Experimental Evidence for the Structure of Lysozyme in the Transition State of Folding

Shin-ichi SEGAWA

School of Science and Technology, Kwansei Gakuin University,

2-1 Gakuen, Sanda, Hyogo 669-1337, Japan

PREFACE

Protein folding is the physical process by which a polypeptide chain spontaneously folds into its characteristic three-dimensional structure from an unfolded conformation. Amino acid residues interact with each other through different types of interactions to produce a specific tertiary structure, which is determined by the amino acid sequence. In 1965, the three-dimensional structure of lysozyme was solved with the X-ray diffraction method. I, an undergraduate student in a physics course, thought it would become possible in the near future to elucidate the protein folding in atomic detail. My doctoral research started in 1970. I imagined that the ultimate goal of researches on protein folding was predicting the three-dimensional structure characteristic to the amino acid sequence or designing novel proteins. At that time, a few laboratories had got started on fast reaction kinetics of protein unfolding and refolding, and also I started kinetic studies of lysozyme folding by means of the temperature-jump method. In the era from the 1970s to the 1980s, many new powerful techniques had been developed aiming at the elucidation of protein folding mechanism. Two-dimensional NMR spectroscopy, site-directed mutagenesis, a highly sensitive micro-calorimeter, hydrogen/deuterium exchange method combined with the NMR spectroscopy, and so on became available for the investigation of protein folding. However, it was a very difficult problem to infer the folding pathway because of the two-state nature of the folding transition and the conformational heterogeneity of the unfolded state. I noticed that it was rather practicable to get information about the structure of protein in the transition state of two-state kinetics. Consequently, I have studied the folding mechanism of lysozyme directing our attention to the transition state. In order to study the structure of lysozyme in the transition state at the atomic resolution, we prepared various types of disulfide-deficient variants of lysozyme as protein samples with a marginal stability, and have investigated them by chemical kinetics and/or NMR structural studies. In March of 2014, I retired from Kwansei Gakuin University in which I had been for 38 years since 1976. On the basis of my last lecture held on Feb. 24 of 2014, I organized my studies on the protein folding into this article.

Sanda, Hyogo, JAPAN Shin-ichi SEGAWA
May, 2015

ABSTRACT: The folding mechanism of protein is not unveiled yet in atomic detail of a specific protein. It is primarily because the folding reaction is a two-state transition between the native (N) and unfolded (U) states. Short-range and long-range interactions among adjacent residues in the native structure simultaneously work together upon the protein folding, and their cooperativity makes it a two-state transition. It was exemplified by computer simulations that the more cooperative the specific interactions were, the more definite the folding pathway was. However, it was almost impossible due to the two-state kinetics to infer the folding pathway by detecting directly folding intermediates. Therefore, we thought it was rather practicable to get information about the structure of protein in the transition state by evaluating the effect of either varying the solvent condition or modifying the protein on folding and unfolding rates. Consequently, we have studied kinetics of lysozyme folding focusing our attention on the transition state. At the beginning, we reexamined whether the transition state really existed or not in the protein folding composed of vast numbers of elementary steps. Computer simulations of island-model proteins afforded evidence that a single transition state actually existed as a turning point between N and U states, and the transition state theory was applicable to the analysis of folding and unfolding rate constants. By examining the change in heat capacity from the N state to the transition state, we derived a conclusion that the transition state of lysozyme was like a "dry molten globule" with diminished non-covalent interactions, but without water inside. Further, we obtained useful information about the transition state of lysozyme folding by evaluating effects on folding and unfolding rates of the modification of specific residues. Later, this study was more generalized and formulated as the φ-value analysis. Finally, we performed NMR observation of different species of disulfide-deficient variants of lysozyme. Characteristics of the dynamic structure of each three-disulfide variant of lysozyme determined by NMR observations were entirely consistent with those of the transition state obtained from their φ-value analysis. It supports the idea that the transition state is the limit of structural fluctuations occurring in the native state. By examining the marginal state of the disulfide-deficient lysozymes, we succeeded in elucidating the structure of lysozyme in the transition state in atomic detail. It is summarized in the section 3 of Chapter X.

Contents

V. CHARACTERIZATION OF THE TRANSITION STATE

VI. NMR STRUCTURAL STUDIES ON 3SS-VARIANTS OF LYSOZYME: LIMITS OF THE NATIVE STRUCTURE

VII. STRUCTURAL FLUCTUATIONS STUDIED BY THE H/D EXCHANGE METHOD

VIII. TWO-DISULFIDE VARIANTS ARE ON THE BORDER LINE BETWEEN ORDERED AND DISORDERED STATES

IX. GLYCEROL-INDUCED FOLDING OF 2SS[6-127, 64-80]

X. LATENT STRUCTURES IN UNSTRUCTURED 1SS-VARIANTS

Abbreviations

GuHCl	guanidinium hydrochloride
triNAG	N-acetyl-D-glucosamine trimer
DMSO	dimethyl sulfoxide
N state	native state
U state	unfolded state
MG	molten globule
T state	transition state
ΔH	change in enthalpy
ΔE	change in energy
ΔC	change in heat capacity
k_f	rate constant of folding
k_{uf}	rate constant of unfolding
Cys6-Cys127	disulfide bridge between Cys6 and Cys127
4SS-lyz	recombinant hen lysozyme containing four authentic disulfide bridges and an extra N-terminal Met
3SS-varinat	three-disulfide variant of lysozyme lacking any one of four authentic disulfide bridges
2SS-variant	two-disulfide variant of lysozyme lacking any two disulfide bridges and leaving the others
1SS-variant	one-disulfide variant of lysozyme preserving any one disulfide bridge
0SS-variant	lysozyme variant lacking all of the disulfide bridges
6-127 rcm-lyz	Cys6-Cys127 reduced carboxymethylated derivative of lysozyme
des(6-127) variant	3SS-variant lacking disulfide bridge Cys6-Cys127: C6S/C127A
des(30-115) variant	3SS-variant lacking disulfide bridge Cys30-Cys115: C30A/C115A
des(64-80) variant	3SS-variant lacking disulfide bridge Cys64-Cys80: C64A/C80A
des(76-94) variant	3SS-variant lacking disulfide bridge Cys76-Cys94: C76A/C94A
2SS[6-127, 30-115]	2SS-variant preserving Cys6-Cys127 and Cys30-Cys115
1SS[6-127]	1SS-variant preserving Cys6-Cys127
α-domain	N-terminal half: residues 1-39 and C-terminal half: 88-129
β-domain	peptide segment from 40 through 87
A-helix	peptide segment from 4 through 15
B-helix	peptide segment from 24 through 36
C-helix	peptide segment from 89 through 99

D-helix	peptide segment from 108 through 115
β3_{10}-helix	3_{10}-helix existing in the β-domain: residues 80-84
ct3_{10}-helix	3_{10}-helix existing near at the C-terminal end: residues 120-124
β1-strand	peptide segment from 41 through 46
β2-strand	peptide segment from 50 through 54
β3-strand	peptide segment from 57 through 61
Loop	peptide segment from 62 through 79
N-terminus	amino terminus
C-terminus	carboxyl terminus
2D and 3D	two- and three-dimensional
CD	circular dichroism
H/D	hydrogen/deuterium
NMR	nuclear magnetic resonance
NOE	nuclear Overhauser effect
HSQC	heteronuclear single-quantum coherence
NOESY	nuclear Overhauser effect spectroscopy
NN(*i*,*i*+1)	^{1}HN-^{1}HN NOE cross-peak between *i*-th and (*i*+1)-th residues
αN(*i*,*i*+1)	^{1}HC$_{\alpha}$-^{1}HN NOE cross-peak between *i*-th and (*i*+1)-th residues
αN(*i*,*i*+3)	^{1}HC$_{\alpha}$-^{1}HN NOE cross-peak between *i*-th and (*i*+3)-th residues
^{15}N-^{15}N	sequential connectivity between ^{15}N nuclei of adjacent residues

Abbreviations for amino acids

Ala: A	alanine
Arg: R	arginine
Asn: N	asparagine
Asp: D	aspartic acid
Cys: C	cysteine
Glu: E	glutamic acid
Gln: Q	glutamine
Gly: G	glycine
His: H	histidine
Ile: I	isoleucine
Leu: L	leucine
Lys: K	lysine
Met: M	methionine
Phe: F	phenylalanine
Pro: P	proline
Ser: S	serine
Thr: T	threonine
Trp: W	tryptophan
Tyr: Y	tyrosine
Val: V	valine

I. INTRODUCTION

1. Phillips's scenario of lysozyme folding

In 1970, my studies on protein folding started when I was a graduate student of the University of Tokyo. For preparations of my research work, I used to read a paper entitled "The three-dimensional structure of an enzyme molecule" written by David, C. Phillips, which had been published in *Scientific American* (1). The enzyme is lysozyme, whose three-dimensional structure had been solved with the X-ray diffraction method by D. C. Phillips and his colleagues at the Royal Institution in London (2). The article described the structure of lysozyme in atomic detail. For example, hen lysozyme is composed of four α-helices, three β-strands and two 3_{10}-helices (see the caption of Figure 1). The three-dimensional structure is roughly divided into two domains: the α-domain (residues 1-39, and 88-129) and the β-domain (residues 40-87). There exists a deep cleft between the α- and β-domains, which forms the active site of the enzyme. On the basis of the arrangement of atoms in the substrate binding site, the study successfully demonstrated how lysozyme performed its task as an enzyme. Lysozyme is the first enzyme whose function was elucidated at the atomic resolution. Furthermore, the author proposed an attractive hypothesis that the folding of the protein chain to its native conformation began at the N-terminus of the chain even before the synthesis was complete. Parts of the polypeptide chain, particularly those near the N-terminus, may fold into stable conformations that can still be recognized in the final native structure. Although this proposal put forward by him is not accepted now, his insight into the folding of lysozyme obtained from the inspection of its three-dimensional structure is great even today.

In brief, the scenario of lysozyme folding suggested by him is as follows. The first 40 residues from the N-terminus form a compact substructure before the molecule is fully synthesized. The next residues 41 through 54 form the anti-parallel β-sheet, which lies out of contact with the compact submolecule formed by the earlier residues, and the hydrophobic side-chains of Ile55 and Leu56 are buried in the initial hydrophobic pocket. Residues 57 through 86 are folded in contact with the β-sheet structure so that the folded chain forms a structure with two wings lying at an angle to each other. Residues 86 through 96 form a length of α-helix, one side of which is predominantly hydrophobic. This helix lies in the gap between the two wings with its hydrophobic side buried within the molecule. The remaining residues are folded around the globular unit formed by the N-terminal end of the polypeptide chain.

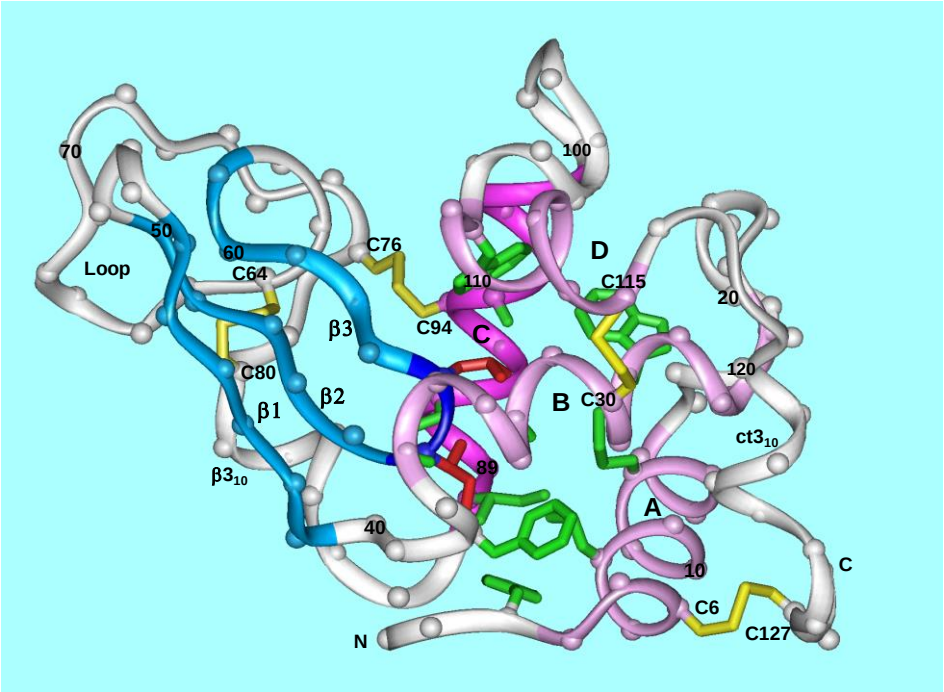

Figure 1. Ribbon diagram of hen lysozyme, which was reproduced on the basis of the X-ray crystallographic structure (PDB 6LYZ). Balls on the ribbon indicate the position of C_α atom of each amino acid residue, and residue numbers are shown every ten residues on the balls. Four α-helices, labeled A (residues 4-15), B (24-36), C (89-99) and D (108-115) are colored magenta, while three β-strands, labeled β1(41-46), β2(50-54) and β3(57-61) are colored light blue. The ribbon of residues 55-56 are colored dark blue and the side-chains of I55 and L56 are represented by red sticks. A loop region (62-79) and 3_{10}-helix (80-84) are referred to as Loop and $\beta3_{10}$. Another 3_{10}-helix (120-124) appears near at the C-terminus ($ct3_{10}$). Yellow sticks represent four disulfide bridges: C6-C127, C30-C115, C64-C80 and C76-C94. Side chains of several residues surrounding I55 and L56 are shown by green sticks (F3, L8, M12, W28, A32, S36, F38, I88, S91, A95 and W108).

A target of my research was experimentally to demonstrate his proposal. I started my research by studying the kinetics of lysozyme folding. I used to think it was important in elucidating the folding mechanism to detect directly folding intermediates by some spectroscopic method, and particularly to describe conformation changes of a specific protein in atomic detail. About 29 years later from that time, I saw the photograph of the first public representation by Dr. Phillips held at the Royal Institution in 1965 in the article entitled "The early history of lysozyme", which had been written by L. N. Johnson and published in *Nature Structural Biology* (3). In the photograph (Figure 2), a long polypeptide chain of lysozyme was hanging down from the ceiling, and there was the three-dimensional wire-model of the lysozyme molecule on the desk beside Dr. Phillips. It can readily be imagined that his primary interest was the elucidation of the protein folding from that time.

Figure 2. First public representation by Dr. Phillips held at the Royal Institution in 1965. This photograph was reprinted from *Nature Structural Biology* **5** (1998), 942-944.

2. Eigen's theory of selforganization of matter

Another scientist who had strong influence on my studies was Manfred Eigen. He won the 1967 Nobel Prize in Chemistry for his studies on the kinetics of extremely fast chemical reactions with relaxation methods. I used to study the kinetics of lysozyme folding by the temperature-jump method (4,5), which had been developed by M. Eigen and his colleagues (6). In 1971, he published an epoch-making paper entitled "Selforganization of matter and the evolution of biological macromolecules" in *Naturwissenshaften* (7). In 1977, he published a trilogy (8,9) entitled "The Hypercycle" in *Naturwissenshaften*, following the preceding paper. The articles put forward a novel theory to explain the evolution of self-reproducing matter and the origin of life. His theory is developed on the network of chemical reactions, particularly self-replication reactions through an autocatalytic reaction. Necessary prerequisites for the evolution are the self-reproduction of molecular species from energy-rich building materials, their decomposition to ingredients, the incomplete precision of self-reproduction, and a natural selection of the advantage of molecular species through a feedback system. Under some constraint such as constant overall population, the natural selection occurs from an ensemble of random molecular species to a major molecular species. He clearly explained the natural selection and the evolution of self-reproducing molecular species not only by

theoretical treatments but also by computer simulations. General physical principles on the evolution of biological macromolecules and the origin of life gave me a deep impression. In the computer game (8) presented by him to exemplify the theory on the natural selection of self-reproducing molecules, I felt some analogy to the protein folding problem in a sense that the goal in the former game was to create a meaningful message from a random sequence of letters, while in the latter, it was to make an ensemble of random structures of polypeptide chain converge into a unique three-dimensional structure. In 1980, I started a study on the computer simulation of island-model proteins, and attempted to get physical principles of protein folding (10).

Finally, my research works have been focused on two themes; experimental studies in atomic detail on the folding mechanism of a specific protein: lysozyme, and the pursuit of physical principles of the protein folding by chemical kinetics and computer simulations. Around 1980, I noticed that it was important to elucidate the structure of protein in the transition state, after that we have studied the protein folding directing our attention to the transition state. We prepared a total of 12 kinds of disulfide bridge-deleted (disulfide-deficient) variants of lysozyme in order to see a marginal state of the tertiary structure of lysozyme. We had performed NMR structural studies of them until I retired from Kwansei Gakuin University in March of 2014. This article is a compilation of the studies.

II. PROTEIN FOLDING PROBLEM

1. Both folding and unfolding simultaneously occur between N and U

In most cases of small globular proteins, folding and unfolding processes simultaneously occur at two-state kinetics (11), except for a quick formation of the so-called molten globule which has been observed just after an abrupt change in solvent. That is to say, upon the folding from fully unfolded state (U) in the presence of concentrated denaturants such as 6.0 M guanidinium hydrochloride (GuHCl), a rapid change into a collapsed non-native state was observed immediately after the removal of denaturants, and then the folding to the native state (N) followed (12). Such a transient state is called the molten globule state (MG). Molten globules are collapsed and generally have some native-like secondary structure unlike the fully unfolded protein, but they have no persistent tertiary structure unlike the native protein. In such a sense, MG became to be considered a popular folding intermediate state (13). Originally, the term of MG came from a state of proteins found under mildly denaturing conditions (14). It is still unclear in most proteins whether the transition between U and MG is a two-state transition or not. It seems a rather continuous transition without cooperativity. On the contrary, the transition from MG to N is highly cooperative. Even in the case of bovine α-lactalbumin which has been extensively studied on the formation of the molten globule, NMR observation demonstrated that the close packing characteristic of the N state emerged all over the molecule in a highly cooperative manner following the rapid formation of the MG (15).

Under a physiological condition, both native and non-native states must coexist, even though the population of non-native species is so small as to be barely detectable. The non-native state that equilibrates with the native state under a physiological condition seems an ensemble of molecular species of various non-native conformations including the molten globule. Now we focus our attention on the folding process from such a non-native state to the native state, and on the reverse as the unfolding process. Although much effort has been expended to characterize the molten globule state, not much attention has been paid to the late stage of folding. If the native species is abbreviated as N, and the non-native one as U, we can only detect either N or U at the rate-determining step toward the native state. No folding intermediate is experimentally detectable at two-state kinetics. Is it a meaningful inquiry to observe the folding pathway?

I noticed that it made a large contribution to elucidating the folding pathway to characterize the transition state in the two-state kinetics between N and U, particularly to detect structural features of protein in the transition state. Since thermodynamic properties

of the transition state are reflected on folding and unfolding rates, I thought it became possible by evaluating the effect of either varying the solvent condition or modifying the protein on folding and unfolding rates. Consequently, we have attempted to obtain structural information of lysozyme in the transition state with such a method (16,17). First, in Chapter III, we reexamine the transition state, which is only a hypothetical state to analyze the two-state kinetics, according to a reaction scheme composed of vast numbers of elementary steps. In Chapter IV, Monte Carlo simulations of model proteins exemplifies that the transition state theory is reasonable to analyze the folding and unfolding rate constants (10). Second, we have prepared various kinds of lysozyme derivatives by introducing an extra cross-link or by constructing different species of disulfide-deficient variants, and evaluated the effects of modifications on folding and unfolding rate constants (17,18). The results are described in Chapter V. Finally, in Chapter VI to X, we examine the structure of lysozyme in the transition state at the atomic resolution, based on NMR observations of the disulfide-deficient variants (19-22).

2. There is a proper sequential order in which building blocks are assembled

Specific interactions among neighboring residues in the native structure are necessary for conducting proteins to a unique native one. They are classified into long-range and short-range groups according to the sequence separation of $|i-j|$ between i-th and j-th amino acid residues. Generally speaking, short-range interactions are required to accelerate the protein folding, and long-range ones are necessary for having proteins stay in the vicinity of the native structure. Cooperativity among these specific interactions, namely their working together, makes the folding two-state kinetics. It is likely that local native conformations of various shapes are generated at various sites along a polypeptide chain and merged with each other, but such structural elements are frequently broken before arriving at the transition state that is the turning point between N and U. Upon the folding transition, most of building blocks tend to emerge at almost the same time with intervening in the non-native chain. However, protein folding cannot be accomplished by assembling them at random, because the steric overlap among them prevents building blocks from merging with each other. Only when they are assembled in some proper order, the reaction may get over the transition state. It takes long time to find the proper order, because a single molecule stays for most of its time in either U or N, even though a single transition between them itself finishes in a short time. This is the scenario of protein folding that I have imagined. The folding pathway is equivalent to the sequential order in which building blocks are successfully assembled into the final native structure. Although the folding is the all-or-none transition between N and U, and seems products of chance,

there must exist some folding pathways. Even though there are multiple folding pathways in principle, a major route must exist among them.

3. Tertiary structures of proteins are like a three-dimensional jigsaw puzzle

It is well known that the helix-coil transition of a cylindrical helix is not a two-state transition. It is referred to as a continuous transition, that is to say, the average helix content of a single molecule continuously increases with the change in the equilibrium state from a coil to a helix. In general, small-size helices are generated at various sites along a polypeptide chain, and merged with each other. In this case, it is possible to initiate the formation of the whole helix at any sites, and there are many equally-probable pathways leading to the final cylindrical helix. Also, it is possible to break the helix at any sites into pieces. Protein folding is analogous to a jigsaw puzzle in terms of the close packing of a lot of small pieces of different shapes. Since an ordinary jigsaw puzzle is built on a plane, it is possible to start packing small pieces at any sites: from a peripheral area or a central part. The sequential order to fit pieces together is not important in completing the picture. In such a sense, the formation of a cylindrical helix is able to be compared to assembling a two-dimensional jigsaw puzzle.

