PROTEINS
STRUCTURE ■ FUNCTION ■ BIOINFORMATICS

Comprehensive secondary-structure analysis of disulfide variants of lysozyme by synchrotron-radiation vacuum-ultraviolet circular dichroism

Koichi Matsuo,[1] Hidenori Watanabe,[2] Shin-ichi Tate,[2] Hideki Tachibana,[3] and Kunihiko Gekko[2]*

[1] Hiroshima Synchrotron Radiation Center, Hiroshima University, Higashi-Hiroshima 739-0046, Japan
[2] Department of Mathematical and Life Sciences, Graduate School of Science, Hiroshima University, Higashi-Hiroshima 739-8526, Japan
[3] Department of Biotechnological Science, School of Biology-Oriented Science and Technology, Kinki University, Kinokawa, Wakayama 649-6493, Japan

ABSTRACT

To elucidate the effects of specific disulfide bridges (Cys6-Cys127, Cys30-Cys115, Cys64-Cys80, and Cys76-Cys94) on the secondary structure of hen lysozyme, the vacuum-ultraviolet circular dichroism (VUVCD) spectra of 13 species of disulfide-deficient variants in which Cys residues were replaced with Ala or Ser residues were measured down to 170 nm at pH 2.9 and 25°C using a synchrotron-radiation VUVCD spectrophotometer. Each variant exhibited a VUVCD spectrum characteristic of a considerable amount of residual secondary structures depending on the positions and numbers of deleted disulfide bridges. The contents of α-helices, β-strands, turns, and unordered structures were estimated with the SELCON3 program using the VUVCD spectra and PDB data of 31 reference proteins. The numbers of α-helix and β-strand segments were also estimated from the VUVCD data. In general, the secondary structures were more effectively stabilized through entropic forces as the number of disulfide bridges increased and as they were formed over larger distances in the primary structure. The structures of three-disulfide variants were similar to that of the wild type, but other variants exhibited diminished α-helices with a border between the ordered and disordered structures around the two-disulfide variants. The sequences of the secondary structures were predicted for all the variants by combining VUVCD data with a neural-network method. These results revealed the characteristic role of each disulfide bridge in the formation of secondary structures.

Proteins 2009; 00:000–000.

Key words: disulfide variants; lysozyme; secondary structure; synchrotron radiation; vacuum-ultraviolet circular dichroism.

INTRODUCTION

Disulfide bridges play an important role in the construction of the tertiary structure of proteins. Most proteins containing disulfide bridges unfold when they are completely reduced even in the absence of a denaturant, and the free energy of folding provides the driving force not only for correct folding but also for correct pairing of disulfide bridges.[1,2] However, it is unknown whether disulfide bridges force the protein into its folded form or whether other forces guide the protein into a conformation in which the disulfide bridges can form easily.[3] The stabilizing force of disulfide bridges is mainly attributable to the loss of conformational entropy of the unfolded polypeptide chain by cross-linking. This effect depends on the location of disulfide bridges: a greater separation between the cross-linked residues in the primary structure will result in a greater decrease in conformational entropy.[4] However, there is

Abbreviations: CD, circular dichroism; NN, neural network; VUV, vacuum-ultraviolet; VUVCD, vacuum-ultraviolet circular dichroism; SS1, disulfide bridge between Cys6 and Cys127; SS2, disulfide bridge between Cys30 and Cys115; SS3, disulfide bridge between Cys64 and Cys80; SS4, disulfide bridge between Cys76 and Cys94; SS12, two disulfide bridges of Cys6-Cys127 and Cys30-Cys115; SS13, two disulfide bridges of Cys6-Cys127 and Cys64-Cys80; SS23, two disulfide bridges of Cys30-Cys115 and Cys64-Cys80; SS34, two disulfide bridges of Cys64-Cys80 and Cys76-Cys94; SSΔ1, deletion of Cys6-Cys127; SSΔ2, deletion of Cys30-Cys115; SSΔ3, deletion of Cys64-Cys80; SSΔ4, deletion of Cys76-Cys94; 0SS-variant, variant with no disulfide bridge; 1SS variant, one-disulfide variant in which one of the four native disulfide bridges (Cys6-Cys127, Cys30-Cys115, Cys64-Cys80, and Cys76-Cys94) is formed; 2SS variant, two-disulfide variant with two of the four disulfide bridges; 3SS variant, three-disulfide variant with three of the four disulfide bridges.

Grant sponsor: Ministry of Education, Science, Sports, and Culture of Japan; Grant number: 20550153.
*Correspondence to: Kunihiko Gekko, Department of Mathematical and Life Sciences, Graduate School of Science, Hiroshima University, Higashi-Hiroshima 739-8526, Japan.
E-mail: gekko@sci.hiroshima-u.ac.jp.
Received 8 January 2009; Revised 2 March 2009; Accepted 10 March 2009
Published online in Wiley InterScience (www.interscience.wiley.com).
DOI: 10.1002/prot.22430

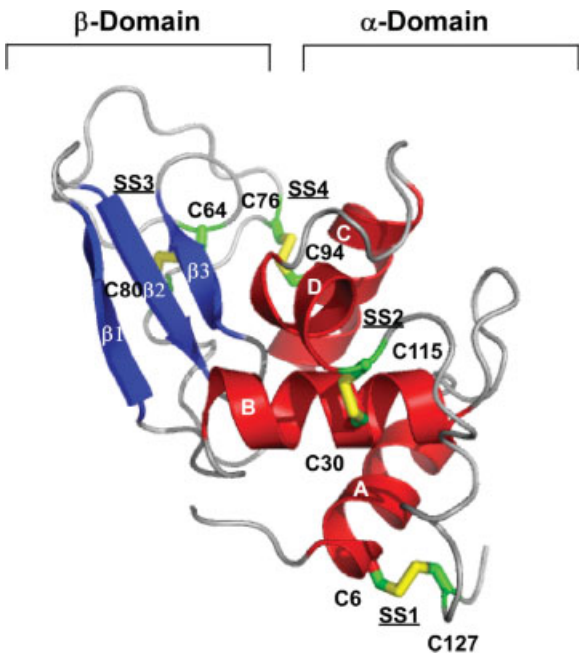

Figure 1

Ribbon diagram of the wild-type hen lysozyme produced from the X-ray structure (PDB code: 6LYZ). This figure was drawn by the program PyMOL.[42] The sequences for α-helices and β-strands were assigned as follows by the DSSP program: A-helix (residues 5-14), B-helix (residues 25-36), C-helix (residues 89-100), D-helix (residues 109-114), β1-strand (residues 43-45), β2-strand (residues 51-53), and β3-strand (residues 58-59). Yellow sections protruding from the backbone indicate the disulfide bridges of Cys6-Cys127 (SS1), Cys30-Cys115 (SS2), Cys64-Cys80 (SS3), and Cys76-Cys94 (SS4).

evidence that disulfide bridges affect the energetics of the folded conformation (native state) to varying extents, such as in the loss of stabilizing interactions and the possible formation of a strain associated with the introduction of cross-links.[5]

Hen lysozyme contains four disulfide bridges: Cys6-Cys127 (SS1), Cys30-Cys115 (SS2), Cys64-Cys80 (SS3), and Cys76-Cys94 (SS4). Figure 1 shows a ribbon diagram of the wild-type hen lysozyme as revealed by X-ray crystallography (PDB code: 6LYZ), in which the secondary structure elements were assigned with the DSSP program. Two disulfide bridges, SS1 and SS2, are located in the α-domain (residues 1–39 and 88–129), SS3 is in the β-domain (residues 40–87), and SS4 is at the interface between two domains. To examine the effects of individual disulfide bridges on the stability and folding kinetics of this protein, various disulfide-deficient variants were constructed by systematically replacing the Cys residues with Ala or Ser residues.[6–12] The structures of these variants have been studied mainly by circular dichroism (CD)[6,8,10] and NMR spectroscopes.[11,12] A general conclusion of these studies is that all species of the three-disulfide variant (3SS) have the native-like structure, all species of the one-disulfide variant (1SS) and a variant having no disulfide bridge (0SS) are largely unstructured, and the two-disulfide variants (2SS) are at the border between the ordered and disordered states. However, the roles of individual disulfide bridges remain controversial since NMR is not suitable for structures exhibiting large fluctuations although it does give detailed information on native and native-like structures.

