that β1 and β2 strands adopt an antiparallel β-sheet. This is also a strong propensity inherent in its amino acid sequence. Moreover, the end of β2-strand is in close proximity to the end of the B-helix (tn55-B). That is to say, NOE contacts of I55−F38, I55−N39, L56-W28 and L56-A32 are firmly preserved in 2SS[6−127, 30−115]. It is noteworthy that N- and C-terminal ends of the β-domain (T40 and N87) and the basement of the β-sheet (I55 and L56) are fixed together to the so-called hinge region, through the specific binding between the C-helix and the N-terminal half of the α-domain.

DISCUSSION

Differences Between 2SS[6−127,30−15] and 2SS[64−80,76−94]. NMR structural analysis demonstrated that the structure of the α-domain in 2SS[6−127, 30−115] is quite similar to that of metLYZ, although the β-domain is largely unstructured. On the contrary, the HSQC spectrum of 2SS-[64−80, 76−94] was characteristic of unfolded proteins. 2SS[64−80, 76−94] not only lacks stable structure in the α-domain due to the loss of both SS1(C6−C127) and SS2-(C30−C115), but it also lacks ordered structure in the β-domain despite the existence of both SS3(C64−C80) and SS4(C76−C94). Therefore, the cooperativity between α- and β-domains appears indispensable to attain the structure of the β-domain, whereas the structure of α-domain may be maintained almost independently from the β-domain if both SS1 and SS2 exist.

The constraint alone due to SS3 and SS4 was insufficient to maintain stability of the β-domain. What is necessary for the stabilization of the β-domain? In metLYZ, some interactions between two domains appear to support the structure of the β-domain. For example, when the C-helix is built into the hydrophobic core through the specific interactions with the N-terminal half of the α-domain, residues 86 and 87 are anchored to the hinge region. Further, side-chains of I55 and L56 corresponding to the basement of the β-sheet are also fixed in the hydrophobic pocket surrounded by A-, B-, and C-helices. If these atom contacts were established in 2SS-[64−80, 76−94], the β-domain might be folded stably due to the presence of both SS3 and SS4. However, the α-domain is actually disrupted in 2SS[64−80, 76−94] so that the β-domain becomes unstructured because of the loss of cooperativity between two domains.

On the other hand, a large number of long-range atom contacts are involved in the α-domain of 2SS[6−127, 30−115]. In particular, when both SS1 and SS2 exist, the D-helix is linked to the B-helix, and the ct3_{10}-helix is close to the B-helix. It is likely that these constraints due to cross-linking facilitate to attain the long-range atom contacts around the regions B−D and ct3_{10}−B located at the interface between N- and C-terminal halves of the α-domain. Indeed many long-range NOEs were obtained in these regions of 2SS-[6−127, 30−115]. Further, close packing of side-chains was found between the C-helix and the N-terminal half of the α-domain in the hydrophobic core of 2SS[6−127, 30−115] as well as in metLYZ. In the presence of disulfide constraints due to SS1 and SS2, once the structure with the nativelike close packing is formed, intradomain atom contacts appear to be enough to maintain stability of the α-domain. However it should be noted that either SS1 or SS2 is indispensable for the preservation of the α-domain even in the presence of two disulfide bridges of SS3 and SS4.

Implications for Lysozyme Folding. It has been found that the collapse to the relatively compact globule occurs before the folding transition of authentic lysozyme (*18*, *19*). The persistent structure acquired in the α-domain during the folding process was similar to the so-called molten globule obtained at equilibrium under some unfolding condition (*20*−*22*). Many experiments have been carried out in lysozyme and α-lactalbumin to prove that folding intermediates contain persistent hydrogen bonded structure in the α-domain but lack stable structure in the β-domain. In particular, two-disulfide variants of α-lactalbumin, αLA-[6−120, 28−111] and αLA[61−77, 73−91], were studied by means of the disulfide exchange reaction and far- and near-UV CD spectra (*23*). They are close homologues to hen lysozyme 2SS[6−127, 30−115] and 2SS[64−80, 76−94], respectively. αLA[6−120, 28−111] and αLA[61−77, 73−91] were found both to be in the so-called molten globule state. Further, it was suggested that the persistent structure acquired in the α-domain has nativelike secondary structure and nativelike topology but no significant tertiary structure. The structural properties of kinetically trapped intermediates have been inferred from such a persistent structure existing in αLA[6−120, 28−111].

On the other hand, in hen lysozyme 2SS[6−127, 30−115] studied by us, the α-domain is tightly folded and maintained stably with leaving the β-domain unstructured. Judging from its NMR spectra, however, the conformation acquired in the α-domain is never the molten globule. Many side-chains are closely packed within the hydrophobic core, as shown from long-range NOEs involved in the regions B−D, ct3_{10}−B, C−J_{AB}, and I88−A. Thus, 2SS[6−127, 30−115] in an equilibrium state appears different from kinetically trapped intermediates, rather similar to the final native structure in which the β-domain is unable to be maintained stably because of the loss of SS3 and SS4. In the folding process of authentic lysozyme with four disulfide bridges, it is likely that kinetically trapped intermediates are in a dynamic molten

globule state and proceed to the final native structure with rearranging the polypeptide conformation.