On the other hand, the tertiary structure of a protein seems to be composed of several building blocks of different shapes, whose structures themselves must be compact and relatively stable in a three-dimensional space. The cooperativity between stabilization energies of building blocks themselves and long-range interactions among them makes the protein folding an all-or-none transition. It means that structures of building blocks being stable by themselves are still preserved in the final native structure of protein, although rearrangements of side chains may be necessary. In order to unfold a compact three-dimensional structure into two building blocks whose original structures remain preserved, a flexible hinge is required. That is to say, at the one or two single bonds making up such a hinge, two rigid blocks must be movable enough to permit significant angular rotation around the hinge bonds without severe steric overlap. Our previous study demonstrated that there were limited numbers of potential flexible hinges in the native structure of proteins, using a simple approximation that proteins were constructed of hard-sphere atoms (23). In other words, we cannot unfold the tertiary structure of proteins into two building blocks around any site except the potential flexible hinges. For example, Figure 3 shows the hinge motion of two building blocks of tuna cytochrome c around the residue 56 (23). Even though each building block, (1-55) or (56-103), is folded native-like in advance, they are able to merge with each other without any obstacles. Indeed, two isolated peptide fragments, (1-44) and (45-103), which were analogous to the building

blocks mentioned above, formed the stable one-to-one fragment-complex with a conformation similar to the native structure of the intact protein (24). It suggests that the

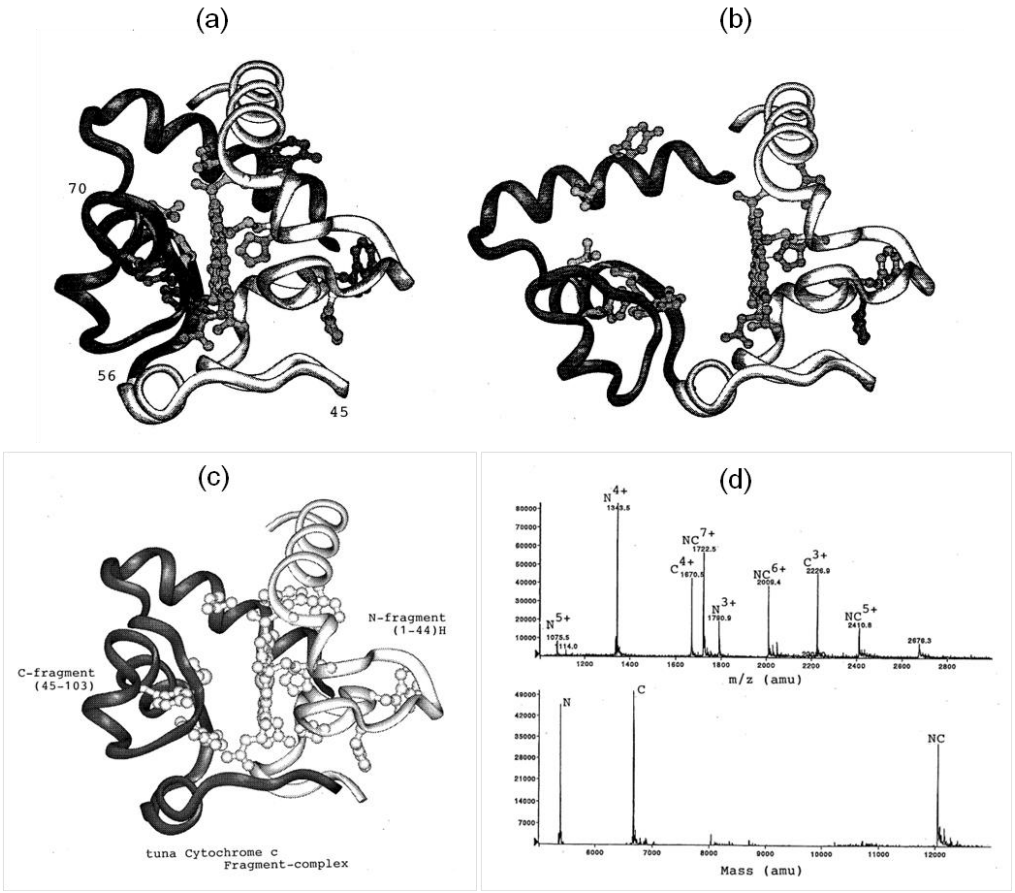

Figure 3. The hinge motion around the residue 56 of tuna cytochrome c (cyt c). The ribbon diagram was reproduced from the X-ray crystallographic structure (PDB 3CYT). (a) The native structure of tuna cyt c. The ribbon of residues 1-55 is colored in white, while that of 56-103 in black. The heme group is connected to the protein framework by covalent bonds with Cys14 and Cys17, and the heme iron is coordinated to His18 and Met80. Suppose that the peptide segment 1-6 is flexible enough, two peptide domains, 1-55 and 56-103, are mutually movable without severe obstacle due to a steric overlap with preserving the native structure of each domain. (b) The hypothetical protein conformation is drawn after the mutual movement around the hinge. The dihedral angles, (φ,ψ), of the residue 56 were changed from the native ones by $\Delta\varphi$= -24º and $\Delta\psi$= -24º. (c) Tuna cyt c fragment-complex. Peptide fragments, 1-44 and 45-103, were prepared by the limited digestion of tuna cyt c by V8 protease and were isolated and purified as the N-fragment (1-44) and the C-fragment (45-103). Although each of them was unstructured alone in water at pH 7.0 and 25° C, one-to-one fragment-complex was formed upon the mixture of N- and C-fragments. CD and NMR spectra of the fragment-complex indicated that its tertiary structure was quite similar to the native one of intact protein (24). (d) The ion-spray mass spectrum of a mixture of equimolar amounts of N- and C-fragments measured at pH 4.0. Symbols of N, C and NC represent dissociated N- and C-fragments and the fragment-complex, respectively. The obtained masses for N, C and NC are 5,371.0, 6,679.0 and 12,050.0. The existence of NC complex suggests that the fragment-complex is stable even in a vacuum.

all-or-none transition of the fragment-complex formation originates in the cooperativity

between intra-domain and inter-domain interactions.

Suppose that a tertiary structure was unfolded into two building blocks, they may be further divided into smaller building blocks by the hinge motion. For example, suppose that several residues from the N-terminus or the C-terminus are movable freely in the native structure of lysozyme, a limited number of potential flexible hinges are found around residues 28, 40, 82 and 108 (23). Such building blocks of lysozyme are shown schematically in Figure 4. Consequently, the protein structure seems the close packing of small building blocks which can unfold without severe obstacle due to the steric overlap. In such a sense, protein folding is able to be compared to assembling a three-dimensional jigsaw puzzle such as a wooden mosaic work.

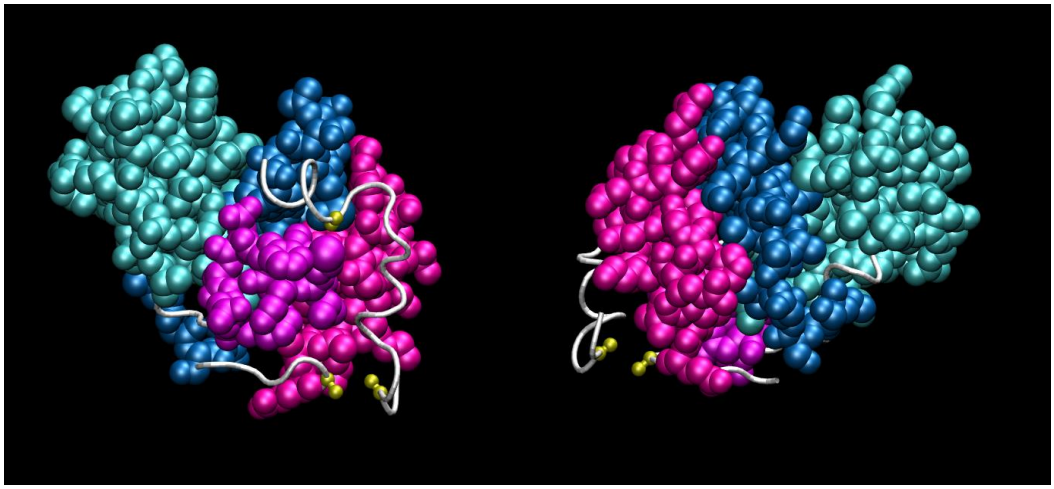

Figure 4. Schematic representation of hen lysozyme composed from four building blocks dissected at potential flexible hinges. Residues 7-28, 29-38, 42-79, 84-108 are represented by spheres with a van der Waals radius of each atom, while residues 1-6, 39-41, 80-83 and 109-129 are represented by tubes tracing respective C_α atoms. Building blocks are distinguished by different colors: the block 7-28 is colored reddish-purple, the block 29-38 purple, the block 42-79 light blue and the block 84-108 dark blue. Left: the view from the same direction as in the ribbon diagram of Figure 1. Right: the back view of the left.

4. An initial combination of building blocks is important to the selection of folding pathways

Probably, the initial assembly of a few building blocks may be very important in a successful folding of proteins that undergo the two-state transition. If it is wrong, it may cause severe obstacle due to the steric overlap in merging building blocks; we cannot merge remaining building blocks with the former structure, as everyone experiences it in a wooden mosaic work. It results in breaking the former structure due to the strong chemical-affinity toward the unfolded state. The initial combination of building blocks must limit following folding events, and the folding pathway must be strongly restricted

by the initial selection of building blocks.

Recently, the concept of relative contact order (CO) was proposed to quantify the topology of protein native structure (25). Contact orders measure the average sequence separation between contacting residues in the native structure. When the native structure is constructed of some compact substructures, the CO is relatively small. In the case of a packed structure without any compact substructure, the portion of long-range interactions becomes larger and the CO results in large. For example, the CO is small in a cylindrical helix, intermediate in a four helix bundle, and large in a compact structure without any secondary structure. Observations on relatively small and single domain proteins demonstrated that the CO well correlated with the logarithm of folding rates; a larger CO indicates a longer folding time. Also a weak correlation was found between the folding transition state placement and the CO; the folding transition state of a protein with a larger CO is placed nearer to the native state (25). In the case of proteins with a large CO, therefore, a large-size assembly of building blocks is required to get over the transition state of folding. Probably it may take longer time to find a proper order upon assembling larger number of building blocks, so that the folding rate significantly slows down. Further, it is likely that the initial combination of building blocks is more limited to complete the folding transition. Now, our interest is in a correlation between folding pathways and the CO of protein.

5. A correlation between the CO and the multiplicity of folding pathways

In 1986, we published a paper (10) entitled “Computer simulation of the folding-unfolding transition of island-model proteins”. Although the details of our studies will be described in Chapter IV, structural fluctuations, the transition state placement, and the folding pathway were examined in connection with contact orders of several types of model proteins. Results obtained there were consistent with those suggested by the present concept of CO, although the concept of CO had been unknown yet at that time. Here we will focus our attention on the correlation between the CO of proteins and the multiplicity of folding pathways. As described later, computer simulations exemplified folding pathways of various types of model proteins. A general tendency suggested by them is as follows; folding pathways are more limited for proteins with a larger CO. For example, a model protein with a CO of 35% (named R-1 in the original paper) exhibited virtually a single major folding pathway, while a model protein with a CO of 19% (named α/β) showed multiple folding pathways. On the other hand, a cylindrical helix with a CO of 7% exhibited no definite folding pathway.

III. TRANSITION STATE THEORY AND A STEADY STATE APPROXIMATION

In most cases, two-state kinetics is applicable to protein folding or unfolding. Actually, folding or unfolding reaction can be well expressed by a single exponential function of time in the case where any intermediate product does not accumulate as much as detectable during the reaction. However, this does not mean that there is no intermediate on the reaction pathway. Protein folding or unfolding generally occurs through vast numbers of elementary reaction steps. Two-state kinetics means that intermediate species on multiple reaction steps are in a steady state. For example, let us consider the following reaction scheme.

$$A_1 \underset{k^-_1}{\overset{k^+_1}{\rightleftarrows}} A_2 \underset{k^-_2}{\overset{k^+_2}{\rightleftarrows}} A_3 \rightleftarrows \cdots\cdots\cdots \underset{k^-_{n-1}}{\overset{k^+_{n-1}}{\rightleftarrows}} A_n \underset{k^-_0}{\overset{k^+_0}{\rightleftarrows}} B_m \underset{r^-_{m-1}}{\overset{r^+_{m-1}}{\rightleftarrows}} \cdots\cdots\cdots\cdots \underset{r^-_2}{\overset{r^+_2}{\rightleftarrows}} B_2 \underset{r^-_1}{\overset{r^+_1}{\rightleftarrows}} B_1$$

Suppose that intermediate species, A_i ($1<i\leq n$) and B_j ($1<j\leq m$) are in a steady state, let us consider the net rate constant, k_p, from A_1 to B_1, and the reverse rate, k_r, from B_1 to A_1. Provided that the rate constant k^+_i of the elementary step from A_i to A_{i+1} is much slower than k^-_i of the reverse step, and that the rate constant r^+_j of the elementary step from B_{j+1} to B_j is much faster than r^-_j, the states, A_n and B_m, are located at the top of Gibbs free-energy profile. Since the populations of molecules in intermediate states from A_2 through B_2 are very small, it is appropriate to postulate that time-derivatives of intermediate populations, $d[A_i]/dt$ and $d[B_j]/dt$, are equal to zero, that is to say, a steady state approximation is good. So that we can shorten the above reaction scheme as follows.

$$A_1 \underset{k^-_p}{\overset{k^+_p}{\rightleftarrows}} A_n \underset{k^-_0}{\overset{k^+_0}{\rightleftarrows}} B_m \underset{k^-_r}{\overset{k^+_r}{\rightleftarrows}} B_1$$

Then, the apparent rate constants, k^+_p and k^-_p, between A_1 and A_n are expressed by

$$k^+_p = k^+_{n-1}\cdot K_1\cdot K_2\cdots K_{n-2}, \qquad k^-_p = k^-_{n-1},$$

where K_i is the equilibrium constant between A_i and A_{i+1}; $K_i = k^+_i / k^-_i$.

In the same manner, k^+_r and k^-_r are written by

$$k^+_r = r^+_{m-1}, \qquad k^-_r = r^-_{m-1}\cdot K'_1\cdot K'_2\cdots K'_{m-2}, \qquad \text{where } K'_j = r^-_j / r^+_j.$$

Since A_n and B_m are also in a steady state, finally the rate constant k_p from A_1 to B_1, and k_r of the reverse step are given by

$$k_p = k^+_0\cdot K_1\cdot K_2\cdots K_{n-1} = k^+_0\cdot K_{A_n\text{-}A_1}, \qquad k_r = k^-_0\cdot K'_1\cdot K'_2\cdots K'_{m-1} = k^-_0\cdot K_{B_m\text{-}B_1},$$

where $K_{An\text{-}A1}$ and $K_{Bm\text{-}B1}$ are the equilibrium constants between A_n and A_1 and between B_m and B_1, respectively. The expressions obtained for k_p and k_r are similar to those obtained from the conventional transition state theory applied to a single elementary step: $A_1 \rightleftarrows B_1$. Although it seems inappropriate to apply the transition state theory to the protein folding composed of many elementary steps, the theory is also applicable to reaction rates with multiple elementary steps, if the steady state approximation is good.

Suppose that k_p and k_r obey the absolute reaction rate theory, states A_n and B_m correspond to a transition state (T) hypothetically equilibrated with A_1 and B_1, respectively. So that $K_{An\text{-}A1}$ or $K_{Bm\text{-}B1}$ corresponds to the hypothetical equilibrium constant of the state T with the state A_1 or B_1. Further, k^+_0 and k^-_0 are equal to k_BT/h in the absolute reaction rate theory, where k_B, h, and T are the Boltzmann constant, the Plank constant, and the absolute temperature, respectively. Although the so-called transition state is only a hypothetical state between A_1 and B_1, A_n or B_m is an intermediate state which actually exists at the top of Gibbs free-energy profile along the reaction pathway under a steady state approximation, and $K_{An\text{-}A1}$ or $K_{Bm\text{-}B1}$ is an actual equilibrium constant. When A_1 and B_1 correspond to U and N, respectively, k_p is the rate constant of folding (k_f) and k_r is that of unfolding (k_{uf}).

Activation energies of folding and unfolding, $E^{\ddagger}_f$ and $E^{\ddagger}_{uf}$, are obtained from Arrhenius plots of k_f and k_{uf}, respectively. On the other hand, in the steady state approximation, $E^{\ddagger}_f$ or $E^{\ddagger}_{uf}$ is represented by

$$E^{\ddagger}_f = -\partial \ln (k^+_0 \cdot K_{An\text{-}A1}) / \partial T^{-1} = \varepsilon^+ + \Delta H_f, \quad E^{\ddagger}_{uf} = -\partial \ln (k^-_0 \cdot K_{Bm\text{-}B1}) / \partial T^{-1} = \varepsilon^- + \Delta H_{uf},$$

where ε^+ or ε^- is the activation energy associated with k^+_0 or k^-_0, and ΔH_f or ΔH_{uf} is the change in enthalpy associated with the equilibrium constant $K_{An\text{-}A1}$ or $K_{Bm\text{-}B1}$. If we postulate that ε^+ or ε^- is negligible compared with ΔH_f or ΔH_{uf}, the activation energy is virtually equal to the change in enthalpy from the initial state to the state A_n or B_m. If we call both A_n and B_m together the transition state T, $E^{\ddagger}_f$ and $E^{\ddagger}_{uf}$ are virtually equal to the change in enthalpy from U to T and from N to T, respectively. From this point of view, we call $E^{\ddagger}_f$ and $E^{\ddagger}_{uf}$ activation enthalpies of folding ($\Delta H^{\ddagger}_f$) and unfolding ($\Delta H^{\ddagger}_{uf}$), respectively.

So far, the transition state theory has been examined according to the reaction scheme composed of vast numbers of elementary steps on the assumption that a steady state approximation is true. In next Chapter, we demonstrate that folding and unfolding rate constants of model proteins determined directly by Monte Carlo simulations really obey the transition state theory, and that the transition state certainly lies at the top of the free-energy profile.

IV. COMPUTER SIMULATION OF ISLAND-MODEL PROTEINS

1. Three-dimensional structures of model proteins

In our computer simulation study of island-model proteins published in 1986 (10), we attempted to exemplify structural fluctuations in the native and unfolded states, folding and unfolding rate constants, the transition state, and the folding pathway. In computer simulations performed by Eigen, the goal of his study was to create a meaningful message from a random sequence of letters (8). According to his manner, we also tried to represent a conformation of protein by a sequence of binary numbers. We postulated that a model protein was a single chain composed of 64 units (residues). Each unit takes eight different substates from 0 to 7 as shown in Figure 6 (a), where 0 is the native state of the unit and states 1 to 7 are non-native states. That is to say, a sequence of 64 octal digits represents a conformation of protein. When all units are in the 0-state, the chain has the native conformation. The three-dimensional structure of a model protein in the native state is represented by a cube composed of 4 x 4 x 4 lattice points (26). Every unit of the chain is located somewhere on the lattice without overlapping. Characterization of the native backbone-structure depends on how to join these lattice points in a single chain. Here, we take up three types of conformation: (a) a cylindrical helix of 2 x 2 x 16 lattice points, (b) an α/β protein composed of two α-helices and two β-sheets and (c) a R-1 protein with an irregular chain in whole, which are shown in Figure 5 (a), (b) and (c), respectively.

2. Specific interactions among neighboring units and the island-model approximation

Specific long-range interactions are required to stabilize the three-dimensional structure of a model protein, that is to say, two units which occupy nearest-neighbor lattice points in the native structure but are not connected by a bond interact specifically with each other to decrease the energy of the system by ε_ℓ (long-range). As a result, both α/β and R-1 proteins have 81 specific long-range interactions among neighboring units, and a cylindrical helix has 61. These specific interactions are well represented by spots on the triangular contact map as shown in the bottom of Figure 5 (a), (b) and (c); the spot on the intersection of the *i*-th row and the *j*-th column represents the specific interaction between the *i*-th and *j*-th residues ($i \geq j$). This triangular contact map symbolizes the feature of the three-dimensional structure of protein in the native state in terms of the distribution pattern of spots. Spots running parallel or perpendicular to the diagonal line

symbolize an α-helix or an antiparallel β-sheet. A triangular contact map rich in spots far away from the diagonal line represents a protein rich in long-range interactions. Since

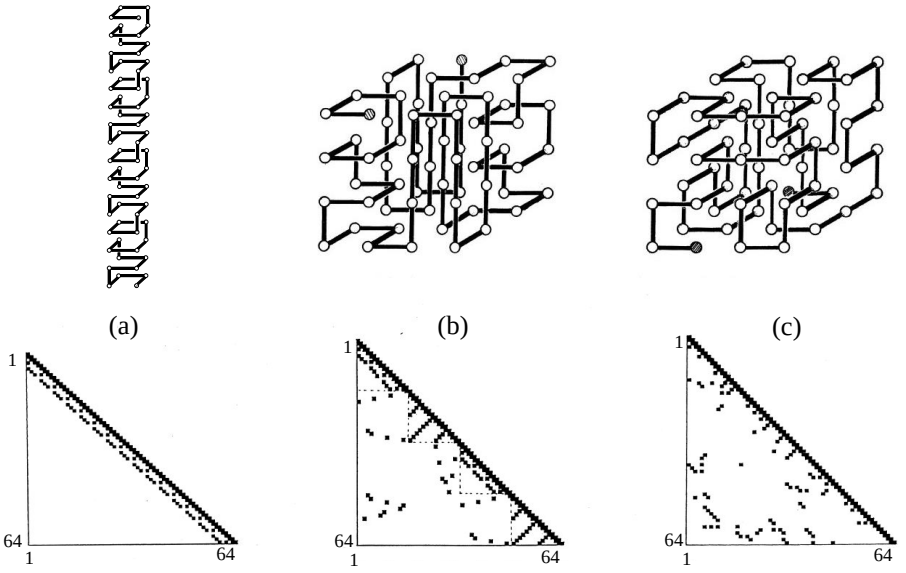

Figure 5. Three-dimensional structures of island-model proteins; (a) cylindrical helix, (b) α/β protein, (c) R-1 protein. All of them are constructed of 64 units (residues) connected by a bond in sequential order. The cylindrical helix is composed of 2 x 2 x 16 lattice points. The α/β protein is composed of two α-helices and two β-sheets, while the R-1 protein has no compact substructure. The diagrams below are triangular contact maps; long-range and short-range interactions between two neighboring units on the lattice points are represented by spots on intersections between them. Besides them, spots on the diagonal line of the contact map represent the intrinsic stabilization energy of the unit. The arrangements of spots typical of the α-helix or the anti-parallel β-sheet are observed in the contact map of the α/β protein. In the case of cylindrical helix, the total number of spots is 188 (details: long-range interactions 61, short-range ones 63, intrinsic ones 64). On the other hand, the total number of spots in each contact map of α/β and R-1 proteins is 208 (details: long-range interactions 81, short-range ones 63, intrinsic ones 64). Therefore, the total stabilization energy of cylindrical helix is 188 ε_0 and that of the α/β or the R-1 protein is 208 ε_0.

contact order (CO) measures the average sequence separation between contacting residues in the native state, the CO of the protein rich in spots far away from the diagonal line is large. For example, the R-1 protein has a CO of 35%, and the α/β protein a CO of 19%, while the CO of the cylindrical helix is only 7%.

When the conformation of the unfolded protein is represented in terms of a sequence of octal digits, several islands of the digit 0 exist intervening in the sequence, as shown in Figure 6 (b). These islands correspond to native local conformations intervening in non-native ones. The island-model approximation (27) means that the specific interaction between a pair of units is disregarded, if the pair is separated by a sequence of non-native units. That is to say, the specific interactions shown by spots on

the triangular contact map are available only when they are involved in the same island. In addition, we postulated that there were two other types of interactions. In the first, the energy of the system decreases by ε_i (intrinsic) when each unit is in the 0-state, and in the second it decreases by ε_s (short-range) when both units adjoining each other through a bond are in the 0-state. In our original paper, we assigned the same value ε_0 of -1024 to all of ε_i, ε_s, and ε_ℓ.