CD spectroscopy has been widely used for analyzing the secondary structures of proteins from the ordered (native) to disordered (denatured) states, because this spectroscopy is very sensitive to local peptide structures (although, unlike NMR, it does not exhibit atomic resolution) and because CD spectra can be measured at a low protein concentration under various solvent conditions.[13] The analysis of secondary structures by CD spectroscopy has improved considerably due to (i) the development of software programs (DSSP and Xtlsstr) capable of extracting the secondary structures from atomic coordinates,[14,15] (ii) the advancements in software (CONTIN, SELCON3, and CDSSTR) for analyzing CD spectra,[16–18] and (iii) the extension of CD measurements to the vacuum ultraviolet (VUV) region.[19] The short-wavelength limit of CD spectroscopy can be successfully extended by using synchrotron radiation as a high-flux source of photons, which yields much more accurate data that cannot be obtained by a conventional CD spectrophotometer.[20–23] Estimations of the contents and segment numbers of the secondary structures of proteins have been considerably improved by extending CD measurements to 160 nm using a vacuum-ultraviolet circular dichroism (VUVCD) spectrophotometer.[24,25] VUVCD spectroscopy has also revealed characteristic differences in the secondary structures of various types of denatured proteins.[26] Further, we have recently succeeded in improving the sequence-based prediction of secondary structures by combining VUVCD data with a neural-network (NN) method,[27] which has provided a more detailed understanding of the secondary structures of proteins.

In the present study, VUVCD spectroscopy was applied to disulfide-deficient variants of hen lysozyme to elucidate the roles of individual disulfide bridges in secondary structure formation. We comprehensively measured the VUVCD spectra of 13 species of disulfide variants (4 species of each of 3SS, 2SS, and 1SS variants, and the 0SS variant) down to 170 nm using a synchrotron radiation VUVCD spectrophotometer. The contents, segment numbers, and sequences of the secondary structures were predicted for all the variants, and the roles of each disulfide bridge in secondary structure formation were discussed comparing with the results of NMR analysis.

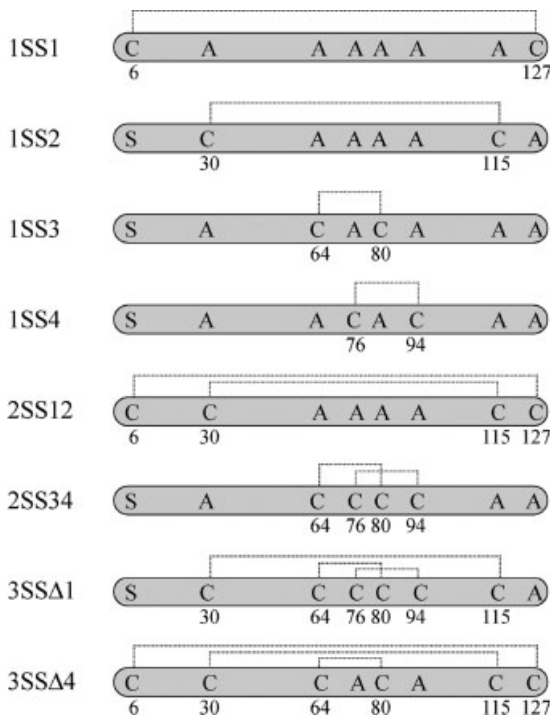

Figure 2

Nomenclature of disulfide variants of hen lysozyme based on their primary structures. The letters "C", "A", and "S" denote cysteinyl, alanyl, and seryl residues, respectively, and the numbers beneath them denote their positions.

MATERIALS AND METHODS

Construction of disulfide variants of hen lysozyme

The 13 species of disulfide-deficient variants of hen lysozyme were prepared by replacing the respective pairs of Cys residues among four disulfide bridges with Ala or Ser residues (C6S, C30A, C64A, C76A, C80A, C94A, C115A, and C127A). The nomenclature of the constructed variants is illustrated in Figure 2. The methods of gene construction, protein expression, and purification of recombinant variants have been detailed previously.[6–10] The N-terminal methionine residue is attached in each variant, because it is expressed with *Escherichia coli*. The wild-type hen lysozyme (crystallized six times) was obtained from Seikagaku Kogyo.

Sample preparation

The freeze-dried samples of the variants were dissolved in double-distilled water. The pH was adjusted to 2.9 ± 0.1 with trifluoroacetic acid or acetic acid. The wild-type lysozyme was dialyzed against double-distilled water (pH 6-7) and used without further purification. The concentration of the wild type was determined spectrophotometrically using a spectrophotometer (UV-250, Shimazu) with an extinction coefficient of 26.4 dL g^{-1} cm^{-1} at 280 nm.[11] This extinction coefficient was also used for the 3SS and 2SS variants, and corrected to 25.6 dL g^{-1} cm^{-1} for 1SS1 and 1SS4, 25.7 dL g^{-1} cm^{-1} for 1SS2 and 1SS3, and 25.2 dL g^{-1} cm^{-1} for 0SS based on the results of the amino-acid analysis of these variants.

VUVCD measurements

The VUVCD spectra of the variants were measured from 260 to 170 nm under a high vacuum (10^{-4} Pa) using the VUVCD spectrophotometer constructed at the Hiroshima Synchrotron Radiation Center (HSRC/HiSOR) and an assembled-type optical cell with CaF_2 windows. The details of the optical devices of the spectrophotometer and the optical cell are available elsewhere.[28,29] The optical path length of the cell was adjusted with a Teflon spacer to 10 μm. The temperature of the cell was controlled at 25°C using a Peltier temperature-control unit. All of the VUVCD spectra were recorded at protein concentrations of 0.1% to 0.4% with a 0.25-mm slit, a 16-s time constant, a 4-nm min^{-1} scan speed, and 4 to 9 accumulations. The spectra were calibrated by normalizing the ellipticities to those measured using a conventional CD spectrophotometer (J-720W, Jasco) in the overlapping wavelength region (260-200 nm) with a 50-μm optical path length, a 20 nm min^{-1} scan speed, and 9 accumulations. The ellipticity was reproducible within an error of 5%, with this error being mainly attributable to noise in the signal and to inaccuracy in the optical path length.

