Theoretical Study on Relaxation Processes of Biopolymers in Aqueous Solution

Masayoshi Takayanagi

Contents

Chapter 1

General Introduction

1.1. Introduction

All of the things in the world are made from a great variety of molecules that are composed of atoms. It is true that, in principle, the fundamental quantum-mechanical dynamics of such various molecules could be calculated by the time-dependent Schrödinger equation. However, if their quantum characteristics might be negligible, the dynamics could be described directly by solving simultaneous Newtonian equations of motion under the classical approximation, which drastically reduces the complexity of the equations. Moreover, for macroscopic systems in equilibrium that consist of an immense number of molecules $\sim 10^{23}$, there exist some theories established to treat them, i.e. thermodynamics and statistical mechanics, in which the macroscopic properties of the systems can be represented by only a small number of thermodynamic quantities (temperature, entropy, pressure, volume, and so forth). Although a number of theoretical schemes for calculating the molecular behaviors have been developed so far, there exist a few exact analytic solutions which could be obtained only for ideal or simple systems in equilibrium. This is the reason why numerical calculations should be carried out, inevitably for realistic complex systems, especially for biopolymers.

Actually, to avoid an extremely long computational time for solving quantum-mechanical equations, many theoretical studies have been carried out with classical molecular dynamics (MD) simulations in which each atom is considered as a point mass, assuming simple interaction functions among them. Further, instead of dealing with a huge number of molecules in macroscopic systems, it is usual to adopt the periodic boundary condition in the MD simulations at the same time. In addition, due to the incredible progress of computational resources in recent decades, a large-scale MD simulations have become possible to be applied to a protein molecule

in aqueous solution that consists of ~10,000 atoms. Therefore, in this thesis, I theoretically analyze an aqueous solution system including both a protein and many water molecules by using MD simulations under the periodic boundary condition.

1.2. Understanding Protein Functions at the Atomic Level

As far as the living things are concerned, one could say that they are composed of biomacromolecules such as DNA and proteins. Since these molecular species are, of course, composed of atoms, it is true that theoretical schemes and equations mentioned in the Introduction can be also applied to understand their functions from the microscopic level.

So far, a lot of scientists have devoted considerable effort to understanding the biological molecules. Starting with the discovery of the double-helical structure of DNA [1] that contains the genetic code, they have explored the genetic information of various living creatures and, finally, the whole sequence of human DNA (i.e. human genome) was clarified in the current century [2].

However, in order to understand biological processes, it is absolutely necessary to investigate not only DNA sequences but also proteins whose amino acid sequences are determined by genetic codes kept in DNA. With regard to the structure of proteins, which is very important for it to function, the first determination was done for myoglobin (Mb) using X-ray crystallography in the late 1950s [3]. Thereafter, various protein structures have been determined due to the progress of experimental techniques, such as X-ray crystallography, NMR, neutron diffraction, and so on. It is quite recently that the full structure of ribosome, which is a very large complex of ribosome RNAs and proteins, was reported [4], showing that experimental methods to

determine the static structures of biomacromolecules is now very efficient.

On the other hand, from the viewpoint of theoretical studies, a computational prediction of the folded structure from its amino-acid sequence is still a very difficult task because of two major difficulties. The first one is due to the long time scale of milliseconds or more required for the folding of a standard protein. Since, according to Anfinsen's dogma [5], a folded structure of protein depends only on its primary sequence, one could, in principle, predict its structure theoretically. However, a straightforward prediction with present full-atomic MD calculations is impossible because of the time scale computable usually at most up to microseconds that is far shorter than the required folding time, milliseconds.

The second difficulty is related to the fact that such other molecules as chaperones and ribosomes are usually involved in the in vivo folding process. It was in fact reported that the folding process in vivo requires necessarily the action of chaperones [6] to prevent misfolding and aggregation. In addition, the "cotranslational folding" [7,8], namely, such a folding that ribosome-bound N-terminal nascent peptide chains would begin to fold during the elongation (Fig. 1.1), is also supposed to play a key role in the incipient folding process. These make the prediction problem more complicated and difficult.