3. Monte Carlo simulation

Eight substates of a unit are linked in a loop and the transition to the two neighbors is allowed. For example, 4-state can convert to 3- or 5-state, and 0-state to 1- or 7-state, as shown in Figure 6 (a). We denote the transition probability from the native state

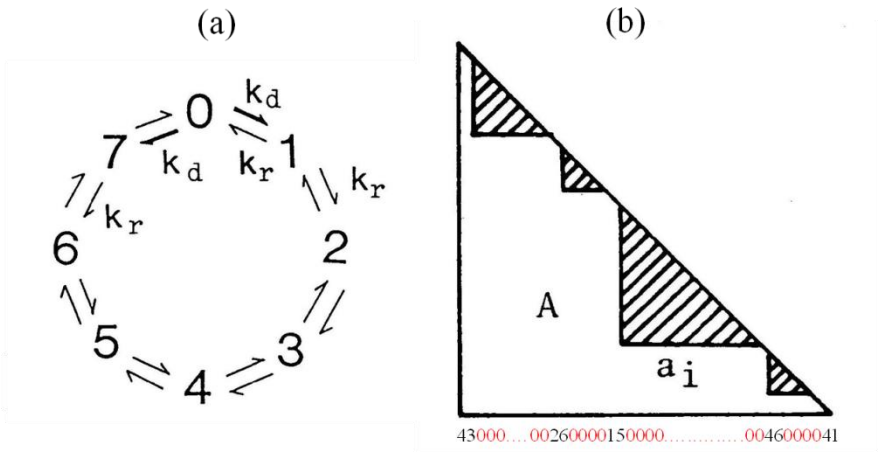

Figure 6. Island-model approximation of specific long-range interactions. When all the units on the lattice have the native conformation, the protein has the native three-dimensional structure illustrated in Figure 5. However, each unit changes its conformation from the native one to another. We postulated that each unit had eight different substates as shown in the diagram (a). The 0-state denotes the native one, and others the non-native states. Arrows represent the reaction steps between substates, and k_d or k_r denotes the transition probability described in the text. When some units are in the non-native state, protein conformations are represented by a sequence of octal digits shown below the triangle (b), in which several islands of the digit 0 are created with being interrupted by non-native digits. For example, four small islands are observed in the triangle (b). Specific long-range interactions between two isolated islands are not available any longer, because the pair of units cannot come into contact on neighboring lattice points. Therefore, stabilization energy of the conformation is determined by counting the number of spots involved in triangles indicated by hatched lines.

(0-state) to non-native states (1- or 7-state) by k_d and that of the reverse process by k_r. Also, the probability of transition among non-native states is k_r. In order to satisfy the canonical distribution, we assume k_d and k_r as follows:

$$k_d = 1/2 \cdot \exp(-\Delta E/\theta), \qquad k_r = 1/2.$$

ΔE is the increase in the energy of a system associated with the transition of a unit from the 0-state to the 1- or 7-state; in $\theta = k_B T$, k_B and T are the Boltzmann constant and the absolute temperature, respectively. The Monte Carlo simulation was performed on a microcomputer for exclusive use.

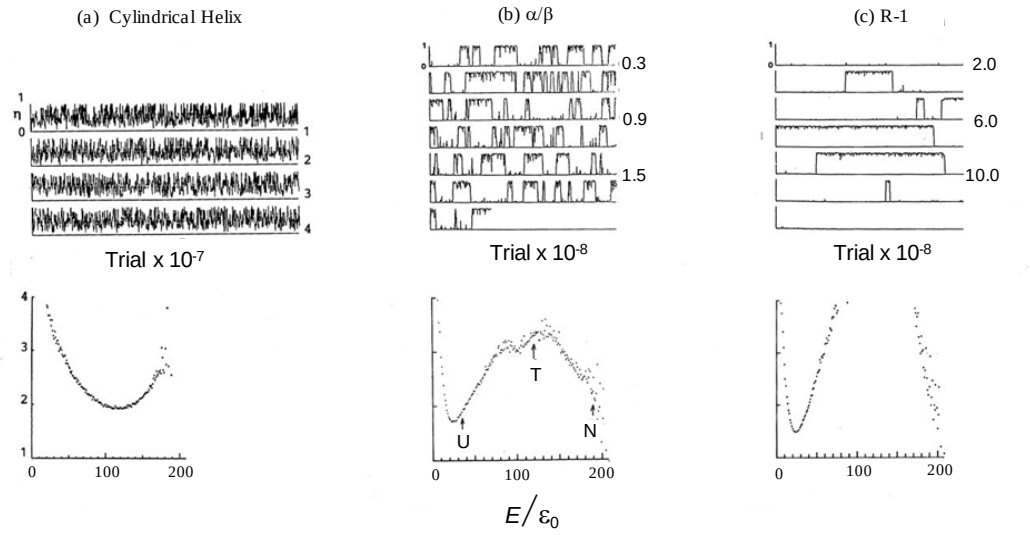

Figure 7. Structural fluctuations of island-model proteins near at the midpoint of folding-unfolding transition. The ordinate is order parameters (η) of protein conformations, and the abscissa trial numbers of Monte Carlo simulations. The order parameter was calculated from the sum of areas of small shaded triangles shown in Figure 6 (b) as follows, $\eta = \Sigma_i a_i / A$, where a_i and A were areas of the i-th small triangle and the whole triangle, respectively. (a) Cylindrical helix at $\theta = 1230$. (b) α/β protein at $\theta = 1480$. (c) R-1 protein at $\theta = 1480$. Total times of trials in Monte Carlo simulations were (a) 4 x 10^7, (b) about 1.9 x 10^8 and (c) about 1.3 x 10^9, respectively. The diagrams below are free-energy profiles calculated from the results of Monte Carlo simulations. The ordinate is proportional to the free energy of protein with the energy E, and given by $-\log_{10} [q(E)/Q]$, where $q(E)$ is the frequency of occurrence of states with the energy from E to $E+\delta E$, and $Q = \Sigma_E q(E)$. The abscissa is the energy of the protein in a unit of ε_0. In the diagram (a), a single free-energy minimum is observed near at the helix content of about 50%. In the diagram (b), the symbols U and N denote the placements of $<E(U)>$ and $<E(N)>$, respectively, while the symbol T denotes the transition state placement determined from the Arrhenius plots of k_{uf} and k_f.

Although the number of all possible conformations of a single chain is $8^{64} \approx 6 \times 10^{57}$, the total number of trials in Monte Carlo simulations was about 1 x 10^9. In the equilibrium state at the midpoint of transition, a model protein repeats the unfolding and folding transitions many times. The above chart of Figure 7 (b) shows a record of the simulation for α/β proteins at $\theta = 1480 = 1.45 \cdot |\varepsilon_0|$, which is close to the temperature at the midpoint of transition. A simulation starts from the fully unfolded state in which all digits are 4. The ordinate represents the order parameter (η) of the protein conformation. When $\eta = 1$, a model protein has the fully folded native conformation. A protein molecule almost always stays in either the unfolded (U) or the folded state (N), while the transition

between them is completed in a moment. Dynamic aspects of the unfolding-folding transition of R-1 proteins are remarkably different from those of α/β; the structural fluctuations in U and N states of R-1 proteins are quite suppressed and the lifetimes of U and N states are both much longer than those for α/β proteins.

On the other hand, a cylindrical helix frequently changes its conformation compared with α/β and R-1 proteins. The differences in dynamic aspects are well represented by free-energy profiles, which were calculated from the frequency of occurrence of states with energy E. The bottoms of Figure 7 (a), (b), and (c) show the free-energy profiles calculated for a cylindrical helix, α/β, and R-1 proteins, respectively. In the cases of α/β and R-1 proteins, there are two free-energy minima at both ends of abscissa (the reaction coordinate), and one free-energy maximum between two minima. In other words, a free-energy barrier exists between N and U states, and the ratio of the population of N state to that of U state varies with the change in temperature. This is a two-state transition typical of the protein folding-unfolding. The free-energy barrier of R-1 proteins is much higher than that of α/β. On the contrary, in the case of the cylindrical helix, there exists a single free-energy minimum near the middle of the reaction coordinate. Its position moves toward more disordered states with the increase in temperature. This is typical of a helix-coil transition and called a continuous transition. The mean value of energy E in the vicinity of N or U was calculated for α/β proteins; $\langle E_N \rangle = 191\varepsilon_0$, $\langle E_U \rangle = 35\varepsilon_0$ at $\theta = 1480 = 1.45 \cdot |\varepsilon_0|$. Therefore, the unfolding enthalpy (ΔH_{uf}) is equal to $156 \cdot |\varepsilon_0|$, where we use the term of the change in enthalpy (ΔH) without any distinction from the change in energy (ΔE).

4. Folding and unfolding are typical Poisson processes

In Figure 7 (b), the order parameter abruptly changes from 0-neighborhood to 1-neighborhood, or in the opposite direction. So that the transition of folding or unfolding itself happens in a short time, but the time span of the U state or the N state varies every try. To take an average, simulations of the folding of α/β proteins were independently repeated 16 times, each time starting from the fully unfolded state. In Figure 8 (a), each chart below shows the trajectory of a single simulation. As a result, a relaxation curve of the order parameter (η) was obtained by totaling them. It was very close to a single exponential curve as shown in the above chart of Figure 8 (a). The same result was also obtained for the unfolding transition. This demonstrates that the folding and unfolding transitions are typical Poisson processes. That is to say, the folding or unfolding transition is a rare event, and its average rate of occurrence is $1/\tau$, where τ is the relaxation time of a single exponential curve. Thus, the rate constants of folding and unfolding, k_f and k_{uf},

were well determined from the relaxation times. On the other hand, folding and unfolding transitions are repeated many times near at the midpoint of transition, therefore, k_f or k_{uf} was determined from the reciprocal of the mean lifetime of the U or the N state.

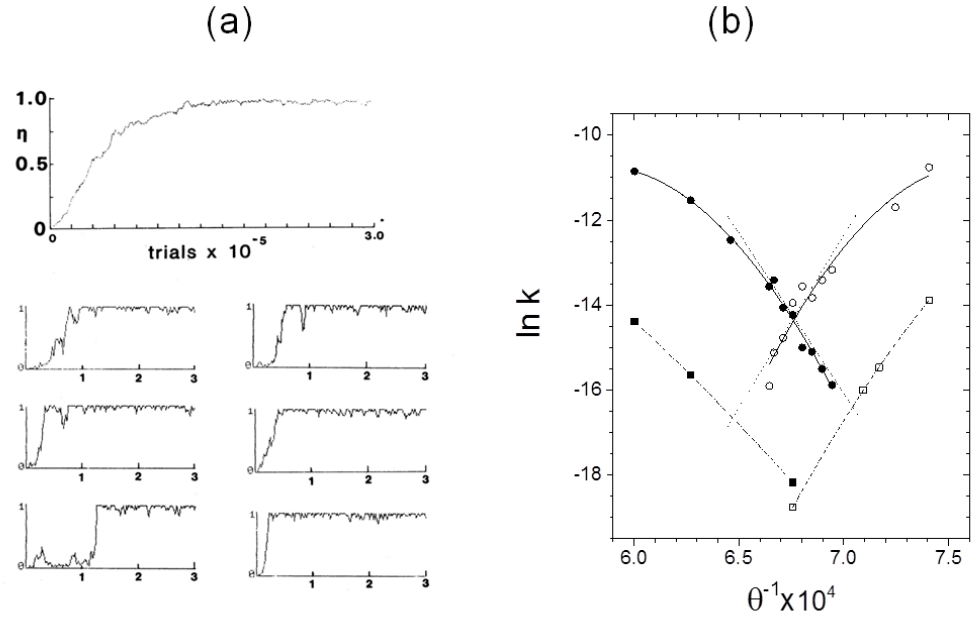

Figure 8. A single folding transition observed in a single protein molecule. Monte Carlo simulations of α/β protein were carried out at $\theta = 1350$ from the beginning conformation of 444…..44. After a while, the protein reaches the native conformation and the order parameter fluctuates around 1-neighborhood. Such simulations were repeated independently 16 times by using different sequences of random numbers. Some results are shown in the bottom of the diagram (a). The folding transition abruptly occurs and the duration staying in the U-state varies every try. A single exponential decay curve was obtained by totaling them as shown in the above chart. It means that the folding transitions are typical Poisson processes. The average rate of occurrence is equal to the reciprocal of the relaxation time of the exponential curve. The relaxation time was 4.9 x 10^4 trials in the diagram (a). Thus, rate constants of k_f and k_{uf} were determined for the α/β and R-1 proteins at various temperatures. The diagram (b) shows their Arrhenius plots. Filled and open symbols represent k_{uf} and k_f, respectively. (○) α/β protein, (□) R-1 protein. The slopes of the tangential lines at the midpoint of the transition give $\Delta H^{\ddagger}_{uf}$ and $\Delta H^{\ddagger}_{f}$ of the α/β protein at $\theta = 1480$: $\Delta H^{\ddagger}_{uf} = 70|\varepsilon_0|$ and $\Delta H^{\ddagger}_{f} = -85|\varepsilon_0|$. These values indicates that the transition state placement of the α/β protein is close to $E = 120\ \varepsilon_0$. On the other hand, $\Delta H^{\ddagger}_{uf}$ of the R-1 protein was nearly equal to $50|\varepsilon_0|$ at $\theta = 1480$.

5. Arrhenius plots of folding and unfolding rates and the consistency with the transition state theory

Figure 8 (b) shows the Arrhenius plots of folding and unfolding rates for α/β and R-1 proteins. Provided that folding and unfolding rate constants are according to the transition state theory, the activation energies are calculated from the slopes:

$$E^{\ddagger}_{uf} = -\partial \ln k_{uf} / \partial \theta^{-1}, \quad E^{\ddagger}_{f} = -\partial \ln k_f / \partial \theta^{-1}.$$

In this case, $E^{\ddagger}_{uf}$ and $E^{\ddagger}_{f}$ are expressed in terms of the change in energy per a single molecule. In the absolute reaction rate theory, a rate constant k is expressed by the product of the hypothetical equilibrium constant of the transition state T with an initial state, $K_{T\text{-}ini}$, and the factor of $k_B T/h$. Hence,

$$E^{\ddagger} = \Delta H_{T\text{-}ini} + k_B T,$$

where $\Delta H_{T\text{-}ini}$ is also expressed in terms of enthalpy change per a single molecule. Since $k_B T$ is generally much smaller than $\Delta H_{T\text{-}ini}$, $E^{\ddagger}$ is virtually identical with the first term. Therefore,

$$E^{\ddagger}_{uf} = \Delta H_{Tuf\text{-}N} = \Delta H^{\ddagger}_{uf}, \quad E^{\ddagger}_{f} = \Delta H_{Tf\text{-}U} = \Delta H^{\ddagger}_{f},$$

where T_{uf} and T_f denote the hypothetical transition state of unfolding and folding, respectively. For the α/β protein, it turns out from the Arrhenius plot that

$$\Delta H^{\ddagger}_{uf} = 70|\varepsilon_0| \text{ and } \Delta H^{\ddagger}_{f} = -85|\varepsilon_0|.$$

Suppose that $H(T)$ denotes the enthalpy in the transition state T, then

$$\Delta H^{\ddagger}_{uf} = H(T_{uf}) - H(N), \quad \Delta H^{\ddagger}_{f} = H(T_f) - H(U).$$

T_{uf} and T_f are not necessarily identical with each other, because folding and unfolding rates are independently determined respectively in Monte Carlo simulations. From our computer simulations, following results were obtained;

$$\Delta H^{\ddagger}_{uf} - \Delta H^{\ddagger}_{f} = 155|\varepsilon_0|, \quad \langle E_U \rangle - \langle E_N \rangle = 156|\varepsilon_0| = H(U) - H(N).$$

Hence,

$$155|\varepsilon_0| = \Delta H^{\ddagger}_{uf} - \Delta H^{\ddagger}_{f} = H(U) - H(N) + H(T_{uf}) - H(T_f) = 156|\varepsilon_0| + H(T_{uf}) - H(T_f).$$

The above relation indicates that $|H(T_{uf}) - H(T_f)|$ is nearly equal to zero, so that both states, T_f and T_{uf}, together may be denoted by T. Thus, Monte Carlo simulations of island-model proteins demonstrated that there existed a single transition state between N and U states according to the transition state theory. In the case of the α/β protein, the transition state placement is near at $E=120\varepsilon_0$ on the reaction coordinate, judging from the value of $\Delta H^{\ddagger}_{uf}$ or $\Delta H^{\ddagger}_{f}$. Certainly, the bottom of Figure 7 (b) shows that the state with the highest free energy is located near at this point. These results are entirely consistent with that the transition state corresponds to a single turning point between N and U with the highest free energy on the reaction scheme. Thus, Monte Carlo simulations verified that the steady state approximation was applicable to the protein folding-unfolding reaction composed of multiple elementary steps, and the transition state theory was valid to analyze the two-state kinetics, as examined in Chapter III. In addition, Figure 8 (b) shows that the $\Delta H^{\ddagger}_{uf}$ of the R-1 protein is nearly equal to $50|\varepsilon_0|$. The value is smaller than that of the α/β protein, and it implies that the transition state placement of the R-1 protein is nearer to the N state than that of the α/β protein.

6. Recording folding pathways of island-model proteins

In terms of statistical mechanics, the most probable conformation of island-model proteins can be determined in an equilibrium state at any temperature. This method is suitable for analyzing the helix-coil transition (27), but not suitable for the protein folding-unfolding transition, because it is the all-or-none transition between N and U. Since a single protein molecule is almost always in either the N or the U state, however, the conformation in an equilibrium state is only the mean of various conformations that exist in the N and U states. Therefore, it is difficult to follow only the conformations which occur at the moment of the transition from U to N, or from N to U. Further, we cannot reveal the time sequence of formation of several local structures by means of statistical mechanics. However, computer simulation makes it possible to follow all the events occurring at the moment of folding transition. What kinds of local structures are formed? What is the time sequence of those events? However, the folding transition happens many times and events in a single molecule vary every try. We attempted to determine statistically the folding pathway from 255 trials of folding transition. We divided 64 units on a single chain into 5 blocks composed of 11 units, and named them A(5-15), B(16-26), C(27-37), D(38-48) and E(49-59), respectively; the beginning 4 units and the last 5 units were not involved in these blocks. Suppose every unit within each block is in the 0-state (native state), we judge that the block is formed, and if any one unit within the block is not in the 0-state, we judge the block is broken. For example, when the blocks A, C and D are formed and the blocks B and E are broken, the conformation of the chain is denoted by A_CD. Once all blocks are broken, the past record of folding events is lost. Only when the folding transition is completed leading to ABCDE, is the final time sequence of the formation of the blocks recorded. In the same manner, unfolding pathways are also recorded.

7. A major folding and unfolding pathways

The results are shown in Figure 9. The thickness of the line is proportional to the probability that the step is selected upon the folding transition. In computer simulations of a cylindrical helix, it is possible to form a short helix anywhere on the chain composed of blocks A through E, and further to continue the helix formation at any sites intervening in the non-native sequence, with compensating the entropy loss that arises from the helix formation. For example, split-type intermediates such as B_D or AB_D appear in the course of the folding transition. Also, the disruption of a cylindrical helix happens anywhere on the chain. As a result, short helices are floating on the chain, and the population of chains with the average helix content becomes predominant near at the

midpoint of transition. Thus, the cylindrical helix undergoes a continuous transition from a coil to a helix, and there exists no transition state. The cylindrical helix is able to be

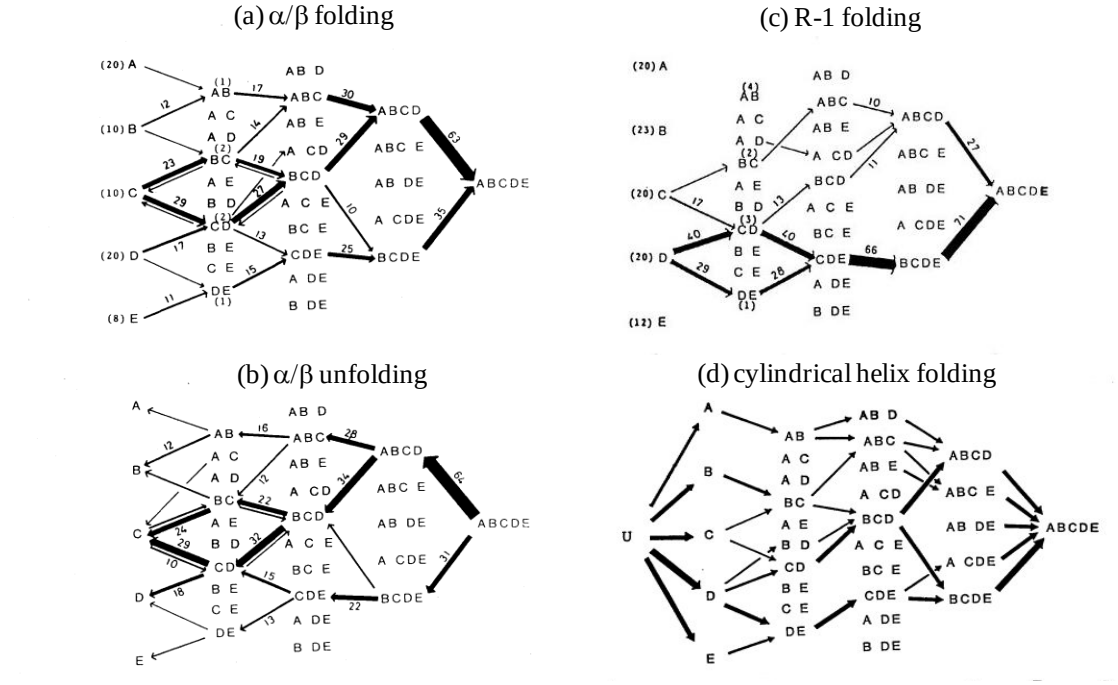

Figure 9. Folding and unfolding pathways of island-model proteins. Note that only when the folding transition is completed is the final pathway recorded. (a) Folding pathways of the α/β protein at $\theta = 1480$. The number on the arrow represents the percentage of the reaction step selected upon the folding transition. The thickness of the arrow is roughly proportional to the percentage. The arrows less than 5 % are not shown in the diagram. For example, 23 % on the arrow from C to BC indicates that the path from C to BC was selected in the 23% out of 255 times folding transitions. (b) Unfolding pathways of the α/β protein at $\theta = 1480$. It is noteworthy that the percentage of the reaction step selected upon the unfolding is very close to that of the reverse step selected upon the folding. (c) Folding pathway of the R-1 protein at $\theta = 1350$. This diagram indicates that the 69% of folding transitions proceeded through the formation of the block D. (d) Helix formation of the cylindrical helix at $\theta = 1230$. The helix formation is able to be initiated anywhere on the chain. Also, split-type intermediates such as ABC_E, AB_DE, A_CDE, AB_E and AB_D appear in the course of the folding transition. So far, the formation of intermediates was analyzed only when the folding transition was completed, however, protein molecules spend long time in the U-state before the folding transition happens. In general, any block from A through E is generated many times, and also joined blocks are generated occasionally in the U state. In diagram (a) and (c), the numbers in the parentheses represent the average number of times each block is generated in the U state before a folding transition. For example, each block was generated with a nearly equal probability in the R-1 protein, but the intermediate A, B or E did not proceed to the native state anymore. Free energies of various intermediate states generally increase abruptly as the intermediate come close to the native state. The formation of intermediates with a small gain in stabilization energy is accompanied with a large increase in free energy. It is likely that the folding reaction proceeds through a valley in the free-energy landscape.

formed through a large variety of routes, that is to say, there exists no major folding pathway (see Figure 9 (d)).