Secondary structure analysis

The secondary structure contents of the variants were analyzed using the modified SELCON3 program[16] and the VUVCD spectra down to 160 nm for the following 31 reference proteins with known X-ray structures[24,25] (their PDB codes are in parentheses): myoglobin (1WLA), hemoglobin (1G08), human serum albumin (1E78), cytochrome *c* (1HRC), peroxidase (1ATJ), α-lactalbumin (1F6S), lysozyme (1HEL), ribonuclease A (1FS3), insulin (4INS), lactate dehydrogenase (9LDT), glucose isomerase (1OAD), lipase (3LIP), conalbumin (1OVT), transferrin (1LFG), catalase (7CAT), subtilisin A (1SBC), α-amylase (1BAG), papain (9PAP), ovalbumin (1OVA), β-lactoglobulin (1B8E), pepsin (4PEP), trypsinogen (1TGN), α-chymotrypsinogen (2CGA), soybean trypsin inhibitor (1AVU), concanavalin A (2CTV), staphylococcal nuclease (1EY0), thioredoxin (2TRX), carbonic anhydrase (1G6V), elastase (3EST), avidin (1AVE), and xylanase (1ENX). The secondary structures of these proteins in crystal form were assigned into four classes (α-helices, β-strands, turns, and unordered structures) using the DSSP program[14] based on the hydrogen bonds

between adjacent amide groups. In this study, the 3_{10}-helix was classified as an unordered structure. The root-mean-square deviation (δ) and the Pearson correlation coefficient (r) between the X-ray and VUVCD estimates of the secondary structure contents of the reference proteins were 0.058 and 0.85, respectively, confirming the high accuracy of the VUVCD estimation.[24,25] The predictive accuracy was much better for α-helix content ($r = 0.93$) than for β-strand content ($r = 0.82$). The numbers of α-helix and β-strand segments were estimated by assuming the presence of four distorted residues per α-helix segment and two distorted residues per β-strand segment.[17]

Sequence-based prediction of secondary structures

The sequences of the secondary structures were predicted by combining the VUVCD data down to 160 nm with the NN algorithm (the VUVCD-NN method).[27] Thirty of the above 31 reference proteins (excluding insulin) were used as the target proteins for the NN analysis. The secondary structures of these proteins were assigned into three classes (α-helices, β-strands, and others) using the DSSP program[14] from the PDB codes. Turns, 3_{10}-helices, π-helices, bends, β-bridges, and blanks were grouped into "others" in the present study.

We adopted the NN algorithm developed by Jones,[30] which predicts the weights (propensity) of 20 amino acids for α-helices and β-strands using evolutional sequence information based on the position-specific scoring matrices generated by the PSI-BLAST algorithm. For the NN calculation, a training data set of 607 chains was prepared from the protein structures in PDB (September 2006 version)[31] by eliminating proteins with short chains, unidentified sequences, sequence similarities, and chain breaks.[27] The secondary structures of these training proteins and their amino-acid sequence information determined by the position-specific scoring matrices program[30] were used to calculate the weights and biases of the 20 amino acids for α-helices and β-strands.

The positions of α-helices and β-strands were assigned on the amino-acid sequence in descending order of the α-helix and β-strand weights of the 20 amino acids until the determined numbers of α-helix and β-strand residues converged to those estimated from the VUVCD analysis. Next, the numbers of α-helix and β-strand segments estimated from the VUVCD analysis were introduced into the NN calculation until the predicted numbers of segments converged to those obtained from the VUVCD estimation. If the predicted numbers of residues and segments for α-helices and β-strands did not converge to the VUVCD estimates, the sequence alignment that minimized the difference between the two estimates was taken as the final solution. The mean predictive accuracy for the sequences of the secondary structures was 74.9% for the 30 reference proteins. The computational protocol for this VUVCD-NN method has been detailed previously.[27]

RESULTS

The disulfide-deficient variants of hen lysozyme are easily unstructured in a medium at a low pH and high temperature, although the wild-type lysozyme is relatively stable.[7–10] Some variants are partly unfolded even at pH 3.9 and 4°C, as judged from their far-UV CD spectra.[8] Most variants are insoluble at a neutral pH and room temperature, although their tertiary structures can be stable. Therefore, in the present study, we measured the VUVCD spectra at pH 2.9 and 25°C, under which conditions the variants are relatively unstable and hence might more distinctly reflect the deletion effects of disulfide bridges. No buffer was used to reduce noise in the spectra. The VUVCD spectrum of the wild type was measured at neutral pH as a reference native state. All the VUVCD spectra obtained were superimposed on those in the far-UV region as measured by a conventional CD spectrophotometer. The VUVCD spectra were constant during the data-acquisition period (∼2 h) under the experimental conditions used, indicating that the synchrotron radiation (0.7 GeV) did not damage the proteins. Thus, the obtained VUVCD spectra are available for comprehensively comparing the secondary structures of disulfide variants.

VUVCD spectra of disulfide variants

Figure 3 shows the VUVCD spectra of the wild-type lysozyme and 13 disulfide variants. Evident characteristic differences in the spectra of these variants indicate differences in their secondary structures. As shown in Figure 3(A), the VUVCD spectra of 3SS variants do not differ greatly from that of the wild type, but they are clearly dependent on the position of the deleted disulfide bridge. The overall features of the VUVCD spectra above 190 nm are very similar to those of the far-UV CD spectra measured at pH 3.9 and 25°C,[6] but the CD bands around 190 nm are distinctly detectable in the VUVCD spectroscopy data. Deletion of the Cys6-Cys127 bridge (3SSΔ1) induces a decrease in intensity and a blue shift of the CD peaks in the VUV and far-UV regions. On the other hand, deletion of Cys30-Cys115 (3SSΔ2) induces an increase in CD intensity that is much more significant in the far-UV region than in the VUV region. The abnormality for 3SSΔ2 was also observed in the far-UV CD spectrum at pH 3.9: the peak intensity at around 220 nm is slightly larger than that for the wild type.[6] However, these results do not necessarily indicate increased secondary structures of this variant because the CD spectra at least in the far-UV region might be perturbed by exciton coupling between the aromatic residues, especially tryptophan resides.[32] Deletions of Cys64-Cys80

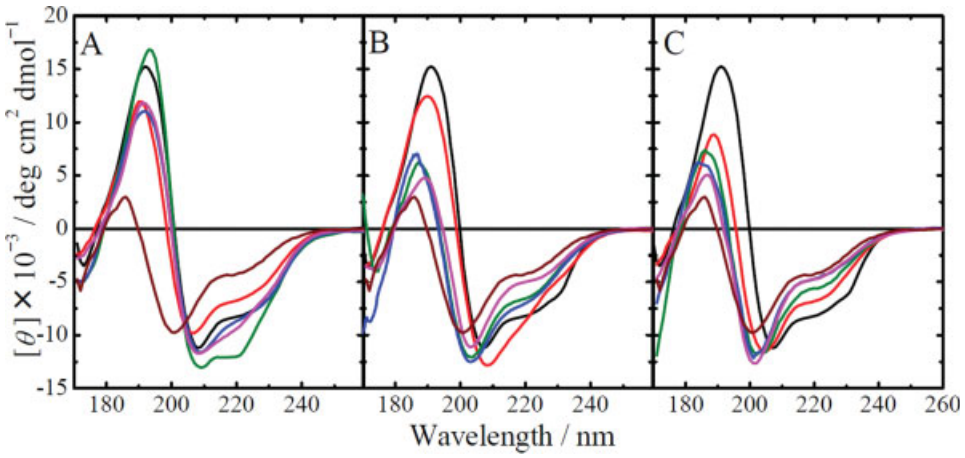

Figure 3

VUVCD spectra of 3SS, 2SS, 1SS, and 0SS variants of hen lysozyme at pH 2.9 and 25°C. Panel A: 3SSΔ1 (red), 3SSΔ2 (green), 3SSΔ3 (blue), and 3SSΔ4 (pink). Panel B: 2SS12 (red), 2SS13 (green), 2SS23 (blue), and 2SS34 (pink). Panel C: 1SS1 (red), 1SS2 (green), 1SS3 (blue), and 1SS4 (pink). The VUVCD spectra for the wild-type lysozyme at neutral pH (black) and 0SS variant (brown) are shown in each panel for comparison.