Furthermore, even if the static structure of a biomacromolecule might be obtained, its functionally important structural dynamics is not completely explained from the thermodynamical point of view. For instance, hemoglobin (Hb), a tetrameric oxygen transport protein whose subunit each has a heme group to bind an oxygen molecule, regulates its affinity for oxygen by changing the quaternary structures: a high affinity R-state and a low affinity T-state [9]. The quaternary changes are triggered by the ligand binding and dissociation. While static crystal structures of T-

and R-states are experimentally known, it is not clear yet how and why the atomistic dynamics of the quaternary changes might proceed.

1.3. Myoglobin: a Model Protein for Understanding Nonequilibrium Dynamics

In this thesis, Mb is used as a model protein for understanding the nonequilibrium dynamics which is important for the function of Mb or other globin proteins. Mb is a monomeric oxygen-storage protein in muscle tissues, containing a heme group as a prosthetic group. The heme can bind a small ligand molecule such as O_2, CO, NO, and so on as the sixth ligand to its centrally-located iron atom. As noted in the previous section, Mb is the first protein whose X-ray crystal structure was determined and a number of subsequent experimental and theoretical studies were thoroughly conducted, meaning that it is relatively easy to verify the validity of the computational results by comparing them with the experiments.

Among them, there are several experiments measuring the properties of N-terminal nascent chains by using Mb or Hb α-subunit, whose folded structure is very similar to that of Mb. For instance, Komar et al. found that the ribosome-bound N-terminal α-globin whose length is longer than 86 amino-acid residues can efficiently bind heme [10]. On the other hand, far-UV circular dichroism (CD) spectra measurements revealed that α-helix ratios of N-terminal fragments of Mb depend on their lengths: the longer chains favor α-helix [11]. These facts imply that Mb might experience cotranslational folding during its synthesis. To investigate the cotranslational folding of Mb, nascent Mb chains were studied in chapter 2 by using MD simulations.

In fact, the cotranslational folding process can be assumed as a nonequilibrium

process. It is well known that the ribosome adds an amino acid to the C-terminal side of the elongating nascent peptide by using a mRNA as a template (Fig. 1.1). Considering the fact that the protein synthesis rate (several tens of milliseconds per residue) is large enough to sample a wide range of conformational space, it is expected that the N-terminal nascent chains would fold into their stable conformation depending on their lengths. That is to say, when a new residue is added to the nascent chain, its conformation would change to adjust its new length. Accordingly, this time-depending behavior of cotranslational folding can be regarded as a nonequilibrium phenomena that proceeds in a long time regime (more than milliseconds).

Incidentally, Mb is often utilized as a model system suitable to investigate relaxation processes of biopolymers. In particular, the photolysis of carbonmonoxy myoglobin (MbCO) triggered by the visible photon absorption of the heme group is one of the most thoroughly studied phenomena (Fig. 1.2). Although the most important ligand responsible to the in vivo function is an oxygen molecule, its photolysis from the heme is accompanied by the geminate ligand rebinding within a few nanoseconds which makes the data analysis and interpretation complicated. On the other hand, an adequately slow rate of the geminate rebinding of MbCO (more than a few microseconds) makes it possible to measure purely dissociated states on a picosecond or nanosecond time-scale. Hence, with these advantageous characteristics, many experimental studies have been carried out using MbCO.

The photodissociation process of MbCO induces not only the ligand dissociation but also the heme group structural changes with vibrational excitation. Immediately after the photolysis, the heme alters its shape from the six-coordinated planar structure to the five-coordinated domed one, while, a part of the absorbed photon energy is

injected into the heme as excess vibrational energy or “heat”. Then, the changes propagate from the heme to the surrounding globin and solvent matrices. Since the nonequilibrium relaxation process is brought about involving structural deformations with vibrational energy diffusion, the photolysis can be considered as a perturbation for the aqueous solution of Mb. It is, therefore, anticipated that the relaxation process triggered by the perturbation should be investigated to shed light on the nonequilibrium dynamics of Mb.