On the other hand, the folding or unfolding of α/β and R-1 proteins is the two-state transition between N and U, and the population of intermediate states is extremely small. When proteins approach the transition state from the U state or the N state, strong chemical-affinity toward the former state works. In the U state of these proteins, it is likely that small islands of digit 0 are generated everywhere on the chain with interrupted by non-native states, but they are not persistent and are floating throughout the chain. Even though a few islands were merged with each other, they almost always collapse and return to the former state because of too much increase in free energy. However, the folding transition is completed on rare occasions. In Figure 9, only the paths selected at the moment of folding transition are recorded. At that time, it is noteworthy that proteins do not repeat going backwards and forwards. It is because the birth and death of the same block is not left on record, although it is repeated frequently. This observation means that a partial unfolding has a strong tendency to return to a just one prior intermediate, so that the folding pathway is narrowed down and seems to go straight. Consequently, all the blocks are successively formed, and the folding transition seems to finish in a moment, as if the folding transition were products of chance. However, the sequential order of folding events is definitely determined. The structure formed by merging a few blocks must be so compact as to make the number of spots on the triangular contact map the most. Since free energy of proteins abruptly increases with the progress of folding, the formation of joined blocks must gain enough stabilization energy to restrain the free energy from increasing too much, as the former structure grows into a larger one. Such a combination of blocks must be limited so as to follow a valley in the free-energy landscape. In α/β proteins, it seems ABC or BCD or CDE, and in R-1 proteins it may be strictly limited to CDE. As a result, there exists a major folding pathway in these proteins. It is C - (BC or CD) - BCD - ABCD - ABCDE upon the folding of α/β proteins at θ = 1480. But the folding transition occasionally happens through several other pathways. In R-1 proteins, the selection of a major folding pathway is much stricter: D - (CD or DE) - CDE - BCDE - ABCDE.

When the unfolding pathway of α/β proteins was analyzed at θ = 1480, and compared with the folding pathway at the same temperature, an important fact was found; the probability of the reaction step selected upon unfolding was entirely the same as that upon folding, as shown in Figure 9 (a) and (b). Although it was somewhat a surprise that the descent to the U state from the T state had been the same as the ascent to there, the reaction may proceed with ensuring the detailed balance of elementary steps even upon

descending the free energy slope. Results of Monte Carlo simulations exemplified that folding and unfolding transitions progressed through the same route, probably following a valley in the free-energy landscape, and that a single transition state existed as the turning point between N and U states.

8. Assembly and disassembly of a three-dimensional jigsaw puzzle

As mentioned in Chapter II, the helix-coil transition is like a two-dimensional jigsaw puzzle, while the protein folding is like a three-dimensional one. Let us consider the assembly or disassembly of a three-dimensional jigsaw puzzle. When we disassemble it, it is not necessarily easy to take any building block out of the three-dimensional structure. In general, there is a key block which is easy to be taken away from the complete form of the jigsaw puzzle. After the removal of the key block, the three-dimensional structure may easily collapse into pieces of small building blocks. Which part on the chain is the key block? The disruption of a tiny local structure frequently occurs but it is restored soon. On the other hand, the unfolding of a substantial area is accompanied with the loss of many long-range interactions, so that such a conformation change around some flexible hinge hardly happens, although the area is broken into small pieces, once it has happened. Probably the key block may be determined in an even balance between them. It is likely to be favorable that some substantial area is broken accompanying not so much increase in energy. Anyway, it is noteworthy that this key block corresponds to the last one assembled in the folding pathway. The exit of the folding pathway is also narrowed to a specific building block.

In the assembly of the three-dimensional jigsaw puzzle, if the initial combination of building blocks is inappropriate, the steric overlap with an additional building block prevents the former structure from growing into a larger compact one, and it results in the complete collapse of the former structure for retrying. The initial combination of building blocks tends to determine success or failure of building a three-dimensional jigsaw puzzle. Such a proper combination of building blocks may be determined so as to satisfy the cooperativity between intra-block and inter-block interactions, otherwise they cannot merge into a larger structure due to the steric overlap. Upon the folding transition, building blocks of various shapes are generated and collapsed repeatedly throughout the chain. When the combination of building blocks mentioned above is realized, a larger-size compact structure may be constructed by assembling them in a proper sequential order. In the case of three-dimensional jigsaw puzzle, both the entrance and the exit of the folding pathway are likely to be narrowed to some limited places.

Consequently, in a cylindrical helix compared to a two-dimensional jigsaw puzzle, the CO is small, the folding rate is very fast, there is no transition state, and there is no definite folding pathway. On the contrary, α/β and R-1 proteins compared to a three-dimensional jigsaw puzzle undergo a two-state transition and have some limited folding pathways. In the α/β protein, there exist some compact substructures: α-helices, β-sheets and the combined α/β structure. Therefore, its CO is not so large and the cooperativity between intra-block and inter-block interactions is relatively easy to be satisfied. It results in that the folding rate is not as slow as that of R-1 proteins, and a few other folding pathways exist besides a major one. However, the three-dimensional structure of R-1 proteins with a large CO is irregular in whole, and there exists no compact substructure. Therefore, short-range interactions alone must keep incompact substructures native-like until the long-range interactions start to work. It implies that more extensive and strong cooperativity is required among specific interactions, namely almost all of them working together. The initial combination of building blocks may be strictly limited to gain as many long-range interactions as possible, while the key block to start the unfolding may be also restricted to minimize the initial loss of long-range interactions. It means that both the entrance and the exit of folding pathway are narrowed. Also, the structure in the transition state must be constructed of a larger number of residues to get enough stabilization energy. As a result, it follows that the folding rate slows down extremely, the transition state placement is nearer to the native state, and a major folding pathway becomes more predominant over others. These characteristics of R-1 protein all originate in its large CO.

V. CHARACTERIZATION OF THE TRANSITION STATE

1. Analysis of folding and unfolding rates characterizes the transition state

The transition state (T) is a turning point between N and U with the highest free energy along the folding pathway. We cannot observe the transition state directly because of the extremely small population. It can be characterized only by evaluating the effect of either varying the conditions or modifying the protein on the rates of folding and unfolding. Since rate constants are determined by the activation free-energy from the U or the N state to the T state, analysis of folding or unfolding rates gives information about thermodynamic properties of the transition state. Especially, the difference in enthalpy from the U or the N state to the T state, $\Delta H^{\ddagger}_{f} = H(\mathrm{T}) - H(\mathrm{U})$, $\Delta H^{\ddagger}_{uf} = H(\mathrm{T}) - H(\mathrm{N})$, are determined from the Arrhenius plot of folding or unfolding rates under the steady state approximation described in Chapter III, where we postulated that the transition state of the folding was virtually identical with that of the unfolding, and that the activation energy $\varepsilon^{\pm}$ associated with a single elementary step passing the peak of free-energy profile was negligible.

In 1984, we published papers on kinetics of lysozyme folding and unfolding (16). Figure 10 shows Arrhenius plots of folding and unfolding rates measured in solutions containing various concentrations of GuHCl. The changes in enthalpy between T and N states, $\Delta H^{\ddagger}_{uf}$, were determined at various temperatures. As a result, $\Delta H^{\ddagger}_{uf}$ were found to be independent of temperature and concentrations of GuHCl, in spite of the fact that the change in enthalpy between U and N, ΔH_{uf}, significantly depended on them (28). These facts indicate that interactions between protein and different kinds of solvent molecules vary only slightly during the first stage of unfolding (from N to T), in other words, the transition state is as compact as the native one. Particularly, the fact that $\Delta H^{\ddagger}_{uf}$ remains constant (about 200 kJ/mol) in the wide temperature range indicates that the change in heat capacity is nearly equal to zero at the first stage of unfolding, according to the relation of $\Delta C^{\ddagger}_{uf} = \partial(\Delta H^{\ddagger}_{uf})/\partial T$. As a necessary consequence, $\Delta C^{\ddagger}_{f}$ is nearly equal to the change in heat capacity from U to N: - 6.7 kJ/mol (28). The temperature dependence of k_f calculated using this value is shown in Figure 10, and it entirely agrees with observed data. The result of $\Delta C^{\ddagger}_{uf} = 0$ presents an important feature of the transition state about the interactions with water.

Experimental facts that ΔC_p per unit mass of protein is proportional to the number of pairs of nonpolar contacts per unit mass of protein suggest that ΔC_p is attributable to structuring water around the nonpolar groups exposed (28). Therefore, it is clear that the second stage of unfolding (from T to U) is associated with hydration of internal groups

of the protein, while the first stage should correspond to disruption of the internal bonds maintaining a compact native structure of a protein. As proposed by Privalov in his review article (28), the transition state can be considered as a "dry molten globule" in protein, i.e., a state with diminished non-covalent interactions, but without water inside (29).

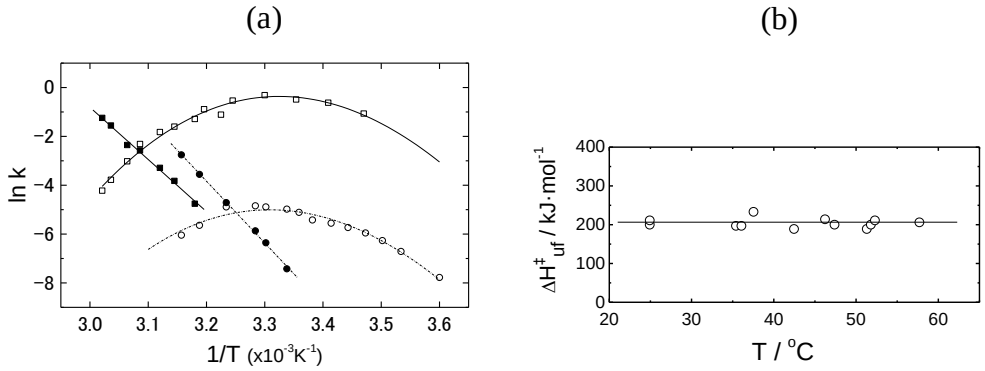

Figure 10. Folding and unfolding kinetics of wild-type lysozyme were measured at various concentrations of GuHCl, and at various temperatures. (a) Arrhenius plots of folding and unfolding rates. Only typical experimental results are shown for simplification. Rate constants were measured at pH 2.6 with a temperature-jump or a concentration-jump of GuHCl. (□) data in 0.7 M GuHCl, (○) data in 2.4 M GuHCl. Filled and open symbols denote k_{uf} and k_f, respectively. The best fitting curve for ln k_f vs. 1/T was calculated by using $\Delta C^{\ddagger}_f$ = -6.7 kJ·K^{-1}·mol^{-1}. (b) Temperature-dependence of $\Delta H^{\ddagger}_{uf}$ measured under various experimental conditions. These data indicate that $\Delta C^{\ddagger}_{uf}$ is equal to zero in the wide range of temperature.

2. Effects on folding and unfolding rates of the modification of specific residues

In the second paper (17) published in 1984, we also evaluated the effects of an introduction of a specific intra-chain cross-link to lysozyme. An extra cross-link between Glu35 and Trp108 was introduced to lysozyme by a chemical modification. The native structure of protein is little affected by the modification (30), and kinetic experiments of its folding and unfolding were performed. The cross-linking increased the folding rate greatly (by a factor of about 500), but decreased the unfolding rate only slightly (by a factor of about 1/10). It is likely that the cross-linking restricts possible conformations in the U state to some conformations. In the intact lysozyme without the extra cross-link, suppose that surroundings of Glu35 and Trp108 are disordered in the T state, conformations of their surroundings must be equally restricted by the introduction of the cross-link in both T and U states. It follows that the folding rate is accelerated little by the cross-linking. Therefore, the observation of folding rates indicates that Glu35 and Trp108 of intact lysozyme already come into contact with each other in the T state, so that the local structure surrounding them must have been affected little in the T state by the

introduction of the cross-link. On the other hand, the unfolding rate is affected only slightly, because both N and T states are equally influenced by the cross-linking.

In general, if the modification such as the introduction of an extra cross-link largely accelerates the folding rate and has no effect on the unfolding rate, it indicates that surroundings of the modification site are structured in the T state as well as in the N state. On the contrary, if the effect of the modification on the folding rate is little and the unfolding rate is largely diminished, surroundings of the modification site in the T state are as disordered as in the U state, so that the modification influences the T state as significantly as the U state. Later, the transition state in barnase folding was characterized in more detail by using protein engineering techniques instead of the chemical modification to alter contacting residues in the native state, and evaluating its effects on the rates of folding and unfolding (31). Now such techniques are generally formulated and called the φ-value analysis (32).

The effects of specific ligand binding on folding and unfolding rates were also examined in our paper mentioned above. N-acetyl-glucosamine trimer, triNAG, is an inhibitor that binds to the active site of lysozyme. In the presence of 8 mM triNAG, folding and unfolding rates of lysozyme were measured. The apparent folding rate in the presence of triNAG was nearly equal to the intrinsic folding rate in the absence of triNAG, while the apparent unfolding rate was significantly diminished. This means that specific interactions between triNAG and the cleft of lysozyme molecule work in the T state as little as in the U state, and become available only at the last stage of folding from T to N.

3. Transition states in the folding of 3SS-variants of lysozyme

In order to characterize the transition state of lysozyme folding, a specific disulfide bridge was modified by the amino acid substitution, and effects on the folding and unfolding rates were evaluated (18). Four species of three-disulfide variants of hen lysozyme (3SS-variants of lysozyme) were prepared by replacing two Cys residues with Ala or Ser: C6S/C127A, C30A/C115A, C64A/C80A, and C76A/C94A (33). For convenience, we call them des(6-127), des(30-115), des(64-80), and des(76-94) variants, respectively. In addition, the recombinant hen lysozyme containing four authentic disulfide bridges and the extra N-terminal Met was studied as a standard of reference. The recombinant lysozyme is called 4SS-lyz. It was confirmed that the folding rate of 4SS-lyz was the same as that of wild-type lysozyme. Kinetic experiments were carried out in the presence of 0.24 mM triNAG, because the change in the intrinsic fluorescence associated with the refolding was too small to determine the rate constant accurately. It was confirmed experimentally that folding rates of 3SS-variants of lysozyme were

independent of the concentration of triNAG. As shown in Figure 11, the folding rate of des(64-80) variant is the fastest and almost the same as that of 4SS-lyz. Folding rates of des(30-115) and des(6-127) variants are also close to that of 4SS-lyz, although the former are a little slower than the latter. Only the folding rate of des(76-94) variant is much slower than the others. Since unfolding rates of 3SS-variants are dependent on the concentration of triNAG, apparent unfolding rates were determined in the presence of a fixed concentration of triNAG (0.24mM). As shown in Figure 11, the unfolding rate of des(64-80) variant is the fastest and those of des(6-127) and des(30-115) variants are the second, and that of des(76-94) variant is markedly slower than the others. Intrinsic unfolding rates in the absence of triNAG could be determined from the dependence of apparent unfolding rates on triNAG concentrations. Although they were only determined at 1.46 M GuHCl, the results mentioned above remained unaltered qualitatively.

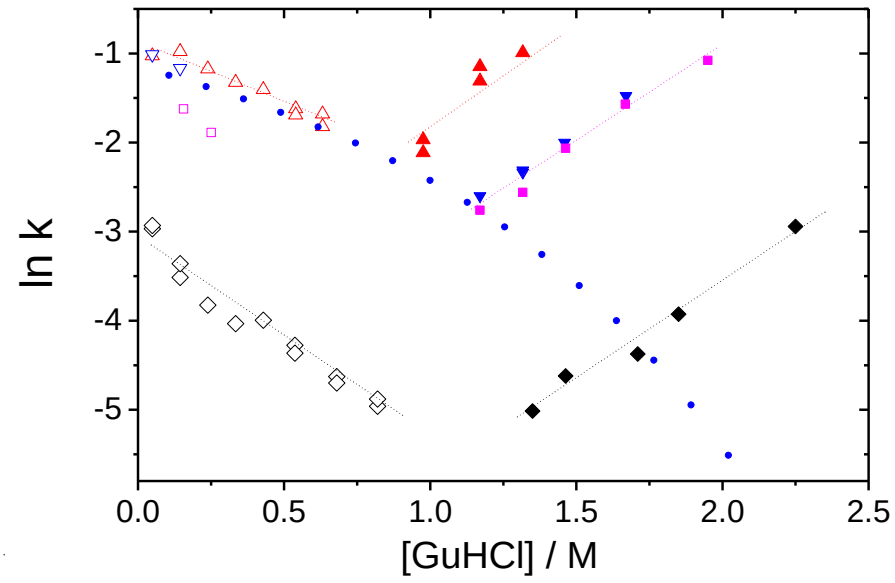

Figure 11. Dependence of folding and unfolding rates of 3SS-variants of lysozyme on the concentration of GuHCl at pH 3.0 and 4.0°C. (•) Folding rates of 4SS-lyz are shown as a standard of reference. They were measured in the absence of triNAG and in the presence of 0.24 mM triNAG. Folding rates were not influenced by the presence of triNAG. Filled and open symbols denote k_{uf} and k_f of various 3SS-variants of lysozyme, respectively. (▽) des(6-127), (□) des(30-115), (Δ) des(64-80), (◊) des(76-94). The unfolding rates of 3SS-variants were measured in the presence of 0.24 mM triNAG.

Two disulfide bridges located close to each other, Cys64-Cys80 and Cys76-Cys94, are opposite extremes in effects on the folding and unfolding rates. The fact that the folding rates of des(64-80) and 4SS-lyz are almost the same suggests that surroundings

around residues 64 and 80 remain disordered in the transition state of des(64-80). Therefore, the cross-linking between residues 64 and 80 contributes to the decrease in conformational entropy not only of the U state, but also of the T state. It follows that surroundings around the Cys64-Cys80 in the T state of 4SS-lyz remain as disordered as those in the U state of 4SS-lyz. On the other hand, the fact that only the folding rate of des(76-94) is extremely reduced implies that the peptide chain from 76 to 94 is structured such that both residues come close to each other in the T state of des(76-94). Therefore, surroundings around Cys76-Cys94 in the T state of 4SS-lyz seem as structured as those in the N state of 4SS-lyz. The fact that only the unfolding rate of des(76-94) is markedly slower than the others is also consistent with the structural features of the transition state mentioned above. Thus, kinetic data on folding and unfolding rates verify that surroundings of Cys64-Cys80 are just opposite of those of Cys76-Cys94 in the transition state.

As a conclusion, lysozyme folding must be accomplished through the transition state in which the peptide chain from 57 through 75 mainly intervening between Cys64 and Cys80 remains largely flexible, regardless of the presence or absence of Cys64-Cys80. On the other hand, the peptide chain from 76 through 94 must be structured firm such that its two ends come close to each other in the T state, regardless of the presence or absence of Cys76-Cys94. Both variants of des(6-127) and des(30-115) are similar to des(64-80) rather than des(76-94) in characteristics of the T state. As a result, it is likely that the D-helix adjacent to Cys30-Cys115 and the C-terminal 3_{10}-helix adjacent to Cys6-Cys127 remain disordered in the T state. On the other hand, kinetic observation of the cross-linked lysozyme mentioned above indicated that the interior of the α-domain including Glu35 and Trp108 was structured firm in the T state as well as in the N state.

VI. NMR STRUCTURAL STUDIES ON 3SS-VARIANTS OF LYSOZYME: LIMITS OF THE NATIVE STRUCTURE

1. Protein structures with a marginal stability

At the first stage of unfolding reaction before reaching the transition state, a partially unfolded state has a strong chemical-affinity toward the native state, that is to say, in each elementary step the rate constant toward the native state is much faster than that of the reverse reaction. At the second stage of unfolding after passing the transition state, the situation is reversed, and each reaction step has a strong chemical-affinity toward the unfolded state. The transition state is a turning point in the protein unfolding, and is the limit of structural fluctuations in the native state. As described already, a kinetic study such as the φ-value analysis was effective to get the information about the transition state, but it was impossible to observe directly the structure of proteins in the transition state. If we were able to prepare a protein derivative with a marginal stability by applying some chemical modifications to the protein, and realize the three-dimensional structure just before the unfolding transition in an equilibrium state, it directly might make a great contribution toward the characterization of the transition state in atomic detail.

For example, when we delete the disulfide bridge from the intact lysozyme one by one, lysozyme derivatives lose the thermal stability of the tertiary structure, then become partially unstructured, and finally lose the tertiary structure completely. Wild-type lysozyme has four disulfide bridges: Cys6-Cys127, Cys30-Cys115, Cys64-Cys80, and Cys76-Cys94. We prepared various types of disulfide-deficient variants: four species of 3SS-variants, three species of 2SS-variants, four species of 1SS-variants, and 0SS-variant lacking all of the disulfide bridges. All of the 3SS-variants had the tertiary structures similar to that of the wild-type (18,20,33), while all of the 1SS-variants were unstructured (22). 2SS-variants are on the border line between the folded and unfolded states (19,21, 34). Below, we will describe details of experimental results on each variant. Three-dimensional structures of 3SS-variants of hen lysozyme lacking one disulfide bridge (C30A/C115A, C64A/C80A, and C76A/C94A) were studied by using the NMR spectroscopy (20). These variants are called des(30-115), des(64-80), and des(76-94) variants, respectively. NMR studies of C6S/C127A were not carried out, because a 3SS-variant lacking Cys6-Cys127 had been prepared by the chemical modification and studied in detail by the X-ray crystallographic method (35). It is the Cys6-Cys127 reduced carboxymethylated derivative of lysozyme (6-127 rcm-lyz), and its three-dimensional structure was almost the same as that of the wild-type except for a few residues from the C-terminus.