(3SSΔ3) and Cys76-Cys94 (3SSΔ4) exert similar effects on the spectra, with a decreased peak intensity at around 190 nm and a slightly increased intensity at 208 to 225 nm.

The VUVCD spectra of three 2SS variants (2SS13, 2SS23, and 2SS34) clearly differ from those of 3SS variants: the CD intensity decreases markedly over the entire wavelength range accompanying the blue shift of the CD peaks at around 190 and 208 nm [Fig. 3(B)]. However, the 2SS12 variant without Cys64-Cys80 and Cys76-Cys94 has a spectrum similar to those of the 3SSΔ3 and 3SSΔ4 variants, suggesting that the 2SS12 variant retains a considerable amount of secondary structures while the other three 2SS variants are largely unfolded. This distinguishing feature of 2SS12 among the 2SS variants is supported by the far-UV CD spectra differing between the 2SS12 and 2SS34 variants at pH 3.9 and 4°C.[10]

The VUVCD spectra of the four 1SS variants and the 0SS variant are shown in Figure 3(C). The CD intensity of 1SS variants decreases in the order 1SS1 > 1SS2 > 1SS3 ≈ 1SS4, accompanying the blue shift of the 190-nm band. The overall features above 190 nm are similar to the reported far-UV CD spectra at pH 4.0 and 4°C.[8] It is evident that the CD intensities at 190 nm and in the far-UV region are lowest in the 0SS variant among the disulfide variants examined, suggesting the greatest disruption of the secondary structures. The VUVCD spectrum of 0SS above 200 nm is nearly identical to the reported far-UV CD spectrum at pH 3.9 and 4°C[8] and to the spectrum of the completely reduced and carboxymethylated lysozyme.[33] Thus, the VUVCD spectra down to 170 nm provide much more detailed information about the secondary structures of the variants than do the far-UV CD spectra down to ~190 nm.

Contents and segment numbers of secondary structures

The secondary structure contents of the wild-type and variant lysozymes were estimated from their VUVCD spectra (see Fig. 3) using the SELCON3 program with the VUVCD spectra and X-ray structures for the 31 reference proteins. The estimated contents of α-helices, β-strands, turns, and unordered structures are listed in Table I, together with the secondary structure contents of the wild type assigned by the DSSP program from the X-ray structures. For ease of comparison, the secondary structure contents in Table I are depicted in Figure 4 as proportional histograms with each secondary structure content normalized to a total content of 100%.

To get the numbers of α-helix and β-strand segments from the CD data, we need information on the end effects of α-helices and β-strands. However, since the operation of these effects on the CD is not definitely evaluated, two approximation methods have previously been used to estimate the numbers of α-helix and β-strand segments from the CD spectra. Pancoska *et al.*[34] used a matrix descriptor of secondary structure segments for the NN-based analysis of protein CD spectra. Sreerama *et al.*[17] empirically determined that on average two and one distorted residues at each end of an α-helix and a β-strand, respectively, contributed to the CD of the segments, these numbers of distorted end residues satisfying the structural interpretation. The results of these two analyses are comparable, and hence we used the method of Sreerama *et al.* to estimate the numbers of α-helix and β-strand segments of each variant from the α-helix and β-strand contents. The results of this estimation are given in Table I and shown in the histograms of Figure 4.

Table I
Secondary Structure Contents, Segment Numbers, and Conformational Entropies of Disulfide Variants of Lysozyme

	Contents (%) of				Segment numbers of		
Variant	α-Helix	β-Strand	Turn	Unordered	α-Helix	β-Strand	ΔS_c (cal mol^{-1} K^{-1})
X-ray	31.8	6.3	27.9	34.0	4	3	
Wild	34.7 ± 1.5	8.0 ± 3.0	23.0 ± 2.2	34.3 ± 2.1	5	3	−67.04
3SSΔ1	23.3 ± 0.8	19.9 ± 1.4	21.7 ± 1.0	33.8 ± 1.8	3	5	−49.47
3SSΔ2	34.8 ± 0.9	16.1 ± 1.0	19.7 ± 0.6	31.7 ± 1.9	5	4	−50.92
3SSΔ3	28.0 ± 0.8	17.4 ± 1.7	23.3 ± 1.3	31.6 ± 1.4	5	4	−52.51
3SSΔ4	30.6 ± 1.4	12.8 ± 2.5	21.4 ± 3.4	36.4 ± 3.1	5	4	−52.25
2SS12	30.5 ± 1.2	14.3 ± 2.3	23.3 ± 1.8	34.6 ± 1.6	4	4	−37.71
2SS13	15.3 ± 0.8	23.2 ± 1.4	21.3 ± 0.6	41.7 ± 0.3	2	7	−35.92
2SS23	15.9 ± 0.9	22.4 ± 1.5	21.5 ± 0.8	42.5 ± 0.4	2	7	−34.67
2SS34	13.4 ± 0.9	24.8 ± 2.6	20.6 ± 0.9	40.6 ± 1.9	2	7	−30.50
1SS1	19.2 ± 1.1	20.3 ± 2.2	19.2 ± 1.0	39.7 ± 2.8	2	6	−21.19
1SS2	15.9 ± 0.1	22.8 ± 0.1	22.7 ± 0.1	42.4 ± 0.2	2	7	−20.14
1SS3	11.0 ± 1.3	26.0 ± 3.2	20.0 ± 1.2	43.3 ± 1.6	2	7	−15.16
1SS4	9.8 ± 1.4	24.6 ± 2.2	21.6 ± 1.4	43.6 ± 2.2	1	7	−15.51
0SS	6.8 ± 1.3	29.8 ± 2.0	20.6 ± 1.6	42.7 ± 1.8	1	7	0.00

Sequence-based prediction of secondary structures

CD or VUVCD spectroscopy itself in principle yields no information on the sequences of the secondary structures, and hence an algorithm exploiting the correlations between the X-ray structures and amino acid sequences of many proteins is necessary for the sequence-based prediction of secondary structures (α-helices, β-strands, and others). There are various types of NN algorithm and computational technique that could be applied,[30,35–37] and we used the NN algorithm of Jones[30] due to the technical convenience of combining it with VUVCD data. Since no propensity data of the 20 amino acids for the secondary structure elements of unfolded proteins are available, we predicted the sequences of the secondary structures of the disulfide variants by constraining the NN calculation with the propensity data for native proteins and with the contents and segment numbers of α-helices and β-strands obtained from the VUVCD spectra. Figure 5 compares the predicted sequences of α-helices, β-strands, and others with those determined from the X-ray structure (PDB code: 6LYZ), in which the sequences for α-helices (A, B, C, and D) and three β-strands (β1, β2, and β3) might have differed slightly from those previously used,[11,12] because the X-ray structure was assigned by the DSSP program in the present study. For the wild-type lysozyme, 99 of the 129 total amino acid residues agreed with the sequences determined from the X-ray structure, indicating a predictive accuracy of 76.7%. The numbers of α-helix and β-strand segments predicted at the sequence level did not necessarily agree with those predicted from the VUVCD data (Table I) because the sequence alignment that minimized the difference between the two estimates was taken as the final solution when the predicted numbers of segments did not converge to the VUVCD estimates (see Materials and Methods section).