When the constraints due to disulfide bridges are removed from the polypeptide chain, residual structures existing in it are stabilized through interactions inherent in the amino acid sequence. In 2SS[6−127, 30−115], the polypeptide segment ranging from 30 to 115 is quite free from the disulfideconstraints except around residues 30 and 115 so that the β-domain is unstructured as a whole. However, there remain some important residual structures in this region. For example, β1 and β2 strands were preserved stably as an antiparallel β-sheet, and the C-helix was properly built into the hydrophobic core, as shown in long-range NOE contacts. In addition, I55 and L56 were embedded between B- and C-helices. In the folding process of lysozyme, supposing that A- and B-helices are formed and folded into a nativelike conformation, the N-terminal half of α-domain may be greatly stabilized by the association of the C-helix through close packing of many side-chains within the hydrophobic core. Also, the association of the C-helix may serve to establish the basic framework of the β-domain by anchoring the N-terminal end of the C-helix (residues 88−91) to the vicinity of residues 39−41 and 54−56. Thus, the C-helix properly built into the hydrophobic core appears to have a crucial role to attain the cooperativity between two domains. Making these arrangements seems a good chance for the folding transition.

Four species of three-disulfide variant of lysozyme clearly showed the cooperative folding-unfolding transition induced by the addition of guanidine hydrochloride (*7*). From kinetic investigation of folding and unfolding reactions, it was found that not only the folding rate but also the unfolding rate is extremely slow in the variant lacking SS4 compared with the rates of the other three. These facts are consistent with the previous study that the most abundant folding intermediate formed during the oxidative refolding of lysozyme is the three-disulfide species lacking SS4 (*24*). Kinetic data for the folding-unfolding reaction were interpreted as follows; even in the absence of SS4, it is required for the folding transition to freeze the backbone conformation of residues 76−94 into nativelike topology. That is to say, the attainment of these constraints is the cause for slowing down only the folding rate of the variant lacking SS4. Supposing the backbone conformations of residues 41−56 (antiparallel β1-β2) and 88−99 (C-helix) are established immediately before the folding transition, the residue 76 may be drawn near the residue 94 on the C-helix built into the α-domain at the final stage of folding transition, even though the loop region remains flexible.

ACKNOWLEDGMENT

We thank Dr. K. Akasaka and Dr. K. Inoue for valuable advice about NMR measurements.

SUPPORTING INFORMATION AVAILABLE

Tables of ^{1}H chemical shift assignments of metLYZ and 2SS[6−127,30−115] at 25 °C and pH 3.8. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. van Mierlo, C. P. M., Darby, N. J., Neuhaus, D., and Creighton, T. E. (1991) *J. Mol. Biol. 222*, 373−390.
2. van Mierlo, C. P. M., Darby, N. J., and Creighton, T. E. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 6775−6779.
3. van Mierlo, C. P. M., Darby, N. J., Keeler, J., Neuhaus, D., and Creighton, T. E. (1993) *J. Mol. Biol. 229*, 1125−1146.
4. Sawano, H., Koumoto, Y., Ohta, K., Sasaki, Y., Segawa, S., and Tachibana, H. (1992) *FEBS Lett. 303*, 11−14.
5. Tachibana, H., Ohta, K., Sawano, H., Koumoto, Y., and Segawa, S. (1994) *Biochemistry 33*, 15008−15016.
6. Tachibana, H. (2000) *FEBS Lett. 480*, 175−178.
7. Yokota, A., Izutani, K., Takai, M., Kubo, Y., Noda, Y., Koumoto, Y., Tachibana, H., and Segawa, S. (2000) *J. Mol. Biol. 295*, 1275−1288.
8. Tachibana, H., Oka, T., and Akasaka, K. (2001) *J. Mol. Biol. 314*, 311−320.
9. Bax, A. (1989) *Ann. Rev. Biochem. 58*, 223−256.
10. Cavanagh, J., Fairbrother, W. J., Palmer, A. G., III, and Skelton, N. J. (1996) *Protein NMR Spectroscopy: Principles and Practice*, Academic Press, San Diego, CA.
11. Piotto, M., Saudek, V., and Sklenar, U. (1992) *J. Biomol. NMR 2*, 661−665.
12. Kay, L. E., Keifer, P., and Saarinen, T. (1992) *J. Am. Chem. Soc. 114*, 10663−10665.
13. Marion, D., Driscoll, P. C., Kay, L. E., Wingfield, P. T., Bax, A., Gronenborn, A. M., and Clore, G. M. (1989) *Biochemistry 28*, 6150−6156.
14. Buck, M., Boyd, J., Redfield, C., MacKenzie, D. A., Jeenes, D. J., Archer, D. B., and Dobson, C. M. (1995) *Biochemistry 34*, 4041−4055.
15. Redfield, C., and Dobson, C. M. (1988) *Biochemistry 27*, 122−136.
16. Schwalbe, H., Fiebig, K. M., Buck, M., Jones, J. A., Grimshaw, S. B., Spencer, A., Glaser, S. J., Smith, L. J., and Dobson, C. M. (1997) *Biochemistry 36*, 8977−8991.
17. Wang, Y., Bjorndahl, T. C., and Wishart D. S. (2000) *J. Biomol. NMR 17*, 83−84.
18. Matagne, A., Chung, E. W., Ball, L. J., Radford, S. E., Robinson, C. V., and Dobson, C. M. (1998) *J. Mol. Biol. 277*, 997−1005.
19. Dobson, C. M., Evans, P. A., and Radford, S. E. (1994) *Trends Biochem. Sci. 19*, 31−37.
20. Kuwajima, K. (1989) *Proteins: Struct. Funct. Genet. 6*, 8−103.
21. Ptitsyn, O. B. (1995) *Adv. Protein. Chem. 47*, 83−229.
22. Balbach, J., Forge, V., van Nuland, N. A. J., Winder, S. L., Hore, P. J., and Dobson, C. M. (1995) *Nature Struct. Biol. 2*, 865−870.
23. Wu, L. C., Peng, Z. Y., and Kim, P. S. (1995) *Nat. Struct. Biol. 2*, 281−286.
24. van den Berg, B., Chung, E. W., Robinson, C. V., Mateo, P. L., and Dobson, C. M. (1999) *The EMBO J. 18*, 4794−4803.

BI0113418