2. Sequential NOE connectivities between adjacent residues

The structures of 3SS-variants were thoroughly examined based on NOE contacts

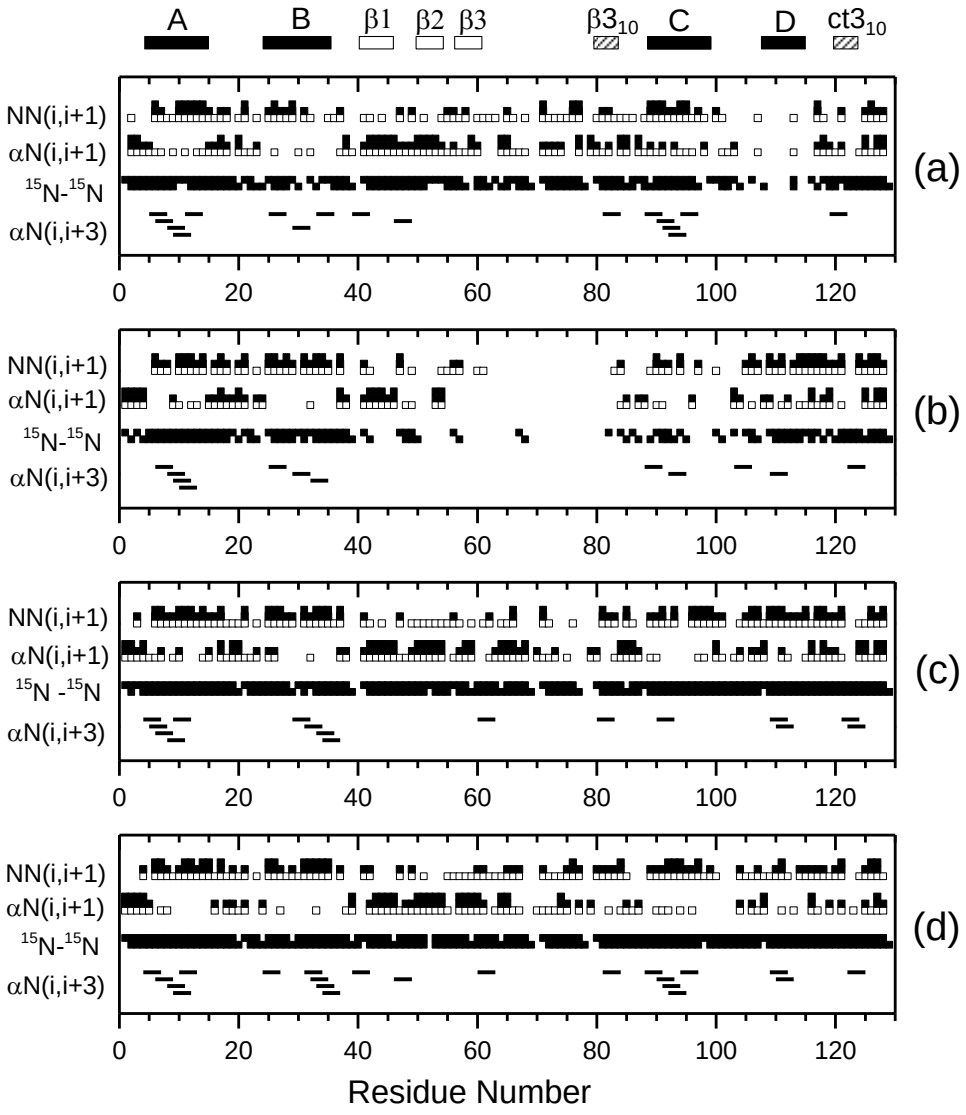

Figure 12. Summary of sequential and short-range NOE connectivities of 3SS-variants of lysozyme. (a) des(30-115), (b) des(64-80), (c) des(76-94), (d) 4SS-lyz. NN(*i*, *i*+1), αN(*i*, *i*+1), and αN(*i*, *i*+3) were detected in the 3D NOESY-HSQC spectra (36) measured at pH 3.8 and 25°C. ^{15}N-^{15}N connectivities between neighboring residues were obtained from the 3D HSQC-NOESY-HSQC spectra (36) measured under the same condition. NOE cross-peaks, NN(*i*, *i*+1) and αN(*i*, *i*+1), are classified into three groups according to their relative intensities, as described in the text. In the line of ^{15}N-^{15}N connectivities, upper and lower squares at the residue *i* correspond to the connectivity from *i* to *i*+1 and that from *i*-1 to *i*, respectively. For reference, the secondary structures observed in the wild-type lysozyme are shown along the residue numbers. A, B, C and D denote α-helices, and β1, β2, and β3 are β-strands. β3$_{10}$ and ct3$_{10}$ denote 3$_{10}$-helices in the β-domain and near the C-terminus, respectively. The assigned chemical shifts were deposited in the BMRB database with accession numbers of 6415 (des(30-115)), 5803(des(64-80)), 5804(des(76-94)), and 7160(4SS-lyz).

determined by NMR spectroscopy (20), which were sequential NOE connectivities found between adjacent residues and long-range NOE contacts found between *i*-th and *j*-th residues of $|i-j| \geq 5$. Figure 12 represents sequential connectivities of des(30-115), des(64-80), and des(76-94) variants. NN(*i*,*i*+1) and αN(*i*,*i*+1) denote ^{1}HN-^{1}HN and ^{1}HC$_{\alpha}$-^{1}HN NOE cross-peaks between *i*-th and (*i*+1)-th residues, respectively. ^{15}N-^{15}N represents sequential connectivities between ^{15}N nuclei of adjacent residues obtained from 3D-HSQC-NOESY-HSQC spectra (36). 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. Short lines in Figure 12 represent observed αN(*i*,*i*+3) connectivities between *i*-th and (*i*+3)-th residues, which are strong indications that the polypeptide backbone adopts a regular helical structure (37). Sequential NOE connectivities of 4SS-lyz, which is the recombinant hen lysozyme containing four authentic disulfide bridges and an extra N-terminal Met, are also shown as a standard of reference in the same Figure (Figure 12(d)). For comparison, secondary structures determined by X-ray analysis of the wild-type lysozyme are indicated along the residue numbers. In general, NN(*i*,*i*+1) and αN(*i*,*i*+1) are indications of α-helix and β-strand, respectively (37). The correspondence between NOE data and X-ray data was very good.

Strong NN(*i*,*i*+1) and αN(*i*,*i*+1) connectivities observed for des(76-94) variant show that the secondary structures of this variant are quite similar to those of the wild-type. However, αN(*i*,*i*+3) connectivities disappear within the C-helix (residues 89-99) of des(76-94) unlike the C-helix of des(30-115) or 4SS-lyz. This may be a minor influence of the deletion of Cys76-Cys94 from the C-helix. On the contrary, in des(64-80) variant, strong αN(*i*,*i*+1) are lost in the β2-strand (50-54) and β3-strand (57-61), and further almost all sequential NOE connectivities disappear from the region of residues 62-80. This verifies that the β3-strand and the loop region (62-79) in the β-domain are disordered in des(64-80). In des(30-115) variant, NN(*i*,*i*+1) are weakened within the B-helix (24-36) and D-helix (108-115). In particular, ^{15}N-^{15}N connectivities are lost within the D-helix, and αN(*i*,*i*+3) are weakened within the B-helix. This shows that the B-helix is partly disturbed, and the D-helix is disordered. However, strong αN(*i*,*i*+1) detected in the entire region of the β-sheet (41-61), strong NN(*i*,*i*+1) found in the A-helix (4-15) and C-helix, and especially αN(*i*,*i*+3) detected within the C-helix altogether indicate that the structural disorder generated within the B-helix or the D-helix is localized there, and does not propagate toward the A-helix or the C-helix.

3. A map of long-range NOE contacts

Long-range NOE contacts are important to the analysis of the three-dimensional structure of protein. NOE cross-peaks are usually detected between a pair of protons within a distance of 0.5 nm (37). In general, long-range NOE cross-peaks tend to be observed between side-chain protons of two distant residues in the amino acid sequence. In the spectral regions below 1.0 ppm, between 5.2 and 7.0 ppm, and above 9.0 ppm along the F2 axis of 2D ^{1}H-^{1}H NOESY spectrum, we could find many cross-peaks well isolated from each other. As a result, 193, 169, 117, and 150 long-range NOE contacts were obtained for 4SS-lyz, des(76-94), des(64-80), and des(30-115) variants, respectively. For example, Figure 13 shows the map of long-range NOE contacts of 4SS-lyz. Since multiple NOE contacts are found between *i*-th and *j*-th residues, for example ^{1}HN(*i*)-^{1}HCβ(*j*), ^{1}HCβ(*i*)-^{1}HCγ(*j*), etc., symbols denoting the NOE contact between a pair of residues vary in sizes and colors according to the number of detected NOE contacts (see the caption of Figure 13). A contact map is divided into three regions: residues 1-39, 40-87, and 88-129. The region from 1 through 39 is called the N-terminal half of the α-domain, from 88 through 129 the C-terminal half of the α-domain, and from 40 through 87 the β-domain.

4. Long-range NOE contacts in 4SS-lyz

First, we describe NOE contacts obtained in 4SS-lyz as a standard of reference. Figure 1 or 19 is helpful for a better understanding. Within the N-terminal half of the α-domain, there are three groups of NOE contacts: J_{AB}, A-B, and hinge. 'J_{AB}' denotes a group of NOE contacts found near the joint (J_{AB}) connecting the A-helix with the B-helix: M12-L17, L17-L25, D18-L25, N19-L25, L17-W28, and Y20-W28. 'A-B' denotes a group of NOE contacts obtained at the interface between A- and B-helices: L8-V29, L8-A32, and A9-V29. Further, 'hinge' represents NOE contacts at the sites where both ends of the N-terminal half of the α-domain come close to each other: V2-N39, F3-N39, F3-F38, R5-F38, and L8-F38, because residues 38-40 correspond to the so-called hinge connecting two domains with each other. Within the C-terminal half of the α-domain, 'J_{CD}' denotes a group of NOE contacts found near the joint (J_{CD}) connecting the C-helix with the D-helix: I98-A107, I98-W108, V99-M105, and V99-W108.

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 (120-124, ct3_{10}) and the B-helix, between the C-terminus and the A-helix, and further between the C-helix and the joint J_{AB}. They are referred to as 'B-D', 'ct3_{10}-B', 'ct-A', and 'C-J_{AB}', respectively in Figure 13. Long-range NOE contacts, N27-W111,

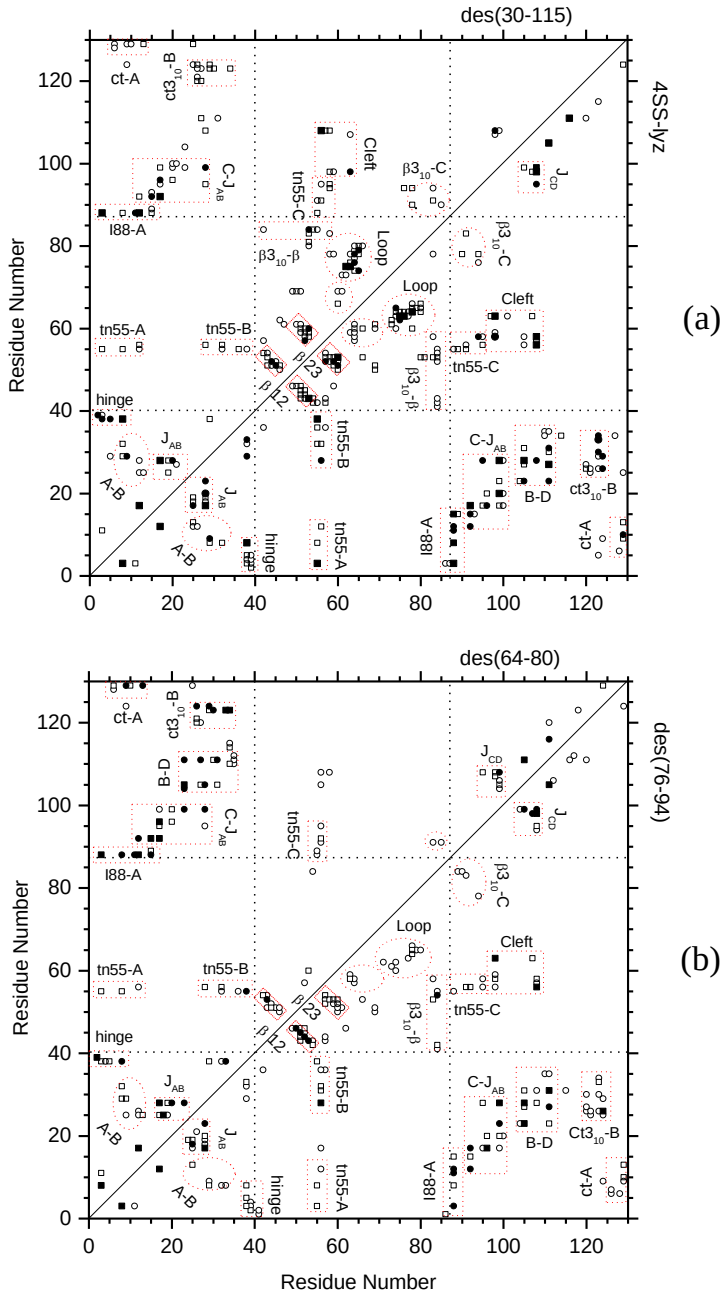

Figure 13. Long-range NOE contact maps of 3SS-variants. (a) NOE contact maps of des(30-115) and 4SS-lyz are above and below the diagonal line, respectively. (b) NOE contact maps of des(64-80) and des(76-94) are above and below the diagonal line, respectively. Long-range NOE contacts are NOE cross-peaks found between protons of the *i*-th and *j*-th residues ($|i-j| \geq 5$), and in most cases obtained in 2D ^{1}H-^{1}H NOESY spectrum (pH 3.8 and 25°C), particularly in the spectral regions below 1.0 ppm, between 5.2 and 7.0 ppm, and above 9.0 ppm. According to the number of NOE cross-peaks detected between a given pair of residues, symbols of NOE contacts vary in size and color, (○) 1; (□) 2-3; (●) 4-5; (■) over 6. NOE contacts are grouped and enclosed by dotted lines, and labeled by some abbreviations (see text).

A31-W111, E35-A110, W28-M105, Y23-W111, Y23-M105, and W28-W108 are involved in 'B-D', which form a part of hydrophobic core in the α-domain. On the other hand, NOE contacts involved in 'C-J_{AB}' form the other part of hydrophobic core, and are especially important in maintaining the structure of the α-domain. That is to say, residues on the C-helix facing the center of the α-domain are close to residues within the joint J_{AB}, and the respective side chains come into contact with each other. The highlight of them is V92-(M12, H15, L17), K96-(L17, Y20), and V99-(Y20, W28). NOE contacts between I88 and the A-helix ('I88-A') may be added to them (see Figure 13). These NOE contacts predominate at the interface between N- and C-terminal halves of α-domain.

Also within the β-domain of 4SS-lyz, there exist a lot of long-range NOE contacts. In Figure 13, 'β12' denotes a group of NOE contacts between β1-strand (41-46) and β2-strand, and 'β23' between β2- and β3-strands. In addition, many long-range NOEs are involved in the loop region referred to as 'Loop'. Further, NOE contacts of (Q41, A42, T43)-L84 or Y53-(A80, L83, L84) are found in the region between the 3_{10}-helix (80-84) and the β-sheet, and referred to as 'β3_{10}-β', where β3_{10} means the 3_{10}-helix existing in the β-domain.

On the other hand, inter-domain contacts are found between the β-domain and the N-terminal half of α-domain. Only I55 and L56, which are located at the hairpin turn from the β2-strand to the β3-strand and referred to as 'tn55', participate in the long-range NOE contacts. In Figure 13, 'tn55-A' and 'tn55-B' represent a group of NOE contacts between tn55 and the A-helix, and between tn55 and the B-helix, respectively. Indeed, side chains of I55 and L56 have long-range NOE contacts with hydrophobic residues of F3, L8, M12, W28, A32, and F38. In addition, there exist many NOE contacts between the β-domain and the C-helix. For example, tn55 also participate in the contact with the C-helix ('tn55-C'): I55-(I88, T89, S91), L56-(A95, W108). NOE contacts are found around the active site of the enzyme with a cleft. They are I58-C94, I58-I98, I58-W108, W63-I98, and W63-A107, and referred to as 'Cleft'. Besides, 'β3_{10}-C' denotes a group of NOE contacts found between the β3_{10}-helix and the C-helix: I78-(A90, C94) and L83-S91. It is very important to the folding of lysozyme that tn55 (I55 and L56) is embedded in the hydrophobic pocket surrounded by A-, B-, and C-helices, because the basement of the β-sheet, namely tn55, is anchored to the α-domain, and it facilitates to attain the cooperativity between two domains. Also, the C-helix plays a crucial role in stabilization of the tertiary structure through NOE contacts with many structural elements: C-J_{AB}, J_{CD}, tn55-C, Cleft, and β3_{10}-C.

5. Difference between 3SS-variants and 4SS-lyz in their contact maps

The contact map of the des(76-94) variant is quite similar to that of 4SS-lyz. Many long-range NOE contacts still exist in 'Loop' and 'Cleft'. The differences between des(76-94) and 4SS-lyz are limited only in the vicinity of residues 74-78. It is noteworthy that NOE contacts in 'Loop' and 'β3_{10}-β' still exist despite the loss of Cys76-Cys94.

On the contrary, in the contact map of the des(64-80) variant, NOE contacts disappear entirely from the loop region. Further they are lost in 'Cleft' and 'β3_{10}-β'. In the β-sheet of des(64-80), NOE contacts in 'β12' just persist, but those in 'β23' are lost. Probably, Cys64-Cys80, rather than Cys76-Cys94, may significantly contribute to the stabilization of the β-sheet through interactions with the β3_{10}-helix, because Cys64-Cys80 combines the β3_{10}-helix with the peptide segment extending from the β3-strand to the residue 64. Despite the loss of stability of the entire β-domain in des(64-80), NOE contacts in 'β12', 'tn55-A', 'tn55-B', and 'tn55-C' are still preserved. In addition, NOE contacts found in 'C-J_{AB}', 'B-D', and 'ct3_{10}-B', which are located at the interface between N- and C-terminal halves of the α-domain, give definitive evidence that the three-dimensional structure of the α-domain of des(64-80) is virtually the same as that of the wild-type.

The contact map of the des(30-115) variant is similar to that of 4SS-lyz except for 'B-D'. The interface between B- and D-helices, which forms a part of the hydrophobic core within the α-domain, is considerably disturbed in des(30-115). However, the NOE contacts found in 'A-B' indicates that the B-helix itself is maintained firm, and the contact with A-helix is preserved native-like. Further, many NOE contacts are definitely found in 'C-J_{AB}', which forms another part of the hydrophobic core. NOE contacts between residues near the C-terminus and the A-helix or the B-helix are secured despite the disordered D-helix.

6. Structural information on 3SS-variants of lysozyme

Distance-information among residues in lysozyme derivatives, which were directly obtained from sequential NOE connectivities and long-range NOE contact maps, characterizes their three-dimensional structures well. In des(76-94), both α- and β-domains are tightly folded into a native-like structure, and the perturbation in the structure is localized only in the vicinity of residue 76. On the other hand, the structure of the β-domain of des(64-80) is largely flexible, but the overall chain-fold appears to be kept native-like. In fact, the polypeptide segment of residues 40-54 takes a hairpin turn around G49 such that β1- and β2-strands run antiparallel to each other. Further, the hairpin turn 'tn55', which corresponds to the basement of the β-sheet, is properly embedded in the

pocket surrounded by A-, B-, and C-helices. Since both ends of the β-domain (Thr40 and Asp87) also come close to the same place, it looks like a pivot of the β-domain. That is to say, the β-domain itself is entirely unstructured in des(64-80), but the pivot of the β-domain is anchored firm to the interface between α- and β-domains.

The structure remaining in the α-domain of des(30-115) seems marginal in the stability, because a half of the hydrophobic core within the α-domain is disrupted at the interface between B- and D-helices, while the other half of the hydrophobic core is secured in the region between the C-helix and the joint J_{AB}. This marginal stability may originate in the presence of Cys6-Cys127 and/or the presence of the stable A-helix, as indicated by NOE contacts in 'ct-A', 'ct3_{10}-B', 'A-B', and 'I88-A'.

Besides, 6-127 rcm-lyz, namely the des(6-127) variant, had been studied by the X-ray crystallographic method. Although a few residues near the C-terminus were disturbed, the A-helix neighboring on the C-terminus was not affected at all by the loss of Cys6-Cys127. As a result, the entire structure in the α-domain of des(6-127) was maintained as firm as in the wild-type. These results suggest that the presence of the stable A-helix is important to maintain the hydrophobic core composed of A-, B-, and C-helices.

VII. STRUCTURAL FLUCTUATIONS STUDIED BY THE H/D EXCHANGE METHOD

1. NMR studies on structural fluctuations in proteins

NMR studies based on NOE data elucidated the static structures of 3SS-variants of lysozyme. All of the four species of 3SS-variants, including 6-127 rcm-lyz as a des(6-127) variant, had the native-like structure, although local structures were disturbed according to the site of the deletion. Besides the static structures, information on dynamic structures of 3SS-variants is necessary to see the limit of structural fluctuations occurring in the native state. Identifying local structures being easy to fluctuate in protein may give important information about the transition state. Amide hydrogen exchange combined with NMR measurements is a straightforward approach to the study of structural fluctuations in protein (38). Hydrogen bonding of amide hydrogens and shielding them from solvent generally protect the hydrogen to deuterium (H/D) exchange reaction. Proteins fluctuate around the time average structure and occasionally expose buried groups to solvent. Thus, H/D exchange reactions are mediated by local fluctuations in protein structure (local unfolding) or the global unfolding of protein. H/D exchange reactions of 3SS-variants were carried out at pH* 3.8 (the uncorrected pH electrode reading) and 25º C, and measured by following the time course of ^{1}H-^{15}N HSQC spectra. The protection factor (PF) of an individual amide hydrogen was determined by comparing the exchange rate (k_{prot}) of an amide measured in a protein with the intrinsic rate (k_{int}) expected in a random coil model (39): PF = k_{int} / k_{prot}. PF is an indicator of structural flexibility around each amino acid residue. Amide hydrogens that are not exchangeable till the global unfolding occurs have the maximum protection factor of $1/K_{uf}$, where K_{uf} is the equilibrium constant of unfolding. Since the stability of the global structure of 3SS-variants is greatly reduced by the loss of a single disulfide bridge, the maximum value of PF of 3SS-variants markedly decreases compared with that of 4SS-lyz. The values of K_{uf} at 25° C and pH 3.8 were determined from calorimetric experiments on respective proteins: $8.6x10^{-8}$ (4SS-lyz), $4.3x10^{-4}$ (des(76-94)), $2.3x10^{-3}$ (des(64-80)), and $2.7x10^{-3}$ (des(30-115)) (18). Protection factors of each 3SS-variant and 4SS-lyz are plotted against the residue number in Figure 14 (20). In 4SS-lyz, amide hydrogens involved in the center of A-, B-, C-helices, and β-sheet have protection factors of more than 10^6 as shown in Figure 14 (d). Thus, exchange reactions of these amide hydrogens are mediated by the global unfolding, while amide hydrogens located at the peripheries of these secondary structures, 3_{10}-helices, and the D-helix are exchanged through structural fluctuations generated locally.

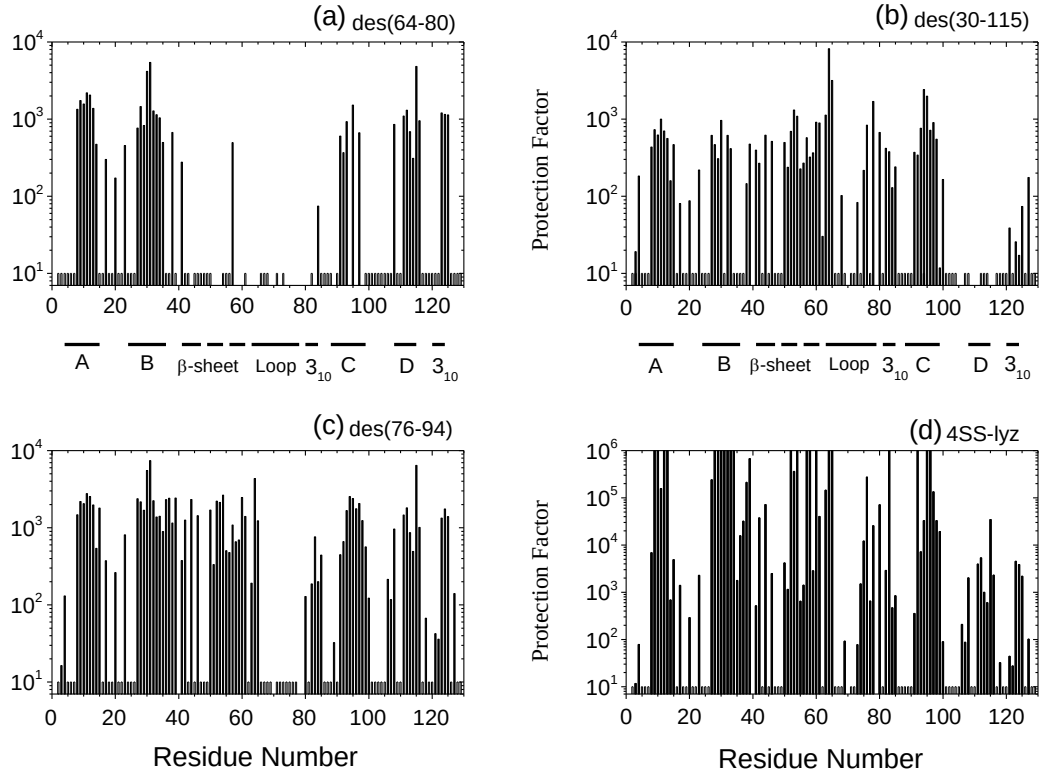

Figure 14. Protection factors of individual amide hydrogens of 3SS-variants of lysozyme against the H/D exchange. (a) des(64-80), (b) des(30-115), (c) des(76-94), (d) 4SS-lyz. Exchange reactions were followed with the change in the integrated volumes of ^{1}H-^{15}N cross-peaks of 2D HSQC spectra measured in D_2O at pH* 3.8 and 25°C. The protection factor (PF) of an individual amide hydrogen was determined by the equation PF = k_{int} / k_{prot}, where k_{prot} was the exchange rate observed in a protein, and k_{int} was the intrinsic exchange rate predicted according to Bai's semi-empirical equation (39). Short open columns represent amide hydrogens whose cross-peaks are detectable in a H_2O solution but not in D_2O because of the fast exchange rate. For reference, the secondary structures in the wild-type lysozyme are shown below residue numbers.