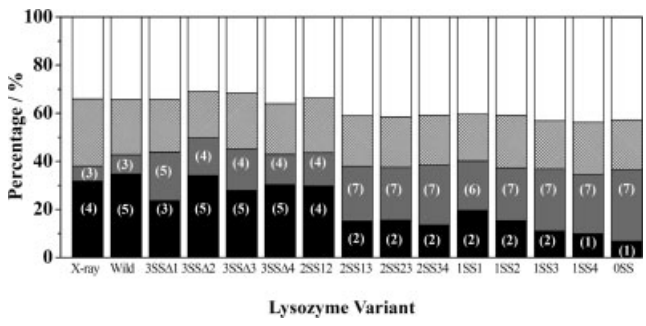

Figure 4
Histogram of the secondary structure contents of disulfide variants of hen lysozyme. α-Helices (black), β-strands (gray), turns (hatching), unordered structures (white). The numbers of α-helix and β-strand segments are shown in the parentheses.

DISCUSSION

The VUVCD spectra of disulfide variants of lysozyme clearly demonstrate that the disulfide bridges play important roles in the structural stability and folding of this protein. The characteristic contributions of each disulfide bridge to the secondary structures were quantitatively evaluated based on the analyses of secondary structure (Table I, Figs. 4 and 5). These results are useful for comparing partly unfolded structures of these variants since NMR is not suitable for structures exhibiting large fluctuations, such as the 1SS and 0SS variants.

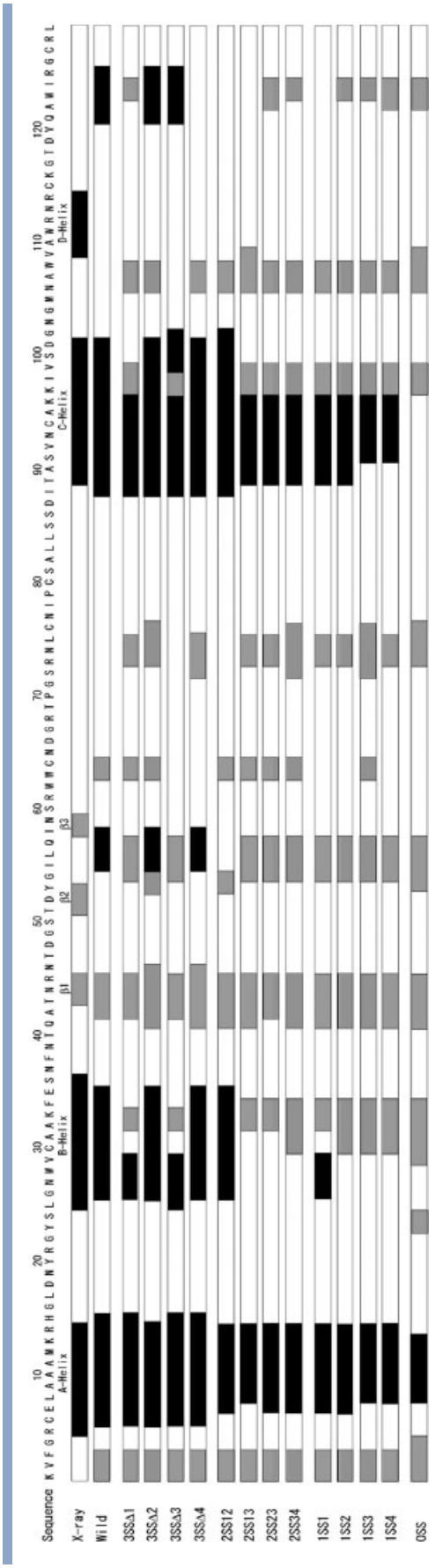

Figure 5
Sequence-based secondary structures of disulfide variants of hen lysozyme predicted by the VUVCD-NN method. The α-helices, β-strands, and unordered structures are shown in black, gray, and white, respectively. The secondary structure regions of the wild type assigned with the DSSP program from the X-ray structure (PDB code: 6LYZ) are indicated along the residue numbers for A-, B-, C-, and D-helices and for β1-, β2-, and β3-strands.

Overall secondary structures of disulfide variant

The results in Table I and Figure 4 provide general features of the secondary structures of the disulfide-deficient variants: α-helices are broken, β-strands increase, turns are almost unchanged, and unordered structures increase due to the deletion of disulfide bridges. The 3SS variants have native-like secondary structures, but the α-helix content in most of the 2SS variants is approximately half that in the native state, and is greatly reduced in the 1SS and 0SS variants. Thus, there seems to be a border between ordered and disordered structures around the 2SS variants, as suggested previously.[11,12] The secondary structure contents of the 3SS, 2SS, and 1SS variants are evidently dependent on the position of disulfide bridges: deletion of SS1 (3SSΔ1) exerts the largest effect of disrupting α-helices among 3SS variants, whereas the SS1 bridge exerts the largest effect of forming α-helices among the 1SS variants. No additivity in the effects of each disulfide bridge on secondary structure formation is apparent, indicating that each disulfide bridge contributes to structural stability in a delicate balance of positive and negative cooperativity.

It is noteworthy that even the 0SS variant contains a considerable amount of secondary structures, and that many β-strands are formed in all the variants. Such large amounts of residual secondary structures have been commonly observed in various types of denatured proteins.[26] Interestingly, there are strong correlations between the β-strand and α-helix contents ($r = -0.95$) and between the unordered-structure and α-helix contents ($r = -0.90$) for all the variants examined, as shown in Figure 6. The increase in unordered structures with decreasing α-helices is reasonable, but the reason for the increases in β-strands is unclear because NMR analyses provided no direct evidence for β-sheets in the 0SS variant. Therefore, the most likely explanation is that most of the β-strands in the variants as well as in other denatured proteins possessing the dihedral angles of β-strands would not contain interpeptide hydrogen bonds, although some β-sheets could remain intact in the 3SS variants, as demonstrated by NMR analyses.[12] Such excess β-strands could form intermolecular β-sheets under appropriate conditions (*e.g.*, high concentrations of protein and salt), initiating amyloid formation. In fact, the 0SS variant easily forms amyloid fibrils,[38] and other disulfide variants have the potential to form amyloid protofibrils under more extreme conditions (unpublished data).