In addition, the photodissociation process of MbCO is also a typical example in which structural changes play a significant role in the functions of the same globin family. Actually, it is well known that Mb is very similar to each subunit of Hb and its allostery is also induced by the ligation state of the heme (the quaternary structural changes between T- and R-states as described previously in section 1.2). Thus, rich information would be provided through the detailed analyses of the MbCO photolysis, on the microscopic mechanism of biological functions of Hb. In fact, by comparing the crystal structural differences among Mb and Hb subunits or between their deoxy and CO-ligated forms, Guallar et al. revealed that the ligand dissociation triggers the displacements in common E and F helices, i.e. the EF clamshell rotation, in both proteins [12]. This motion is considered to play an essential role in the T-R quaternary changes because the F-helix is located in the interface region between the α and β subunits [13]. With a final goal to understand the functionally important process, as a first step, I analyzed the photodissociation process of MbCO by executing MD calculations in chapter 3.

1.4. Perturbation Ensemble Method

The main purpose of this thesis is to understand the nonequilibrium behaviors of biopolymers from the microscopic point of view. For this end, MD calculations of

Mb were executed and analyzed. To compare the computational results of MD calculations with the experimental ones, it is expected that physical quantities must be estimated from the obtained MD trajectories with a statistically significant precision.

Usually, for equilibrium systems, the procedure to calculate some quantity A is based on the time averaging over an enough long MD simulation as

$$\overline{A} = \frac{1}{\tau_{\mathrm{run}}} \sum_{i=1}^{\tau_{\mathrm{run}}} A(i) \tag{1.1}$$

where τ_{run} is a total simulation time step. However, this procedure cannot be applied to the relaxation processes of the photodissociated MbCO which I treat in chapter 3, because it shows nonequilibrium, or time-dependent characteristics. In addition, the structural change brought about by the ligand dissociation are smaller than the thermal fluctuations at room temperature and thus the analysis of just a single MD trajectory yields little information. Therefore, I adopt an alternative approach in which many MD calculations, or a set of ensemble MD (EMD) simulations, are executed for the ensemble average. In this procedure, the quantity $A^i(t)$, the value $A(t)$ of ith MD trajectory at time t, was ensemble averaged over N_{MD} trajectories to calculate the time-dependent average as

$$\langle A(t) \rangle = \frac{1}{N_{\mathrm{MD}}} \sum_{i=1}^{N_{\mathrm{MD}}} A^i(t) \tag{1.2}$$

If the system might be in equilibrium and both τ_{run} and N_{MD} might be enough large, the averages calculated from the Eq. (1.1) and (1.2) would become identical due to the ergodic hypothesis. However, for nonequilibrium processes, only the latter one is applicable to obtain time-dependent values.

Although the time-dependent properties can be calculated by Eq. (1.2), statistical

errors are inevitable due to the limitation of a number of MD simulations with current computational resources. In addition, even if the number of MD simulations is large enough to cancel out the thermal fluctuations, biased initial ensemble of the MD calculations would contaminate the statistical averages. To avoid this problem, a subtraction procedure between perturbed MD (PMD) and unperturbed MD (UMD) was introduced. In the procedure, it should be noted that a pair of PMD and UMD is executed from the identical initial atomic coordinates and velocities for all the atoms except that only the former receives a perturbation at $t = 0$. After executing many (more than several hundreds) pairs of PMD and UMD, or a pair of perturbed ensemble MD (PEMD) and unperturbed ensemble MD (UEMD), the ensemble average over the differences between PMD and UMD is then expressed as

$$\begin{aligned}\left\langle \delta\Delta A(t)\right\rangle &= \frac{1}{N_{\mathrm{MD}}}\sum_{i=1}^{N_{\mathrm{MD}}}\delta\Delta A^{\mathrm{PMD},i}(t) \\ &= \frac{1}{N_{\mathrm{MD}}}\sum_{i=1}^{N_{\mathrm{MD}}}\left\{\left(A^{\mathrm{PMD},i}(t)-A^{\mathrm{PMD},i}(0)\right)-\left(A^{\mathrm{UMD},i}(t)-A^{\mathrm{UMD},i}(0)\right)\right\}\end{aligned} \tag{1.3}$$

where $A^{\mathrm{PMD},i}(t)$ and $A^{\mathrm{UMD},i}(t)$ are the quantity A of ith PMD and UMD trajectories at time t respectively. I call this procedure as the *perturbation ensemble* (*PE*) *method*.