2. Dynamic perturbations of local structures in 3SS-variants of lysozyme

In des(76-94), most amide hydrogens located in all the helices, the β-sheet, and the C-terminal 3_{10}-helix are exchanged with protection factors nearly equal to $1/K_{uf}$ (about 2300). The stability of the global structure of protein is greatly diminished by the deletion of the disulfide bridge, while local structures are not destabilized so much as long as the tertiary structure is secured. As the stability of the global structure decreases, exchange rates controlled by the local unfolding become comparable to those dominated by the global unfolding. It follows that amide hydrogens in the D-helix and C-terminal 3_{10}-helix, which were exchanged through the local unfolding in 4SS-lyz, are exchanged through the global unfolding in the des(76-94) variant. Particularly in the regions where PFs are less

than 2000, protection factors remain unaltered from those of 4SS-lyz. For example, the protection factors in the $\beta 3_{10}$-helix of des(76-94) are nearly equal to those in 4SS-lyz. It indicates that the stability of this region is maintained locally without Cys76-Cys94 as well as in 4SS-lyz. It is likely that this stability is brought by the presence of Cys64-Cys80, which combines the $\beta 3_{10}$-helix (residues 80-84) with Cys64 in the region following to the β3-strand. Also, the entire β-sheet is protected well against the H/D exchange through the interactions between the $\beta 3_{10}$-helix and the β-sheet, as it was already pointed out based on the information of NOE contacts. Only the protection factors of residues 74-78 decrease markedly in des(76-94). These results suggest that the folded structure of des(76-94) is as compact as the wild-type except for the immediate vicinity of residue 76, despite the significant decrease in the stability of the global structure.

In H/D exchange reactions of des(64-80), 45 amide hydrogens disappeared from the first HSQC spectrum after the start of H/D exchange reaction besides those exchanged rapidly in 4SS-lyz. They are primarily the amide hydrogens of residues 42-85 and residues 97-100 on the C-helix. Actually, the entire region of the β-sheet in des(64-80) fluctuates frequently around the time average structure. Further in des(64-80), not only the loop but also the β3-strand and $\beta 3_{10}$-helix are released from the constraint of the disulfide bridge. As a result, it is likely that the entire region of loop and the β3-strand are disordered, and interactions between the β-sheet and the $\beta 3_{10}$-helix are lost. Indeed, amide hydrogens of residues 80-85 are not protected at all in des(64-80), whereas they have definite protection factors in des(76-94). On the contrary, in the α-domain of des(64-80), A-, B-, and D-helices, and C-terminal 3_{10}-helix have protection factors which are dominated by the global unfolding ($1/K_{uf} = 440$). The C-helix is also protected, although its C-terminal part is somewhat disturbed. Thus, the α-domain in des(64-80) is protected against the H/D exchange as firm as in des(76-94).

In des(30-115), A-, C-helices, and the entire region of the β-domain including the loop and the $\beta 3_{10}$-helix have protection factors nearly equal to $1/K_{uf}$ (about 370). These regions seem to be folded firm till the amide hydrogens are exposed into solvent upon the global unfolding. On the other hand, protection factors in the C-terminal 3_{10}-helix decrease considerably compared to those of other 3SS-variants. In particular, amide hydrogens are no longer protected within D-helix and at the interface of B-helix (A31, F34, E35, S36, and N37) with D-helix.

3. Detailed structure of proteins in the transition state

Useful information about the transition state could be obtained from the analysis of rate constants of folding and unfolding of 3SS-variants. In brief, lysozyme folding seems

to be accomplished through the transition state in which the peptide chain from 57 through 75 mainly intervening between residues 64 and 80 remains unstructured, while the peptide chain from 76 through 94 is structured firm such that its two ends come close to each other. Also, the D-helix adjacent to Cys30-Cys115 and the C-terminal 3_{10}-helix adjacent to Cys6-Cys127 remain disordered, while the interior of the α-domain including Glu35 and Trp108 is structured firm in the transition state.

On the other hand, NMR data such as NOE contacts and/or protection factors against amide H/D exchange give following pictures for each 3SS-variant. The β-domain in the folded structure of des(64-80) is largely flexible except for the pivot of the β-domain including Ile55 and Leu56. However, the α-domain structure is virtually the same as that of the wild-type, and in essential the tertiary structure of des(64-80) is kept native-like. While the entire region of the β-domain of des(76-94) is structured firm except for the peptide segment 74-78 despite the loss of the cross-linking between residues 76 and 94. Although two variants are opposite extremes in structural flexibility in the β-domain, they share the structural feature that residue 76 come close to residue 94 in common. Even though the region including the β-sheet, the loop, and $\beta 3_{10}$-helix becomes largely flexible, the tertiary structure of protein is barely secured as long as the peptide chain from residues 76 through 94 is kept native-like. From the point of view that the transition state is the limit of structural fluctuations occurring in the native state, characteristics of dynamic structures of des(64-80) and des(76-94) variants are consistent with those of the transition state obtained from the analysis of kinetic data.

Since both surroundings of Cys64-Cys80 and Cys76-Cys94 overlap each other, it introduces some confusion to imagine the detailed structure of the β-domain in the transition state. In order to fold the peptide chain 76-94 such that both ends come close to each other, it may be required that the turn from $\beta 3_{10}$-helix to the C-helix is kept native-like, even if the $\beta 3_{10}$-helix fluctuates around the time average structure. On the other hand, it is allowable in the transition state that the loop region (62-79) is disordered and the entire β-sheet (41-61) are largely flexible, although it must be secured that β1- and β2-strands run anti-parallel to each other, and the pivot of the β-domain is properly anchored to the pocket in the α-domain. It seems a prerequisite to accomplish the folding transition that residues 76 and 94 come close to each other without the cross-linking between them. The origin may be explained as follows; the structures indispensable for the folding transition such as the pivot of the β-domain may be induced by drawing the residue 76 near to the residue 94 on the C-helix, because the peptide chain from 41 through 61 has a strong propensity to take a hairpin turn around residues 49 and 55 (see Figure 19).

Further, kinetic data indicate that the D-helix and the C-terminal 3_{10}-helix are disordered in the transition state, but the structural stability in the interior of the α-domain is maintained firm. Detailed analysis of the structures of des(30-115) and des(6-127) variants indicates that the D-helix and C-terminal 3_{10}-helix are disordered, but the stability of the A-helix is not lost, and the NOE contacts between A- and B-helices are secured firm despite the disordered interface between B- and D-helices, and further the C-helix is properly set between A- and B-helices. Since the stability in the α-domain is maintained through long-range interactions between C-helix and the joint region J_{AB} connecting the A-helix with the B-helix, the tertiary structure of protein is preserved even if the D-helix and the C-terminal 3_{10}-helix are disordered. Consequently, the detailed structures of des(30-115) and des(6-127) variants are consistent with the picture of the transition state obtained from the analysis of kinetic data.

VIII. TWO-DISULFIDE VARIANTS ARE ON THE BORDER LINE BETWEEN ORDERED AND DISORDERED STATES

1. CD spectra of three species of 2SS-variants

All of 3SS-variants had the native-like tertiary structure, while all of 1SS-variants were unstructured. Two-disulfide variants are on the border line between ordered and disordered states. We constructed three species of two-disulfide variants of lysozyme (2SS-variants): 2SS[6-127, 30-115] containing two disulfide bridges of Cys6-Cys127 and Cys30-Cys115, 2SS[64-80, 76-94], and 2SS[6-127, 64-80]. In 2SS[6-127, 30-115], two disulfide bridges exist in the α-domain, while 2SS[64-80, 76-94] has one disulfide bridge in the β-domain and the other one at the interface between two domains. In 2SS[6-127, 64-80], one disulfide bridge is included in each domain. Judging from their CD spectra shown in Figure 15 (40), 2SS[6-127, 30-115] appears to have a substantial amount of secondary structure and a definite tertiary structure in water, whereas 2SS [64-80, 76-94] and 2SS[6-127, 64-80] have a limited amount of secondary structure but no tertiary

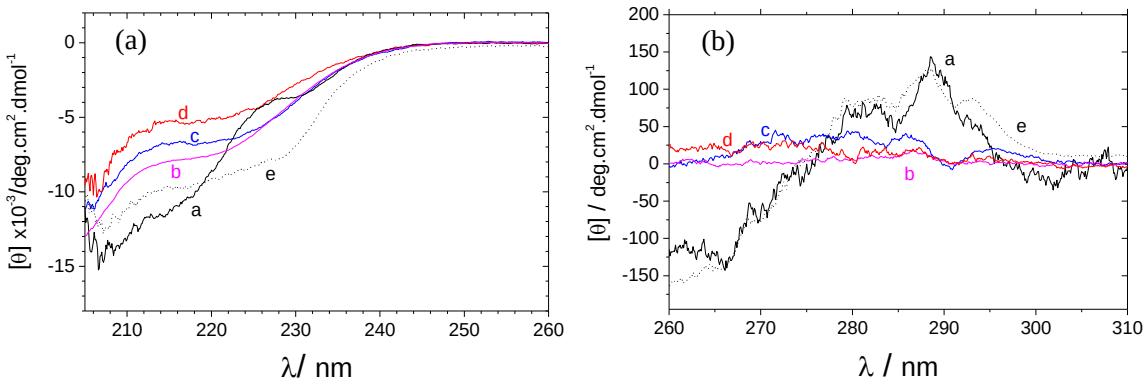

Figure 15. CD spectra of 2SS-variants measured in water at pH 3.8 and 5.0°C. (a) Far-UV CD spectra of 2SS-variants. a: 2SS[6-127, 30-115], b: 2SS[6-127, 64-80], c: 2SS[64-80, 76-94], d: 0SS-variant, e: wild-type. (b) Near-UV CD spectra of 2SS-variants. a: 2SS[6-127, 30-115], b: 2SS[6-127, 64-80], c: 2SS[64-80, 76-94], d: 1SS[64-80], e: wild-type. The far-UV CD spectrum of 2SS[6-127, 30-115] is different from others near at 230 nm. On the other hand, the near-UV CD spectrum of 2SS[6-127, 30-115] is quite similar to that of the wild-type. It is noteworthy that the far-UV CD spectrum of 2SS[6-127, 64-80] is somewhat close to that of the wild-type, but the near-UV CD spectrum is lost as entirely as those of 2SS[64-80, 76-94] and 1SS[64-80].

structure in water. The former belongs to variants with a tertiary structure, while the latter to those with an unstructured conformation. However, 2SS[6-127, 64-80] was different from 2SS[64-80, 76-94] in a solution of concentrated glycerol; in the presence of 50% glycerol, the former had a native-like tertiary structure, but the latter remained

unstructured. The glycerol-induced folding of 2SS[6-127, 64-80] will be described in next Chapter. We carried out NMR measurements of these 2SS-variants to see their detailed structures in water.

2. NMR studies of three species of 2SS-variants

2D-HSQC spectra were measured in water for three species of 2SS-variants (19). The ^{1}H-^{15}N-HSQC spectrum of 2SS[6-127, 30-115] was very similar to that of des(64-80) 3SS-variant (20), although intensities of cross peaks were weakened. On the other hand, HSQC spectra of 2SS[64-80, 76-94] and 2SS[6-127, 64-80] were characteristic of unfolded proteins in a collapsed non-native state. These observations from NMR spectra are consistent with those from CD spectra.

Resonance assignment was practicable for the HSQC spectrum of 2SS[6-127, 30-115]. Sequential NOE connectivities and long-range NOE contacts were found on its 3D-NOESY-HSQC spectrum. They are shown in Figure 16 and 17, together with those of des(64-80) 3SS-variant. Sequential NOE connectivities of the former are virtually the same as the latter except for the fact that peak-intensities of 2SS-variant are weakened in a part of the C-helix including residues A90-A95.

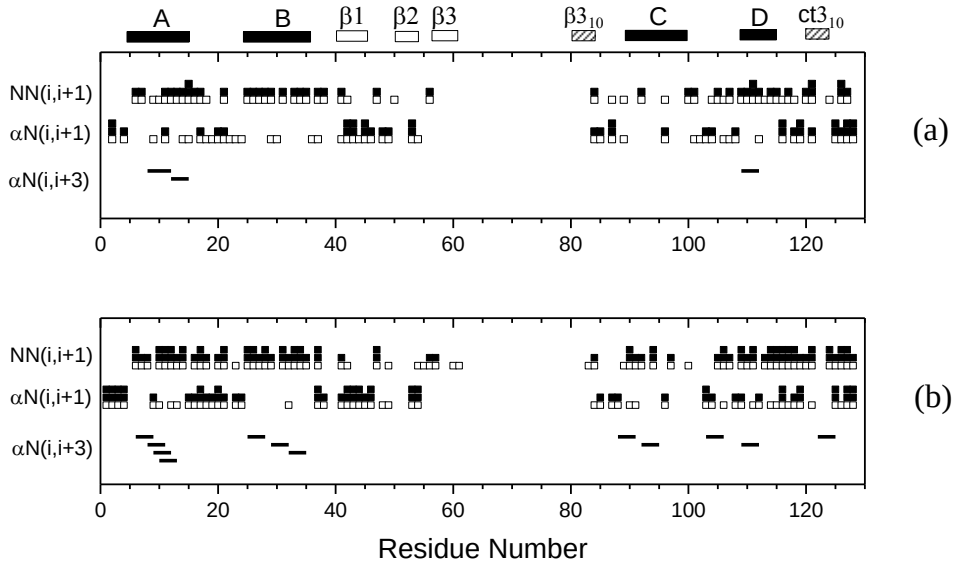

Figure 16. Sequential NOE connectivities of 2SS[6-127, 30-115] (19) and des(64-80) 3SS-variant (20). Symbols used in this figure are the same as those in Figure 12. (a) 2SS[6-127, 30-115], (b) des(64-80) 3SS-variant. The assigned chemical shifts of 2SS[6-127, 30-115] were deposited in the BMRB database with accession numbers of 7159.

The tertiary structure of 2SS[6-127, 30-115] was examined from the NMR distance

information, namely, the NOE contact map of 2SS[6-127, 30-115] was constructed from long-range NOE cross-peaks which had been found mainly in 2D ^{1}H-^{1}H NOESY spectrum. As shown in Figure 17, contact maps of both 2SS[6-127, 30-115] and des(64-80) are almost the same as each other, particularly in the interface between N- and C-terminal halves of the α-domain. It clearly indicates that the α-domain structure is unaltered by the deletion of both Cys64-Cys80 and Cys76-Cys94. Between 2SS[6-127, 30-115] and des(64-80), only a little difference is observed in NOE contacts between C-helix and the turn 'tn55'. It may be because the C-helix including the residue 94 is more disturbed in 2SS[6-127, 30-115] than in des(64-80) 3SS-variant.

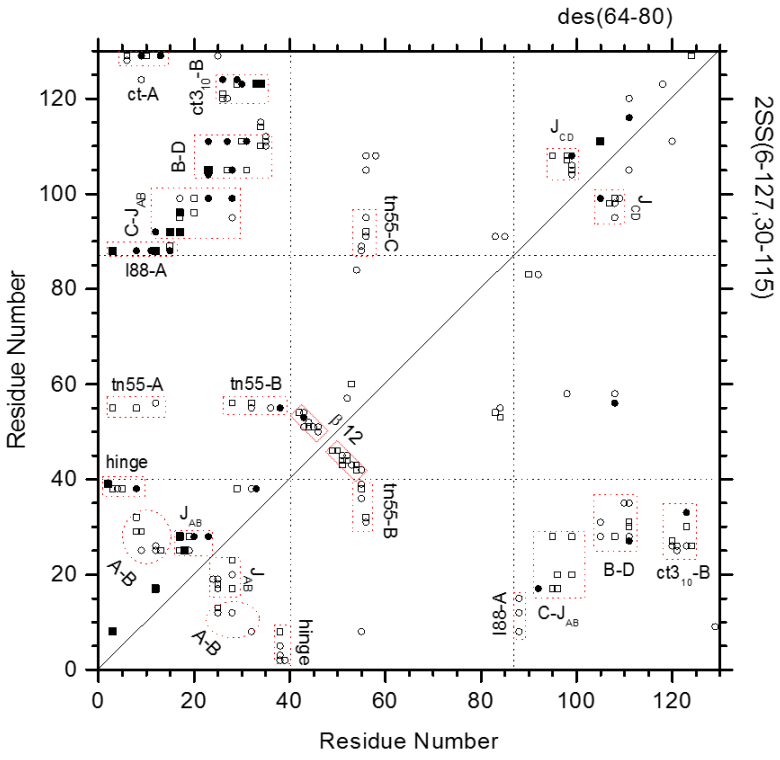

Figure 17. Long-range NOE contact maps of 2SS[6-127, 30-115] and des(64-80) 3SS-variant. Symbols used in this figure are the same as those in Figure 13. The contact maps of des(64-80) and 2SS[6-127, 30-115] are above and below the diagonal line, respectively.

3. Effects on the β-domain structure of the removal of Cys64-Cys80 or Cys76-Cys94

The removal of Cys76-Cys94 from des(64-80) 3SS-variant has little influence on the tertiary structure of des(64-80), although the stability of the tertiary structure diminishes. These results are consistent with those caused by the deletion of Cys76-Cys94 from 4SS-lyz. On the other hand, the removal of Cys64-Cys80 from 4SS-lyz or from

des(76-94) 3SS-variant causes large-scale structural fluctuations in the β-domain, including the entire β-sheet, the loop region, and the $\beta 3_{10}$-helix.

As mentioned above, the deletion of Cys76-Cys94 from des(64-80) 3SS variant contributes only to the destabilization of the C-helix. This effect is more clearly reflected on protection factors against amide H/D exchanges. Indeed as shown in Figure 18, protection factors of the C-helix in 2SS[6-127, 30-115] are much smaller than those of the D-helix and/or the C-terminal 3_{10}-helix, though the former was comparable to the latter in des(64-80) 3SS-variant. Although the peptide chain from 76 through 94 has a propensity for the residue 76 to come close to the residue 94 without a disulfide bridge between them, structural fluctuations within the C-helix grow larger by the loss of Cys76-Cys94. Besides, the entire β-domain, including the β-sheet, the loop, and the $\beta 3_{10}$-helix,

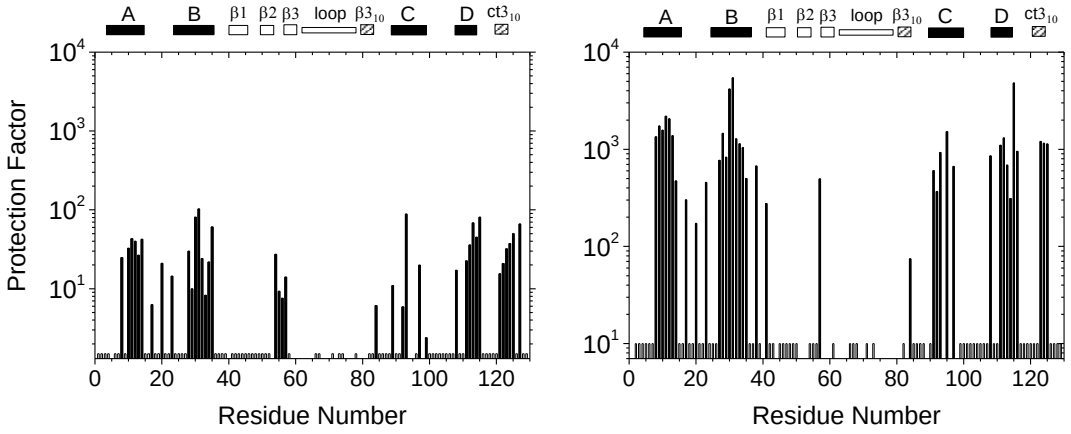

Figure 18. Protection factors of individual amide hydrogens against the H/D exchange measured at pH* 3.8 and 25°C. (Left) 2SS[6-127, 30-115] (41). (Right) des(64-80) 3SS-variant (20). Also in this figure, short open columns represent the amide hydrogens whose cross-peaks are detectable in H_2O but not in D_2O. It is noteworthy that the tertiary structure of 2SS[6-127, 30-115] is less stable than that of des(64-80), but similar regions in the α-domain are protected against the H/D exchange in both variants.

becomes greatly flexible in 2SS[6-127, 30-115] due to the loss of Cys64-Cys80. Nevertheless, the NOE contact map of 2SS[6-127, 30-115] demonstrates that β1- and β2-strands are just maintained as an anti-parallel β-sheet (refer to Figure 17). In addition, the tn55 is located in the vicinity of the B-helix (referred to as 'tn55-B'). Further, the C-helix is properly set between A- and B-helices with side-chains of I88, V92, K96 and V99 buried in the hydrophobic core of the α-domain (referred to as 'C-J_{AB}' and 'I88-A'). These observations suggest that the stability of the pivot of the β-domain is secured but is really marginal in 2SS[6-127, 30-115]. Figure 19 shows a schematic representation of the pivot

remaining in the tertiary structure of des(64-80) 3SS-variant, which remains unaltered also in 2SS[6-127, 30-115]. The disruption of the pivot structure, which seems just the trigger of the global unfolding, may happen, when the residue 76 moves away from the residue 94 to cause a large-scale structural fluctuation of the C-helix.

The removal of Cys76-Cys94 from des(30-115) 3SS-variant results in the production of 2SS[6-127, 64-80]. The fact that 2SS[6-127, 64-80] is unstructured in whole may originate in the instability of the C-helix caused by the deletion of Cys76-Cys94 in addition to the marginal stability of the α-domain without Cys30-Cys115. The pivot structure could be barely maintained in 2SS[6-127, 30-115], but not be preserved in 2SS[6-127, 64-80].