Contribution of conformational entropy

The effects of the numbers and positions of disulfide bridges on the structures of variants can be more quantitatively evaluated using the conformational entropy of the polypeptide chain. Assuming that loop chains

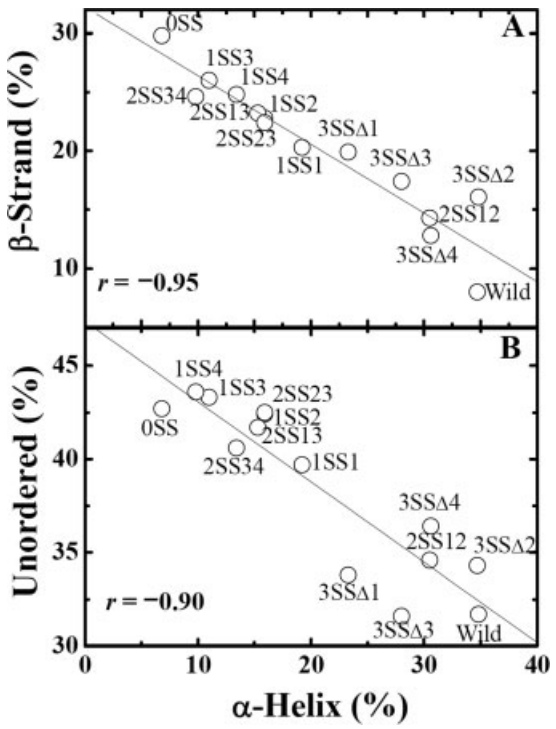

Figure 6

Plots of the β-strand (panel A) and unordered-structure (panel B) contents versus the α-helix content for disulfide variants of hen lysozyme. Solid lines indicate least-squares linear regressions.

conform to Gaussian statistics, the decrease in conformational entropy (ΔS_c) induced by introducing some disulfide bridges into a random chain is

$$\Delta S_c = -R(3.47m + 1.5 \ln |B_m|)$$

where R is the gas constant, m is the number of loops enclosed by the disulfide linkages, and $|B_m|$ the determinant of the matrix whose elements are constructed by the vector connecting sequential Cα atoms of Cys residues participating in cross-linking.[4,8] We used this equation to calculate the ΔS_c value of each variant relative to the 0SS variant, as listed in the last column of Table I. It is evident that ΔS_c becomes more negative as the number of disulfide bridges increases and as they are formed over larger distances in the primary structure when the number of disulfide bridges is constant.

The contribution of conformational entropy to the structure of disulfide variants was examined by plotting the secondary structure contents versus ΔS_c (see Fig. 7). Figure 7(A) shows a strong correlation ($r = -0.88$) between ΔS_c and the α-helix content, indicating that the loss of chain entropy (flexibility) due to cross-linking enhances the formation of α-helices. In contrast, the β-strand content linearly decreases with decreasing ΔS_c ($r = 0.87$) [Fig. 7(B)], suggesting that the chain flexibility is rather favorable for the formation of β-strands. Among each of the 3SS, 2SS, and 1SS variants, the variant having a lower ΔS_c value tends to be richer in α-helices and poorer in β-strands. Thus, the negative correlation observed between the α-helix and β-strand contents (see Fig. 6) can be confirmed in terms of the contributions of conformational entropy. As expected, decreasing ΔS_c decreases the unordered structures ($r = -0.82$) and increases the turns, although the latter correlation is weak ($r = 0.42$) (data not shown). These results constitute evidence that the numbers and positions of disulfide bridges together contribute to the formation of the secondary structures in these disulfide variants dominantly through entropic forces, probably exerting minor effects on the energetics of the folded conformation due to constraints of the polypeptide chain. The entropic forces of disulfide bridges would be a decisive factor not only in the secondary structures but also in the compactness of protein molecules as revealed by the compressibility and expansibility of totally reduced proteins.[33]

Sequences of secondary structures

The secondary structures of three 3SS variants (3SSΔ2, 3SSΔ3, and 3SSΔ4) and two 2SS variants (2SS12 and 2SS34), which have the native-like structure, have been

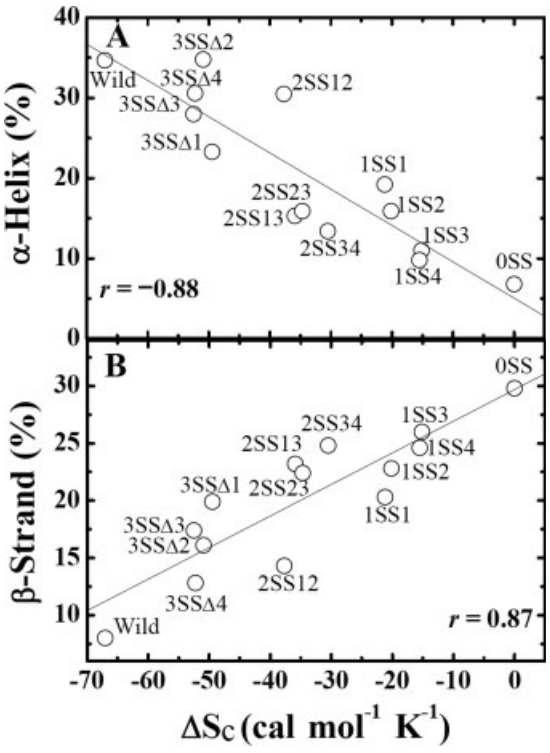

Figure 7

Plots of the α-helix (panel A) and β-strand (panel B) contents versus ΔS_c. Solid lines indicate least-squares linear regressions.

studied by NMR spectroscopy.[11,12] However, such detailed NMR analysis has not been possible for the 1SS and 0SS variants exhibiting large fluctuations due to severe overlaps and/or the extensive line broadening of the NMR signals. At present, there is no other experimental technique for determining the sequences of the unordered structures of disulfide variants or denatured proteins, with only theoretical predictions being possible based on certain assumptions. FTIR and Raman spectroscopy would require a much higher resolution for assigning the amide bands, and X-ray solution scattering is not sensitive to the fine structure of denatured proteins, although the persistence length and the radius of gyration may be simulated by modeling the composition of the structural elements. Therefore, the VUVCD-NN method used in this study, which involves information on the unordered structures, will be a valuable technique for predicting the sequences of variants based on the standards of the native structure although it naturally has the limits of prediction.

As shown in Figure 5, the segments predicted for A-, B-, and C-helices in the wild-type lysozyme mostly agreed with those determined from the X-ray and NMR data. However, a D-helix did not appear at residues 109–114, but it could be predicted to be residues 121–125, which have also been assigned as an α-helix in NMR analysis.[11,12] The segments of β-strands are successfully predicted for β1 (residues 43–45) but not for β2 (residues 51–53) or β3 (residues 58–59). Instead, residues 55–58 (between β2- and β3-strands) are predicted to be an α-helix and residues 63–64 are predicted to be a β-strand. N-terminal residues 1–3 are also predicted to be a β-strand, which is consistent with the NMR data.[11,12] Thus, the deletion effects of disulfide bridges on the secondary structures of lysozyme are predictable at the amino acid sequence level with an accuracy of 76.7%.

The 3SS variants preserve most of the α-helices and β-strands, but their locations in the sequence differ considerably from the structure of the wild type. Large parts of A- and C-helices are preserved, but B- and D-helices might be partly unordered or disrupted. The β1-strand is preserved or slightly elongated, and the regions between the β2- and β3-strands appear to form α-helices or β-strands. Deletion of SS1 (3SSΔ1) among the 3SS variants exerts the greatest effect on the stability of α-helices, probably due to the highest entropic contribution of cross-linking (ΔS_c): the C-terminal residues of B- and C-helices are unordered, and D-helix might be disrupted. This result seems inconsistent with the conclusions of NMR[39] and X-ray[40] studies that reduction and carboxymethylation of Cys6-Cys127 exert no significant effects on the secondary and tertiary structures of lysozyme, except in C-terminal residues. The lower intensity of the VUVCD spectrum for 3SSΔ1 is attributable to the partly unordered structure but not to the contamination of some fully unfolded protein, because this variant is stable at the conditions under which VUVCD measurements were made (pH 2.9 and 25°C), judging from the high similarity with the far-UV CD spectrum at pH 3.8 and 25°C[6] and the pH dependence of thermal stability.[39] The more disordered structures in the VUVCD estimation might be due to electrostatic destabilization of the variant being higher at pH 2.9 than at pH 3.8 for NMR and X-ray analyses.[39,40]