4. Differences between 2SS[6-127, 30-115] and 2SS[64-80, 76-94]

2SS[6-127, 30-115] has a native-like tertiary structure despite the largely disturbed β-domain, but 2SS[64-80, 76-94] is unstructured in whole, judging from CD and NMR spectra. 2SS[64-80, 76-94] loses ordered structures not only in the α-domain but also in

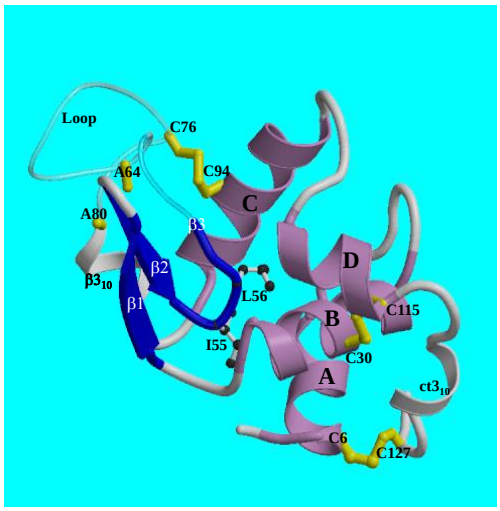

Figure 19. Schematic representation of the pivot of the β-domain remaining in des(64-80) variant. The β-domain including the β-sheet, the loop, and $\beta3_{10}$-helix is entirely unstable in whole, but the β1- and β2-strands are arranged in an anti-parallel, and I55 and L56 are anchored to the pocket surrounded by A-, B-, and C-helices. As a result, besides of I55 and L56, the both ends of the β-domain (T40 and D87) come close to the same place together. This pivot structure is essentially preserved also in 2SS[6-127, 30-115].

the β-domain, despite the existence of both Cys64-Cys80 and Cys76-Cys94. Constraints of these disulfide bridges alone are insufficient to maintain the β-domain firm. What is necessary for the β-domain to be secured? The cooperativity between α- and β-domains appears indispensable to maintain the stable structure of the β-domain. As described

already, the cooperativity arises from interactions between the β3-strand and the C-helix, and the pivot of the β-domain anchored in the α-domain. Thus, the C-helix plays a crucial role in maintaining the cooperativity between α- and β-domains. In 2SS[64-80, 76-94], there is no doubt that the α-domain is disordered by the loss of both Cys6-Cys127 and Cys30-Cys115, and the pivot of the β-domain is unable to get a footing because the hydrophobic pocket composed of A-, B- and C-helices is lost. So that the β-domain leads to the completely disordered state, despite the presence of both Cys64-Cys80 and Cys76-Cys94. Although Cys64-Cys80 plays an important role in structuring the β-domain, and Cys76-Cys94 linked to the C-helix is able to suppress the structural fluctuations occurring in the C-helix, it may be a prerequisite to exhibit their ability that the C-helix is formed neighboring to the β3-strand, and the footing for the pivot of the β-domain to be anchored is secured in the α-domain.

5. 2SS-variants lacking either one of two disulfide bridges in the α-domain

The tertiary structure of 2SS[6-127, 30-115] was barely maintained despite the loss of both Cys64-Cys80 and Cys76-Cys94. It is due to the fact that long-range interactions between N- and C-terminal halves of the α-domain are strong enough due to the presence of two disulfide bridges in the α-domain, so that the pivot of the β-domain is anchored to the pocket in the α-domain. When either one of the disulfide bridges is removed from the α-domain, is the α-domain able to maintain its stable structure?

The tertiary structure of the α-domain with only one disulfide bridge is maintained, if both of Cys64-Cys80 and Cys76-Cys94 are present, namely they are 3SS-variants, des(6-127) or des(30-115). We attempted to prepare 2SS-variants in which either one of Cys6-Cys127 and Cys30-Cys115 was left in the α-domain, and either one of Cys64-Cys80 and Cys76-Cys94 was left to keep the cooperativity between α- and β-domains. In order to maintain the β-domain structure, it is much better to leave Cys64-Cys80 rather than Cys76-Cys94 for reasons mentioned already. On the other hand, the A-helix plays a key role in structuring the α-domain rather than the D-helix. For example, interactions between A- and B-helices are important to keep the configuration of the joint J_{AB} native-like. It is crucial for the C-helix properly to be set between A- and B-helices in constructing the hydrophobic core in the α-domain. Besides, the evidence that the A-helix is indispensable but ten residues from the C-terminal end are nonessential for the maintenance of the α-domain was brought from another experiment on the truncation of amino acid residues from 6-127 rcm-lyz. That is to say, a lysozyme derivative lacking Cys6-Cys127 (6-127 rcm-lyz) was prepared, and then amino acid residues were truncated from the N-terminus or the C-terminus. Although the deletion of residues 120-129 from

6-127 rcm-lyz had little influence on the tertiary structure, the deletion of residues 1-14 led to the complete disruption of the native structure despite the presence of Cys30-Cys115 (42). This is definitive evidence that the A-helix is crucial in maintaining the overall structure of the α-domain. Therefore, we left Cys6-Cys127 rather than Cys30-Cys115 in the α-domain of 2SS-variant to secure the stable A-helix. Consequently, we prepared 2SS[6-127, 64-80] variant to examine a marginal state remaining in the α-domain.

CD and NMR spectra of 2SS[6-127, 64-80] measured in water already indicated that it was unstructured in whole. As mentioned above, it may be because the stability between N- and C-terminal halves of the α-domain is insufficient due to the deletion of Cys30-Cys115, and furthermore the C-helix becomes unstable by the loss of Cys76-Cys94. However, when near-UV CD spectra of 2SS[6-127, 64-80] had been measured in a concentrated glycerol solution, surprisingly they were typical of the folded structure. In next Chapter, we will consider the glycerol-induced folding of unstructured proteins, and examine the tertiary structure in atomic detail.

IX. GLYCEROL-INDUCED FOLDING OF 2SS[6-127, 64-80]

1. Structures induced by glycerol in 2SS-variants of lysozyme

It was an interesting observation that 2SS[6-127, 64-80] variant was unstructured in water but recovered the native-like tertiary structure in a solution containing concentrated glycerol. In order to study the glycerol-induced folding of unstructured proteins, CD spectra of various disulfide-deficient variants were measured in concentrated glycerol solutions. Firstly, far-UV CD spectra were measured at various concentrations of glycerol. As shown in Figure 15, in the absence of glycerol, far-UV CD spectra of 2SS[6-127, 64-80] was virtually the same as those of 2SS[64-80, 76-94]. Also in the presence of 30% (vol.%) glycerol, far-UV CD spectra of both variants were nearly

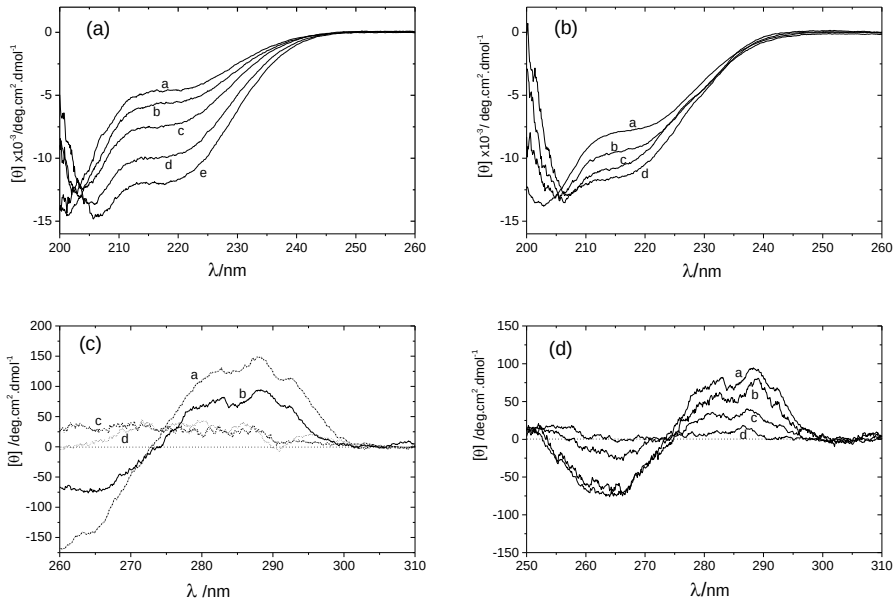

Figure 20. CD spectra of different species of disulfide-deficient variants in solutions containing glycerol (21). All spectra were measured at pH 3.0 and 4°C. (a) Far-UV CD spectra of 1SS[64-80] in the presence of glycerol. a: 0%, b: 10%, c: 30%, d: 50%, e: 70%. Those of 0SS-variant were almost identical with those of 1SS[64-80]. (b) Far-UV CD spectra of 2SS[6-127, 64-80] in the presence of glycerol. a: 0%, b: 10%, c: 30%, d: 70%. (c) Near-UV CD spectra of various variants in the presence of 50% glycerol and those of wild-type lysozyme as a reference. a: wild-type, b: 2SS[6-127, 64-80], c: 1SS[64-80], d: 2SS[64-80, 76-94]. (d) Near-UV CD spectra of 2SS[6-127, 64-80] in the presence of various concentrations of glycerol. a: 50%, b: 30%, c: 10%, d: 0%.

equal to each other, although only the spectrum of 2SS[6-127, 64-80] is shown in Figure 20. In the presence of glycerol, the negative ellipticity at 220 nm, $[\theta]_{220}$, gradually

increased in an absolute value with the increase in glycerol concentrations, as shown in Figure 20 (a) or (b). In the absence of glycerol, absolute values of $[\theta]_{220}$ for 2SS-variants were somewhat larger than those of 1SS-variants and/or 0SS-variant, but they approached each other in a solution containing 70% glycerol. Secondly, near-UV CD spectra of 2SS[64-80, 76-94], 1SS-variants except 1SS[6-127], and 0SS-variant remained nearly equal to zero even in a 50% glycerol solution. On the other hand, those of 2SS[6-127, 64-80] were nearly equal to zero in water, but gradually approached those of the wild-type lysozyme by the addition of glycerol from 0% to 50% (21). These data suggest that the addition of glycerol increases the helical contents of all the variants including 0SS-variant, but the native-like tertiary structure is recovered only in 2SS[6-127, 64-80] variant. As described later, the glycerol-induced helical structures in 0SS-variant or 2SS[64-80, 76-94] seem not to stay at a specific part on the chain unlike the secondary structures induced into 2SS[6-127, 64-80].

2. Mixture of 95% DMSO and 5% D_2O as a quenching buffer

2SS[6-127, 64-80] is unstructured throughout the molecule in water, but the polypeptide chain are gradually frozen into a native-like conformation with increasing glycerol concentrations, as shown in CD spectra. In order to see the detailed mechanism of glycerol-induced folding, NMR spectra were measured in a solution containing 30% glycerol. Cross-peaks of the HSQC spectrum were dispersed as well as in the native state of 4SS-lyz, although their peak-intensities were very weak. However, the peak assignment of the HSQC spectrum of 2SS[6-127, 64-80] was very difficult because of resonance lines broadening out in a glycerol solution. Consequently, we abandoned the peak assignment, and NOESY spectra were unable to give useful information about long-range interactions among residues. Therefore, the observation of glycerol-induced folding was carried out by monitoring the protection factors of amide hydrogens against H/D exchange. Since NMR resonance lines significantly broadened out and overlapped each other in the presence of glycerol, some improvements of experiments were necessary for the study of protection factors.

H/D exchange reactions of 2SS[6-127, 64-80] were carried out in a D_2O solution containing concentrated glycerol at 4° C and pH* 3.0 (uncorrected pH-electrode reading). After the removal of glycerol from the reaction mixture, partially deuterated proteins were freeze-dried. In order to determine H/D exchange rates of amide hydrogens, H/D exchange reaction must be virtually stopped during the acquisition of NMR signals. However, H/D exchange time-constants of 2SS[6-127, 64-80] were much faster than the acquisition time, because the 2SS-variant was inherently unstructured in water. Therefore,

we adopted the mixture of 95% DMSO and 5% D_2O at pH* 5.5 as a quenching buffer in order to measure the NMR spectra of partially deuterated proteins, because the H/D exchange rate had been found to be about 100-fold reduced in the quenching buffer relative to that in water (43). This solvent system had another advantage that it significantly suppressed broadening of the resonance lines of lysozyme variants and enabled us to obtain well-separated NMR spectra with sharp resonance lines. For example, Figure 21 shows the HSQC spectrum measured in a quenching buffer.

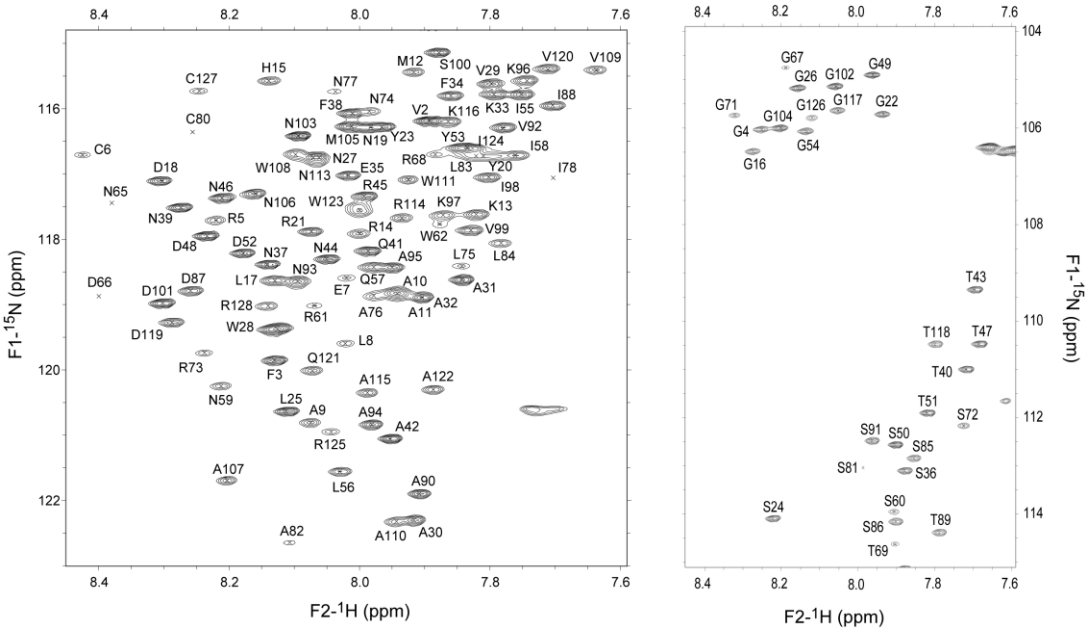

Figure 21. HSQC spectra of 2SS[6-127, 64-80] measured in a mixture of 95%DMSO/5%H_2O at pH 5.5 and 25°C (21). The right chart shows the upper half of the spectrum along the F1-axis, and the left chart the lower half. The cross-peaks of N65, D66, I78 and C80 could be identified but their contour levels were below the threshold. Besides them, intensities of the cross-peaks of residues around Cys6-Cys127 and Cys64-Cys80 were weakened due to line broadening. The assigned chemical shifts were deposited in the BMRB database with accession numbers of 11051(0SS) and 11052(2SS[6-127,64-80]).

3. Protection factors of individual residues in 2SS[6-127, 64-80]

H/D exchange reactions of 2SS[6-127, 64-80] and 0SS-variant were carried out in 20 mM deuterated formate buffer at pH* 3.0 and 4° C in the absence of glycerol. The exchange rates of the same residue in both variants were quite similar to each other. Furthermore, exchange time constants in the formate buffer (τ_{for}) were rather faster than Bai's predicted values due to the low salt concentration of buffer solution. This fact suggests that not only 0SS but also 2SS[6-127, 64-80] has no structure preventing amide hydrogens from being exchanged in the absence of glycerol. Therefore, the ratio τ_{prot}/τ_{for} was used as the protection factor, where τ_{prot} denoted exchange time constants observed

in proteins. In the presence of glycerol, experimental results on H/D exchange reactions of 2SS[6-127, 64-80] are shown in Figure 22. In a solution containing 10% glycerol, protection factors are less than ten at most of residues. With increasing glycerol concentrations, some selected regions are further protected, and their protection factors reach about a thousand in a 30% glycerol solution. The highly protected residues are involved in A-, B- and C-helices and β3-strand, and especially center on Ile55 and Leu56. The profile of protection factors vs. residue numbers is similar to that observed in des(76-94) variant in water. It is noteworthy that protection factors in the C-helix and $\beta 3_{10}$-helix of 2SS[6-127, 64-80] in a 30% glycerol solution are comparable to those of des(76-94) in water. It is likely that the existence of glycerol in addition to Cys64-Cys80 makes a significant contribution to the stabilization of the C-helix and $\beta 3_{10}$-helix in 2SS[6-127, 64-80]. Also, it sounds reasonable that protection factors in the D-helix of 2SS[6-127, 64-80] are much smaller than those of des(76-94), because Cys30-Cys115 does not exist in the former.

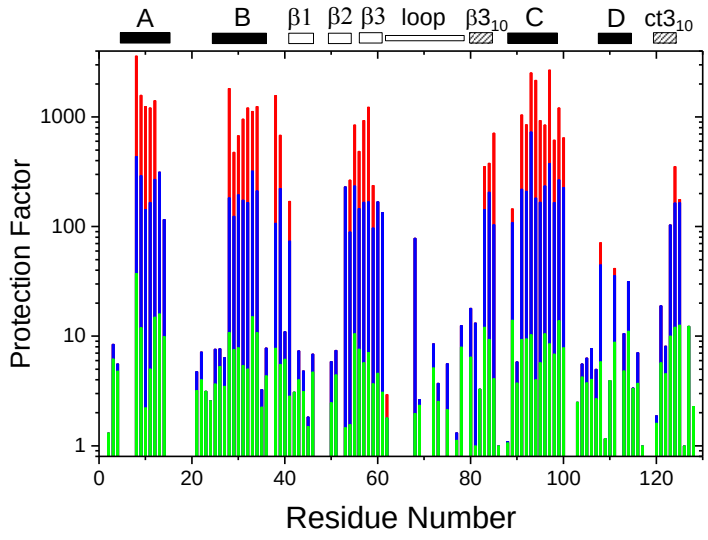

Figure 22. Profiles of protection factors versus the amino acid sequence in 2SS[6-127, 64-80]. Green: 10% glycerol, Blue: 20%, Red: 30%. The locations of structural elements in the wild-type lysozyme are shown by short bars above the diagram.

Glycerol is well known as a solvent additive to stabilize the folded structure of protein through the preferential hydration, that is, a negative binding of glycerol to the protein surface (44,45). Actually in 3SS-variants of lysozyme, it was found that the addition of glycerol significantly increased the stability of the folded structure and

enhanced the efficiency of correct disulfide bridge formation (33,46). The far-UV CD spectra in Figure 20 showed that the addition of glycerol increased the helical contents of all the variants including 2SS[6-127, 64-80] and/or 0SS-variant. As shown in Figure 22, dilute solutions of glycerol tends to induce unstructured protein to recover helical conformations uniformly along a polypeptide chain. For example in the presence of only 10% glycerol, suppose each residue independently changes its conformation, the probability that several consecutive residues are in the form of a helix is low, and helices induced by glycerol seem short and not persistent. It is unlikely that the glycerol-induced helix remains at a specific part on the polypeptide chain and protection factors increase markedly within the segment than others. In order to increase protection factors more, it is necessary for secondary structures induced on a specific segment to be further stabilized through long-range interactions between them. Such a structure with protection factors of about 100 or more seems a persistent three-dimensional structure induced by concentrated glycerol.

The preferential hydration of unstructured protein by glycerol is likely to contribute to the enhancement of hydrophobic interactions among residues. The fact that 2SS[6-127, 64-80] is unstructured in water may be attributable to the disruption of the α-domain due to the penetration of water. The observation of protection factors suggests that long-range interactions are recovered in the α-domain due to the preferential hydration by glycerol, where many hydrophobic residues come close together, including A- and B-helices, and residues on the C-helix facing the joint J_{AB}. In addition, the presence of Cys6-Cys127 facilitates the interactions between the A-helix and the C-terminal 3_{10}-helix, and may still more enhance the effect of preferential hydration by inhibiting expansion of the polypeptide chain. Particularly, the presence of the stable A-helix is likely to be crucial to maintain the configuration among A-, B- and C-helices. This is a reason why the α-domain of 2SS[6-127, 64-80] is stabilized in the presence of concentrated glycerol.

On the other hand, the C-helix properly set between A- and B-helices makes two important contributions to gain interactions between α- and β-domains. One is providing a footing for the pivot of the β-domain to be anchored to the α-domain, and the other is playing a key role in interactions with the β3-strand or the 3_{10}-helix in the β-domain. In the absence of glycerol, the existence of Cys64-Cys80 alone was insufficient to secure the stability of the C-helix and the pivot structure anchored to the α-domain. The preferential hydration by glycerol may enhance the marginal stability of the α-domain without the Cys30-Cys115 and further facilitate the hydrophobic interactions to embed the hydrophobic side chains of Ile55 and Leu56 in the hydrophobic pocket of the α-domain.

X. LATENT STRUCTURES IN UNSTRUCTURED 1SS-VARIANTS

1. Far- and near-UV CD spectra of 1SS-variants

Four species of 1SS-varinats of lysozyme containing each one of native disulfide bridges were constructed and studied by means of CD and NMR spectroscopies. They are 1SS[6-127], 1SS[30-115], 1SS[64-80], and 1SS[76-94]. All of them were unstructured in water, judging from their near-UV CD and ^{1}H-^{15}N-HSQC spectra. Our study showed that the addition of 30% glycerol induced the unstructured 2SS[6-127, 64-80] to form a native-like structure. Why is it observable in a concentrated glycerol solution? It should be noted that the preferential hydration by glycerol did not induce any tertiary structure by itself. Molecules in a native-like conformation might exist in water as a preferential structure

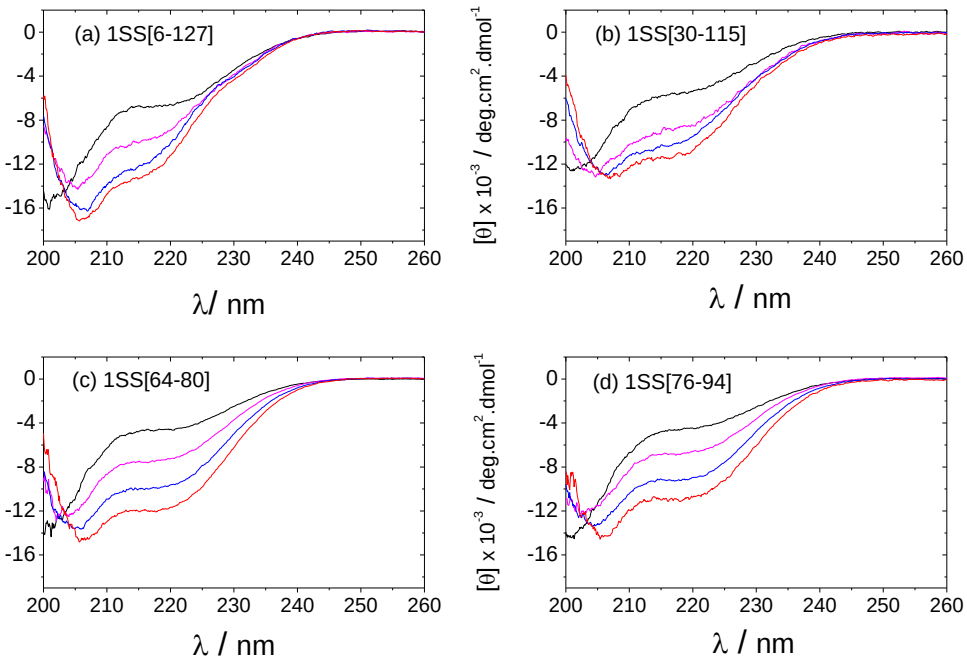

Figure 23. Far-UV CD spectra of 1SS-variants of lysozyme in solutions containing glycerol (22). All spectra were measured at pH 3.0 and 4°C. (a) 1SS[6-127], (b) 1SS[30-115], (c) 1SS[64-80], (d) 1SS[76-94]. In each diagram, glycerol concentrations are 0%, 30%, 50% and 70% in the order of increasing negative intensities.

induced with an assistance of disulfide bridges, but the population of such molecules must have been small enough to escape spectroscopic detection. In the case of 2SS[6-127, 64-80], such a latent structure may have been more stabilized and become just detectable due to the preferential hydration by glycerol. Also in the case of four species of 1SS-variants,

they were almost unstructured in the absence of glycerol, but some compact structure might potentially exist as a preferential structure. We attempted to find such latent structures by the glycerol-enhanced detection.

First, far-UV CD spectra of four species of 1SS-variants were measured, and they are shown in Figure 23. Negative intensities of CD spectra significantly increased in all the variants with the increase in glycerol concentrations from 0% to 70%. In the absence of glycerol, generally the CD spectrum of each 1SS-variant was typical of unstructured lysozyme including 0SS-variant, although values of $[\theta]_{220}$ varied a little. In a solution of 70% glycerol, values of $[\theta]_{220}$ became almost the same for all the variants. It suggests that nearly the same amount of helical content is induced to all the 1SS-variants by the addition of 70% glycerol. The same phenomena were also observed in 2SS- and 0SS-variants.

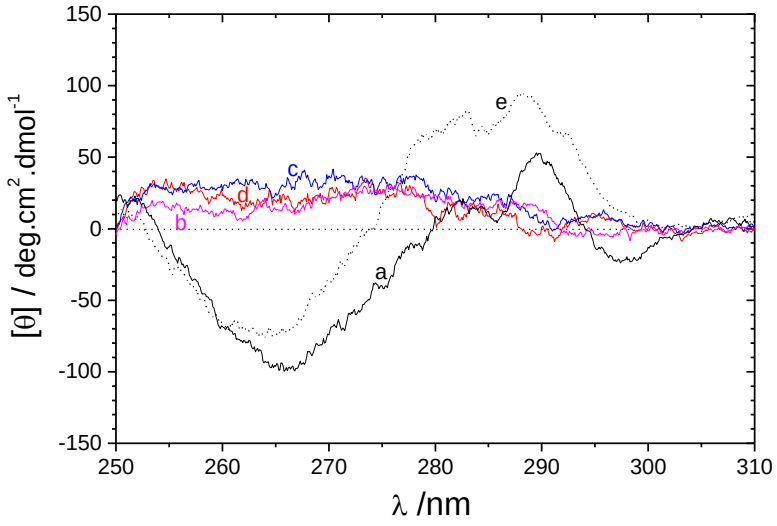

Figure 24. Near-UV CD spectra of 1SS-variants of lysozyme in solutions containing concentrated glycerol (22). All spectra were measured at pH 3.0 and 4°C. (a) 1SS[6-127] in 50% glycerol, (b) 1SS[30-115] in 70% glycerol, (c) 1SS[64-80] in 50% glycerol, (d) 1SS[76-94] in 50% glycerol, (e) the CD spectrum of 2SS[6-127, 64-80] in 50% glycerol is shown for reference.

Second, near-UV CD spectra of four species of 1SS-variants were also measured in solutions containing various concentrations of glycerol. In the absence of glycerol, CD spectra of all the 1SS-variants were nearly equal to zero in the wavelength range of 250 to 310 nm. Figure 24 shows near-UV CD spectra of four species of 1SS-variants in the presence of 50% or 70% glycerol. In cases of 1SS-variants except 1SS[6-127], their CD

intensities were nearly equal to zero at the overall wavelength. Only in the CD spectrum of 1SS[6-127], the negative ellipticity significantly increased near at 265 nm, and the positive one slightly increased near at 290 nm in concentrated glycerol solutions. For comparison, the CD spectrum of 2SS[6-127, 64-80] in the presence of 50% glycerol is shown in Figure 24. Their CD bands near at 265 nm are quite similar to each other between 1SS[6-127] and 2SS[6-127, 64-80]. It is clear that Cys6-Cys127 is responsible for the increase in CD-bands at 265 nm induced by the addition of glycerol, because the near-UV CD spectrum of 2SS[64-80, 76-94] remained nearly equal to zero in a solution of 50% glycerol (see Figure 20). In general, structuring the environment around Phe and Tyr residues and/or a disulfide bridge causes the increase in the negative CD band near at 265 nm (47). In the case of hen lysozyme, most of Phe and Tyr residues are located in the N-terminal half of the α-domain: F3, Y20, Y23, F34 and F38 except for Y53. It is likely that an ordered structure is induced around some of them by the addition of glycerol. On the other hand, the CD intensity near at 290 nm reflecting the environment around Trp residues was not so large in 1SS[6-127] compared with that in 2SS[6-127, 64-80].

2. Protection factors of 1SS-variants

Since CD spectra could not give information at the resolution level of amino acid residues, we attempted to study the glycerol-induced structure by means of H/D exchange experiments of backbone amides combined with NMR measurements. Experimental procedures were the same as those in experiments on 2SS[6-127, 64-80]. Protection factors against H/D exchange were determined for individual residues of 1SS-variants. As shown in Figure 25, they did not exceed ten much more in 1SS-variants except 1SS[6-127]. In all the 1SS-variants, the negative ellipticity at 220 nm, $[\theta]_{220}$, increased in absolute values with the increase in glycerol concentrations, and it approached the same value at 70% glycerol. Despite the high helical contents, the protection factors for H/D exchange reactions remained low (no more than 10) even in a 70 % glycerol solution except for 1SS[6-127]. Helical conformations are induced throughout the polypeptide chain, but there exists no conspicuous area with large protection factors. It is likely that induced helical conformations are short and not persistent such that they float along the polypeptide chain, as well as in 2SS[6-127, 64-80] observed in dilute glycerol solution. On the other hand, in the glycerol-induced structure for 1SS[6-127], there are several residues with a larger protection factor than 100. In such a region, it is likely that some long-range interactions work extra among secondary structures. It is noteworthy that the glycerol-induced structure in 1SS[6-127] is not localized only in the vicinity of Cys6-Cys127, and highly protected regions distribute themselves around A-, B-, C-helices, and

Ile55 and Leu56. The profile of protection factors vs. residue numbers in 1SS[6-127] is similar to that of the glycerol-induced structure in 2SS[6-127, 64-80], although the

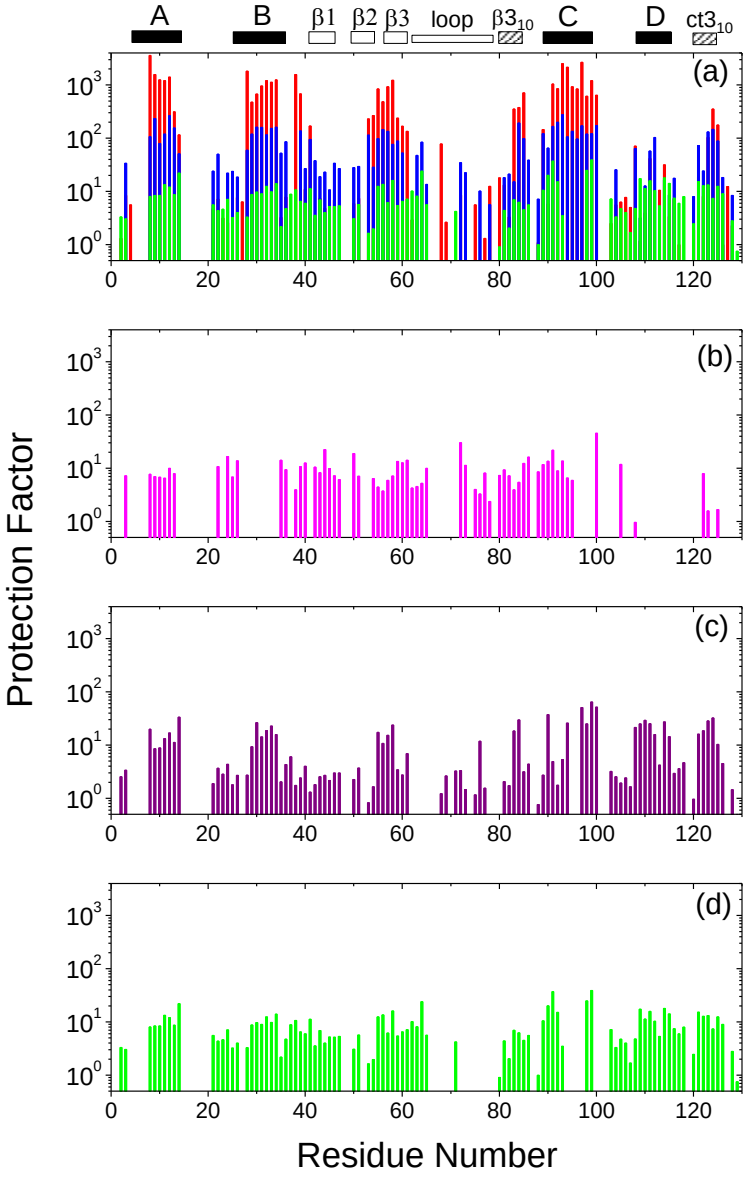

Figure 25. Profiles of protection factors versus the amino acid sequence. The ratio of τ_{prot} / τ_{for} was used as a protection factor, where τ_{prot} and τ_{for} were exchange time constants observed in proteins and those of 0SS-variant observed in formate buffer without glycerol, respectively. (a) Thick lines colored in blue represent protection factors of 1SS[6-127] in 50% glycerol. Extra two thick lines are superposed on them for reference. Red lines stand for protection factors of 2SS[6-127, 64-80] measured in 30% glycerol. Green lines are the same as those shown in the chart (d). (b) Protection factors of 1SS[30-115] in 50% glycerol. Unfortunately, protection factors of several residues in the vicinity of Cys30-Cys115 could not be determined because of too small peak-volumes. (c) Those of 1SS[64-80] in 70% glycerol. (d) Those of 1SS[76-94] in 70% glycerol. The assigned chemical shifts were deposited in the BMRB database with accession numbers of 11459(1SS[6-127]), 11460(1SS[30-115]), 11461(1SS[64-80]) and 11462(1SS[76-94]).

former is less rigid than the latter. This observation implies that Cys6-Cys127 induce a weak tertiary structure into the whole α-domain through the preferential hydration by glycerol. The role of A-helix seems important to recover the long-range interactions in the α-domain, because the amino-acid truncation experiments suggested that the A-helix was indispensable but the C-terminal ten residues were nonessential for the maintenance of the α-domain structure, as mentioned already. Probably, the preferential hydration by glycerol may enhance the hydrophobic interactions in the α-domain as well as in 2SS[6-127, 64-80], although the stability of the C-helix and the structure of the pivot at the interface between α- and β-domains must be greatly weakened due to the loss of Cys64-Cys80.

3. Summary of the transition state of lysozyme folding

We pointed out it was important in elucidation of protein folding to characterize the transition state in the two-state kinetics between N and U, particularly to detect structural features of protein in the transition state. Computer simulations of island-model proteins afforded evidence that the folding pathway was just the reverse process of the unfolding pathway, and a single transition state really existed as a turning point between N and U along the reaction coordinate, and the transition state theory was applicable to the analysis of folding and unfolding rate constants of protein. By examining the change in heat capacity upon the lysozyme unfolding, we derived a conclusion that the transition state of lysozyme was like a "dry molten globule" with diminished non-covalent interactions, but without water inside. Further, we obtained useful information about the transition state of lysozyme folding by evaluating effects on folding and unfolding rates of the modification of specific residues. Later, this study was generalized more and formulated as the φ-value analysis. In summary of the kinetic study of lysozyme, surroundings around residues 64 and 80 remain disordered in the transition state regardless of the presence or absence of Cys64-Cys80, while those around residues 76 and 94 are structured firm such that two residues come close to each other. Further, the D-helix and the C-terminal 3_{10}-helix are unstructured in the transition state, but the interior of the α-domain including A-, B-, and C-helices is folded tightly.

In order to study the structure of lysozyme in the transition state at the atomic resolution, we prepared various types of disulfide-deficient variants of lysozyme as protein samples with a marginal stability, and have studied their three-dimensional structures by means of the NMR spectroscopy. 2SS-variants of lysozyme were on the border line between the folded and unfolded states. Especially, 2SS[6-127, 30-115] variant corresponded to proteins just before the unfolding transition, and 2SS[6-127, 64-

80] just before the folding transition in terms of that the former barely maintained the tertiary structure, and the latter recovered the native-like structure with an assistance of the preferential hydration by glycerol. Besides the information obtained from the marginal structure of these variants, the three-dimensional structures of other 3SS-variants gave important information about the role of each disulfide bridge. From the point of view that the transition state is the limit of structural fluctuations occurring in the native state, characteristics of the dynamic structures of 3SS-variants determined by NMR observations were entirely consistent with those of the transition state obtained from their φ-value analysis.

The detailed structure of lysozyme in the transition state is as follows. In the α-domain, A-, B-, and C-helices exist stably to maintain the configuration for the C-helix properly to be set between A- and B-helices. It is crucial for the stability of the entire α-domain that side-chains of I88, V92, K96 and V99 on one side of the C-helix are buried in the interior of the α-domain. On the contrary, the polypeptide chain near the C-terminal end, including surroundings around the D-helix and the C-terminal 3_{10}-helix, is unstructured. On the other hand, the entire β-domain, including the β-sheet, the loop, and the $\beta 3_{10}$-helix, becomes greatly flexible and the peptide chain from 57 through 75 are disordered even in the presence of Cys64-Cys80, although the peptide chain from 76 through 94 has a strong propensity for the residue 76 to come close to the residue 94. However, β1- and β2-strands still run anti-parallel with a hairpin turn near at Gly49, and the tn55 (Ile55 and Leu56) is anchored to the pocket in the α-domain. The so-called pivot of the β-domain must be barely preserved just before the unfolding transition.

The cooperativity between α- and β-domains must be secured in the transition state to maintain the whole structure of the molecule. The C-helix properly set between A- and B-helices makes two important contributions to gain interactions between α- and β-domains. One is providing a footing for the pivot of the β-domain to be anchored to the α-domain. The other is playing a key role in interactions with the β3-strand and/or the $\beta 3_{10}$-helix. In the transition state, the stability of the pivot of the β-domain is really marginal. If some large-scale structural fluctuation were generated in the region extending from the $\beta 3_{10}$-helix to the C-helix, so that the disruption of the pivot structure happened, it might be a trigger of the global unfolding of lysozyme.

Acknowledgments

I would like to express my special appreciation to Professor Akiyoshi Wada. You were a tremendous mentor for me when I was a graduate student of the University of Tokyo. I would like to thank you for encouraging my research and for leading me to the way of research on protein folding. Also, I would like to thank former members of Dr. Wada's laboratory, particularly Dr. Yuzuru Husimi for serving as my research adviser, and for your brilliant comments and suggestions. After I started to work at Kwansei Gakuin University as a lecturer in School of Science, I could spend my time on studies of protein folding with many graduate students and some research assistants. I would like to thank you for working together as coworkers of my study. All of you have supported me to make my study a great progress. Especially I am deeply grateful to Dr. Yasuo Noda for serving as an adviser to graduate students in my laboratory. For a long time, you have been my best colleague. When I attempted to study the folding of various derivatives of lysozyme, I proposed a joint research of disulfide-deficient variants of lysozyme to Dr. Hideki Tachibana at Kobe University. I deeply thank Dr. Tachibana for gladly agreeing to it. You kindly provided us with the expression system of recombinant lysozyme derivatives, and I thank you for your proper advices on preparation of protein samples. I would also like to thank all of my friends who have worked hard together at Kwansei Gakuin University, particularly thank Professor Koji Yoshimitsu and Professor Yasushi Koyama. Your comments upon our research encouraged me to investigate the protein folding problem more deeply. At the end I would like to express appreciation to my wife Keiko Segawa. You have always supported me with dedication.

References

1. Phillips, D.C. (1966) *Sci. Am.* **215**, 878-890.
2. Blake, C.C., Koenig, D.F., Mair, G.A., North, A.C., Phillips, D.C. and Sarma, V.R. (1965) *Nature* **206**, 757-761.
3. Johnson, L.N. (1998) *Nature struct. biol.* **5**, 942-944.
4. Segawa, S., Husimi, Y. and Wada, A. (1973) *Biopolymers* **12**, 2521-2537.
5. Husimi, Y. and Segawa, S. (1974) *Japan. J. Appl. Phys.* **13**, 1151-1157.
6. Eigen, M. and DeMaeyer, L. (1963) in *Techniques of Organic Chemistry*, Friess, S.L., Lewis, E.S. and Weissberger, A. Eds. (Wiley, New York), vol. **8**, part 2, pp. 895-1054.
7. Eigen, M. (1971) *Naturwissenschaften* **58**, 465-523.
8. Eigen, M. and Schuster, P. (1977) *Naturwissenschaften* **64**, 541-565.
9. Eigen, M. and Schuster, P. (1978) *Naturwissenschaften* **65,** 7-41, 341-369.
10. Segawa, S. and Kawai, T. (1986) *Biopolymers* **25**, 1815-1835.
11. Richards, F.M. (1992) in *Protein Folding*, Creighton, T.E. Ed. (Freeman, New York), pp. 1-58; Creighton, T.E. (1994) in *Mechanisms of Protein Folding*, Pain, R.H. Ed. (Oxford, New York), pp. 1-25.
12. Ptitsyn, O. B. (1992) in *Protein Folding*, Creighton, T.E. Ed. (Freeman, New York), pp. 243-300; Kuwajima, K. (1989) *Proteins: Struct. Funct. Genet.* **6**, 87-103.
13. Kim, P.S. and Baldwin, R.L. (1990) *Annu. Rev. Biochem.* **59**, 631-660; Christensen, H. and Pain, R.H. (1994) in *Mechanisms of Protein Folding*, Pain, R.H. Ed. (Oxford, New York), pp. 55-79
14. Ohgushi, M. and Wada, A. (1983) *FEBS Lett*. **164**, 21-24.
15. 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.
16. Segawa, S. and Sugihara, M. (1984) *Biopolymers* **23**, 2473-2488.
17. Segawa, S. and Sugihara, M. (1984) *Biopolymers* **23**, 2489-2498.
18. Yokota, A., Izutani, K., Takai, M., Kubo, Y., Noda, Y., Koumoto, Y., Tachibana, H. and Segawa, S. (2000) *J. Mol. Biol.* **295**, 1275-1288.
19. Noda, Y., Yokota, A., Horii, D., Tominaga, T., Tanisaka, Y., Tachibana, H. and Segawa, S. (2002) *Biochemistry* **41**(2002), 2130-2139.
20. Yokota, A., Hirai, K., Miyauchi, H., Iimura, S., Noda, Y., Inoue, K., Akasaka, K., Tachibana, H. and Segawa, S. (2004) *Biochemistry*, **43**, 6663-6669.
21. Sakamoto, K., Hirai, K., Kitamura, Y., Yamazaki, K., Yusa, M., Tokunaga, N., Doi, G., Noda, Y., Tachibana, H. and Segawa, S. (2009) *Biopolymers* **91**, 665-675.

22. Noda, Y., Narama, K., Kasai, K., Tachibana, H. and Segawa, S. (2012) *Biopolymers*, **97**, 539-549.
23. Segawa, S. and Richards, F.M. (1988) *Biopolymers* **27**, 23-40.
24. Yokota, A., Takenaka, H., Oh, T., Noda, Y. and Segawa, S. (1998) *Protein Science*, **7**, 1717-1727.
25. Plaxco, K.W, Simons, K.T. and Baker, D. (1998) *J. Mol Biol.* **277**, 985-994.
26. Go, N., Abe, H., Mizuno, H. and Taketomi, H. (1980) in *Protein Folding*, Jaenicke, R. Ed. (Elsevier/North Holland Biomedical Press, Amsterdam), pp. 167-181.
27. Wako, H. and Saito, N. (1978) *J. Phys. Soc. Jpn.* **44**, 1931-1938.
28. Privalov, P.L. (1979) *Adv. Protein Chem.* **33**, 167-241; Privalov, P.L. (1992) in *Protein Folding*, Creighton, T.E. Ed. (Freeman, New York), pp. 83-126.
29. Karplus, M. and Shakhnovich, E. (1992) in *Protein Folding*, Creighton, T.E. Ed. (Freeman, New York), pp. 127-195; Baldwin, R.L., Frieden, C. and Rose, G.D. (2010) *Proteins: Struct. Funct. Bioinfor.* **78**, 2725-2737.
30. Beddel, C.R., Blake, C.C.F. and Oatley, S.J. (1975) *J. Mol. Biol.* **97**, 643-654.
31. Matouschek, A., Kellis, J.T., Jr, Serrano, L and Fersht, A.R. (1989) *Nature* **340**, 122-126.
32. Matouschek, A., Serrano, L. and Fersht, A.R. (1994) in *Mechanisms of Protein Folding*, Pain, R.H. Ed. (Oxford, New York), pp. 137-159.
33. Tachibana, H., Ohta, K., Sawano, H., Koumoto, Y. and Segawa, S. (1994) *Biochemistry* **33**, 15008-15016.
34. Tachibana, H., Oka, T. and Akasaka, K. (2001) *J. Mol. Biol.* **314**, 311-320.
35. Hill, C.P., Johnston, N.L. and Cohen, R.E. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 4136-4140.
36. Cavanagh, J., Fairbrother, W.J., Palmer, A.G., III, Rance, M. and Skelton, N.J. (1996) *Protein NMR spectroscopy: Principles and Practice*, Second Edition, (Academic Press, San Diego, CA).
37. Wüthrich, K. (1986) *NMR of proteins and Nucleic Acids*, (Wiley, New York).
38. Englander, S.W. and Mayne, L. (1992) *Ann. Rev. Biophys. Biomol. Struct*. **21**, 243-265.
39. Bai, Y., Milne, J.S., Mayne, L. and Englander, S.W. (1993) *Proteins: Struct. Funct. Genet.* **17**, 75-86.
40. Miyauchi, H. (2001) Master's Thesis, Kwansei Gakuin Univ.
41. Yamazaki, K. (2005) Master's Thesis, Kwansei Gakuin Univ.
42. Nio, T. (1997) Master's Thesis, Kwansei Gakuin Univ.

43. Zhang, Y., Paterson, Y. and Roder, H. (1995) *Protein Sci.* **4**, 804-814.
44. Gekko, K. and Timasheff, S. (1981) *Biochemistry* **20**, 4667-4676.
45. Lehmann, M.S. and Zaccai, G. (1984) *Biochemistry* **23**, 1939-1942.
46. Sawano, H., Koumoto, Y., Ohta, K., Sasaki, Y., Segawa, S. and Tachibana, H. (1992) *FEBS Lett.* **303**, 11-14.
47. Manning, M.C. and Woody, R.W. (1989) *Biochemistry* **28**, 8609-8613.