Similar disordering in B- and C-helices occurs in 3SSΔ3, but all α-helices except for the D-helix would be preserved in 3SSΔ2. The structure of 3SSΔ4 is very similar to that of the wild type, although residues 72–75 might form a β-strand and D-helix might be unfolded. These predictions are nearly consistent with the results of NMR analyses at pH 3.8 and 25°C[12]: in 3SSΔ2, the D-helix is unstructured and the interface between B- and D-helices is significantly perturbed; in 3SSΔ3, the entire region of the loop (residues 62–79) is considerably disordered and a network of hydrogen bonds within the β-sheet and the 3_{10}-helix (residues 80–84) in the β-domain is disrupted and fluctuating; and the structure of 3SSΔ4 is quite similar to that of the wild type except for the peptide segment of residues 74 to 78. Thus, the 3SS variants essentially exhibit a native-like folded structure, as supported by their residual enzyme activity.[7]

In the 2SS variants, large parts of A- and C-helices are preserved but the B-helix is preserved only in 2SS12 and the D-helix might be totally disrupted. Some residues in B- and C-helices and probably in the D-helix could change to β-strands. The β1-strand is preserved or slightly elongated, as found for the 3SS variants, but the β2- and β3-strands seem to converge in a β-strand. The predicted sequences for 2SS12 are not necessarily consistent with the NMR data,[11] in which the structure of the α-domain in 2SS12 was quite similar to that of the wild type, while the β-domain is largely unstructured and the β1- and β2-strands are maintained stably as an antiparallel β-sheet. NMR data also suggested that the extent of unfolding in the 2SS34 variant made detailed structural analysis impossible.[11] However, the results from the present study indicate that A- and C-helices and β1-strand are partly preserved, although the tertiary structure might be disrupted. These discrepancies between our predictions and NMR data cannot be reasonably explained at present because our experimental conditions differ from those of NMR analyses (pH 3.8 and 95% D_2O) and structures exhibiting large fluctuations are sensitive to the solvent conditions.

The 1SS variants preserve a core part of A- and C-helices but lose B- and D-helices, although in the 1SS1 variant, the four N-terminal residues of the B-helix are conserved probably due to the strongest constraint of chain flexibility. The A- and C-helices are shorter in 1SS3 and 1SS4 than in 1SS1 and 1SS2, suggesting that the structures of SS3- and SS4-associated variants are intrinsically unstable. These predictions are consistent with the previ-

ous results obtained from the far-UV CD spectra and the effective concentration of protein thiol groups.[8] Since no 1SS variants showed near-UV CD bands and bacteriolytic activity,[8] the residual secondary structures would not be sufficient to form a globally folded structure. The 0SS variant loses all of the helices except for a core part of the A-helix, and hence this variant has the most unstable structure among the disulfide variants examined. Such a high stability of the A-helix core is consistent with the helix-probability profile of lysozyme as calculated using the helix-nucleation and helix-propagation parameters for a denatured protein.[41]

An interesting finding in the sequence-based prediction is that the β-strands are distributed over all the sequences, with some of them being located individually in a fixed position. A similar large distribution of β-strand-rich regions has been observed as short-range NOE cross-peaks $\alpha N\ (i,\ i\ +\ 1)$ in NMR analysis.[11,12] The three N-terminal residues and the β1-strand are completely preserved in all variants, as also found in NMR analyses. The propensity for β-strand is strong for residues 54–57, 73–75, 97–99, and 106–108, moderate for residues 30–34, and weak for residues 63–64 and 122–123, depending on the variants or the contents of residual secondary structures. Among these β-strands, the β1-strand can form hydrogen bonds (β-sheets) with the regions (residues 54–57) around the β2- and β3-strands, as detected by NMR in the 3SS and 2SS12 variants. However, it is unlikely that other strands participate in hydrogen bonding with each other since no such stable β-sheets have been observed in NMR investigations.[11] If this is the case, some of these β-strands could form intermolecular β-sheets when the protein concentration is high, leading to amyloid formation as mentioned earlier. Considering that protein unfolding is an indispensable step in amyloid formation, it is probable that the residues whose propensity changes from α-helix to β-strand are candidate initiators for the intermolecular β-sheets of these variants.

Roles of disulfide bridges on secondary structures

The predicted sequence-based secondary structures of the variants could represent useful information on the role of each disulfide bridge in secondary structure formation, although they cannot address the formation of β-sheets. As shown in Figure 5, introducing a single disulfide bridge into the 0SS variant (in which the A-helix is incompletely formed) induces part of a C-helix. The helix-inducing ability of each disulfide bridge increases as it is crossed over larger distances in the primary structure, and the SS1 bridge at the largest distance (1SS1) could further induce a B-helix. The second disulfide bridge introduced into the 1SS variants exerts no significant helix-inducing effects, but SS2 (Cys30-Cys115) clearly enhances the formation of B- and C-helices in 1SS12, probably because Cys30 located in the B-helix depresses the chain flexibility of α-domain through cross-linking with Cys115. Two disulfide bridges, SS4 and SS3, each of which has a weak helix-inducing ability, would cooperatively stabilize the C-helix. In the 3SS variants, the absence of a disulfide bridge has no significant effect on the stability of the A-helix, but either SS1 or SS3 would be indispensable for the completion of B- and C-helices. Although there are limitations in evaluating the nonadditive effects of each disulfide bridge, these speculations could complement NMR studies that take account of the effects of hydrophobic interaction and folding kinetics.[9,11,12]

CONCLUSIONS

VUVCD spectroscopy has revealed the characteristic secondary structures of disulfide variants of lysozyme depending on the positions and numbers of deleted disulfide bridges. In general, the secondary structures were more effectively stabilized through entropic forces as the number of disulfide bridge increased and as it was formed over the larger distance in the primary structure. The sequence-based secondary structures predicted by the VUVCD-NN method provided useful information on the role of a specific disulfide bridge in secondary structure formation. The present study demonstrates that VUVCD spectroscopy is a powerful tool for comprehensively analyzing the secondary structures of proteins from native to denatured states.

ACKNOWLEDGMENTS

The authors thank Professors Hirofumi Namatame and Masaki Taniguchi at the Hiroshima Synchrotron Radiation Center for their technical support of the use of the VUVCD spectrophotometer.

REFERENCES

1. Anfinsen CB, Haber E, Sela M, White FH Jr. The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain. Proc Natl Acad Sci USA. 1961;47:1309–1314.
2. Creighton TE. Role of the environment in the refolding of reduced pancreatic trypsin inhibitor. J Mol Biol 1980;144:521–550.
3. Gilbert HF. The formation of native disulfide bonds. In: Pain RH, editor. Mechanism of protein folding. Cambridge: Oxford University Press; 1994. pp 104–136.
4. Lin SH, Konishi Y, Denton ME, Scheraga HA. Influence of an extrinsic cross-link on the folding pathway of ribonuclease A. Conformational and thermodynamic analysis of cross-linked (lysine^{7}-lysine^{41})-ribonuclease A. Biochemistry 1984;23:5504–5512.
5. Creighton TE. Disulphide bonds and protein stability. Bioessays 1988;8:57–63.
6. Sawano H, Kokumoto Y, Ohta K, Sasaki Y, Segawa S, Tachibana H. Efficient in vitro folding of the three-disulfide derivatives of hen lysozyme in the presence of glycerol. FEBS Lett 1992;303:11–14.

7. Tachibana H, Ohta Y, Sawano H, Koumoto Y, Segawa S. Relationship between the optimal temperature for oxidative refolding and the thermal stability of refolded state of hen lysozyme three-disulfide derivatives. Biochemistry 1994;33:15008–15016.
8. Tachibana H. Propensities for the formation of individual disulfide bonds in hen lysozyme and in the size and stability of disulfide-associated submolecular structures. FEBS Lett 2000;480:175–178.
9. Yokota A, Izutani K, Takai M, Kubo Y, Noda Y, Koumoto Y, Tachibana H, Segawa S. The transition state in the folding-unfolding reaction of four species of three-disulfide variant of hen lysozyme: the role of each disulfide bridge. J Mol Biol 2000;295:1275–1288.
10. Tachibana H, Oka T, Akasaka K. Native-like tertiary structure formation in the α-domain of a hen lysozyme two-disulfide variant. J Mol Biol 2001;314:311–320.
11. Noda Y, Yokota A, Horii D, Tominaga T, Tanisaka Y, Tachibana H, Segawa S. NMR structural study of two-disulfide variant of hen lysozyme: 2SS[6–127, 30–115]-a disulfide intermediate with a partly unfolded structure. Biochemistry 2002;41:2130–2139.
12. Yokota A, Hirai K, Miyauchi H, Iimura S, Noda Y, Inoue K, Akasaka K, Tachibana H, Segawa S. NMR characterization of three-disulfide variants of lysozyme. C64A/C80A, C76A/C94A, and C30A/C115A-a marginally stable state in folded proteins. Biochemistry 2004;43:6663–6669.
13. Berova N, Nakanishi K, Woody RW. Circular dichroism: Principles and applications. 2nd ed. New York, NY: Willy-VCH Press; 2000.
14. Kabsch W, Sander C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers 1983;22:2577–2637.
15. King SM, Johnson WC. Assigning secondary structure from protein coordinate data. Proteins 1999;35:313–320.
16. Sreerama N, Woody RW. Estimation of protein secondary structure from circular dichroism spectra: comparison of CONTIN. SELCON, and CDSSTR methods with an expanded reference set. Anal Biochem 2000;287:252–260.
17. Sreerama N, Venyaminov SY, Woody RW. Estimation of number of α-helical and β-strand segments in proteins using circular dichroism spectroscopy. Protein Sci 1999;8:370–380.
18. Unneberg P, Merelo JJ, Chacón P, Morán F. SOMCD: method for evaluating protein secondary structure from UV circular dichroism spectra. Proteins 2001;42:460–470.
19. Toumadje A, Alcorn SW, Johnson WC Jr. Extending CD spectra of proteins to 168 nm improves the analysis for secondary structures. Anal Biochem 1992;200:321–331.
20. Snyder PA, Rowe EM. The first use of synchrotron radiation for vacuum ultraviolet circular dichroism measurements. Nucl Instrum Methods Phys Res A 1980;172:345–349.
21. Sutherland JC, Emrick A, France LL, Monteleone DC, Trunk J. Circular dichroism user facility at the National Synchrotron Light Source: estimation of protein secondary structure. Biotechniques 1992;13:588–590.
22. Wallace BA, Janes RW. Synchrotron radiation circular dichroism spectroscopy of proteins: secondary structure, fold recognition and structural genomics. Curr Opin Chem Biol 2001;5:567–571.
23. Jones GR, Clarke DT. Applications of extended ultra-violet circular dichroism spectroscopy in biology and medicine. Faraday Discuss 2004;126:223–236.
24. Matsuo K, Yonehara R, Gekko K. Secondary-structure analysis of proteins by vacuum-ultraviolet circular dichroism spectroscopy. J Biochem (Tokyo) 2004;135:405–411.
25. Matsuo K, Yonehara R, Gekko K. Improved estimation of the secondary structures of proteins by vacuum-ultraviolet circular dichroism spectroscopy. J Biochem (Tokyo) 2005;138:79–88.
26. Matsuo K, Sakurada Y, Yonehara R, Kataoka M, Gekko K. Secondary-structure analysis of denatured proteins by vacuum-ultraviolet circular dichroism spectroscopy. Biophys J 2007;92:4088–4096.
27. Matsuo K, Watanabe H, Gekko K. Improved sequence-based prediction of protein secondary structures by combining vacuum-ultraviolet circular dichroism spectroscopy with neural network. Proteins 2008;73:104–112.
28. Ojima N, Sakai K, Matsuo K, Matsui T, Fukazawa T, Namatame H, Taniguchi M, Gekko K, Vacuum-ultraviolet circular dichroism spectrophotometer using synchrotron radiation: optical system and on-line performance. Chem Lett 2001;30:522–523.
29. Matsuo K, Sakai K, Matsushima Y, Fukuyama T, Gekko K. Optical cell with a temperature-control unit for a vacuum-ultraviolet circular dichroism spectrophotometer. Anal Sci 2003;19:129–132.
30. Jones DT. Protein secondary structure prediction based on position-specific scoring matrices. J Mol Biol 1999;292:195–202.
31. Bernstein FC, Koetzle TF, Williams GJB, Meyer EF Jr., Brice MD, Rodgers JR, Kennard O, Shimanouchi T, Tasumi M. The Protein Data Bank: a computer-based archival file for macromolecular structures. J Mol Biol 1977;112:535–542.
32. Grishina IB, Woody RW. Contributions of tryptophan side chains to the circular dichroism of globular proteins: exciton couplets and coupled oscillators. Faraday Discuss 1994;99:245–262.
33. Gekko K, Kimoto A, Kamiyama T. Effects of disulfide bonds on compactness of protein molecules revealed by volume, compressibility, and expansibility changes during reduction. Biochemistry 2003;42:13746–13753.
34. Pancoska P, Janota V, Keiderling TA. Novel matrix descriptor for secondary structure segments in proteins: demonstration of predictability from circular dichroism spectra. Anal Biochem 1999;267:72–83.
35. Adamczak R, Porollo A, Meller J. Combining prediction of secondary structure and solvent accessibility in proteins. Proteins 2005;59:467–475.
36. Wood MJ, Hirst JD. Protein secondary structure prediction with dihedral angles. Proteins 2005;59:476–481.
37. Bondugula R, Xu D. MUPRED: a tool for bridging the gap between template based methods and sequence profile based methods for protein secondary structure prediction. Proteins 2007;66:664–670.
38. Niraula TN, Konno T, Li H, Yamada H, Akasaka K, Tachibana H. Pressure-dissociable reversible assembly of intrinsically denatured lysozyme is a precursor for amyloid fibrils. Proc Natl Acad Sci USA 2004;101:4089–4093.
39. Radford SE, Woolfson DN, Martin SR, Lowe G, Dobson CM. A three-disulphide derivative of hen lysozyme. Structure, dynamics and stability. Biochem J 1991;273:211–217.
40. Hill CP, Johnston NL, Cohen RE. Crystal structure of a ubiquitin-dependent degradation substrate: a three-disulfide form of lysozyme. Proc Natl Acad Sci USA 1993;90:4136–4140.
41. Lewis PN, Scheraga HA. Prediction of structural homology between bovine α-lactalbumin and hen egg white lysozyme. Arch Biochem Biophys 1971;144:584–588.
42. DeLano WL. The PyMOL Molecular Graphics System. Palo Alto, CA, USA: DeLano Scientific; 2002.