In fact, it is true that this "subtraction technique" itself was suggested more than 20 years ago [14]. However, because of the limitation of computational resources at that time, the target system size, the number of PMD and UMD calculations, and their simulation lengths were very restricted. On the contrary, now, with the rapid development of computers, one can calculate several hundreds or even several thousands of MD trajectory calculations including more than ten thousand atoms, which is actually the system size of a small protein like Mb in aqueous solution, for 100 ps time duration.

The PE method was applied for the first time to investigate the photolysis process of MbCO in chapter 3, in which 600 pairs of PMD and UMD trajectories were executed and statistically analyzed. As a consequence, an anisotropic structural expansion of photolyzed MbCO was clearly detected. This is consistent with a corresponding experiment showing the same anisotropic change after the photolysis. In conclusion, it was verified that such a large ensemble calculation of the PE method is statistically meaningful enough and it can reproduce the experimental results reasonably.

1.5. Outline

My main interest is in the nonequilibrium properties of biomacromolecules in aqueous solution. Using Mb as a model system, I have investigated not only a short-time (~100 ps) response of photodissociated MbCO but also a relatively long-time (>1 ms) structural changes of ribosome-bound nascent apoMb chains during elongation. The former includes not only structural deformations of Mb but also thermal diffusion from the heme to the surrounding globin and solvent. The latter is the sequential formation of the folded structure along with the nascent chain elongation. These nonequilibrium processes should play significant roles in their functions and folding itself.

In the present thesis, I investigated Mb by using MD simulations. In chapter 2, the stabilities of N-terminal partial chains from apoMb were investigated by executing 10 ns simulations to elucidate the behaviors of the nascent chains during the elongation. In chapter 3, the overall shape changes of photodissociated MbCO within 100 ps time duration were investigated by calculating three Cartesian components of radius of gyration. To explore the subtle changes smaller than the thermal fluctuations, I formulated and used the PE method in which many pairs of PMD and UMD

calculations were executed for statistical analysis. In chapter 4, the general conclusion was provided including the future perspectives.

References

[1] J.C. Kendrew, G. Bodo, H.M. Dintzis, R.G. Parrish, H. Wyckoff, D.C. Phillips, Nature 181 (1958) 662.

[2] M.M. Yusupov, G.Zh. Yusupova, A. Baucom, K. Lieberman, T.N. Earnest, J.H.D. Cate, H.F. Noller, Science 292 (2001) 883.

[3] (a) J.D. Watson, F.H.C.Crick, Nature 171 (1953) 740; (b) R. Franklin, R.G. Gosling, Nature 171 (1953) 740.

[4] (a) International Human Genome Sequencnig Consortium, Nature 409 (2001) 860; (b) J.C. Venter, et al., Science, 291 (2001) 1304; (c) International Human Genome Sequencing Consortium, Nature 431 (2004) 931.

[5] (a) J. Baldwin, C. Chothia, J. Mol. Biol. 129 (1979) 175; (b) M.F. Perutz, Nature 228 (1970) 726.

[6] (a) F.H. White Jr., J. Biol. Chem. 236 (1961) 1353; (b) C.B. Anfinsen, E. Haber, J. Mol. Biol. 129 (1961) 1361; (c) C.B. Anfinsen, Science 181 (1973) 223.

[7] G. Kramer, V. Ramachandiran, B. Hardesty, Int. J. Biochem. Cell. Biol. 33 (2001) 541.

[8] V.A. Kolb, Mol. Biol. 35 (2001) 682.

[9] R.J. Ellis, S.M. van der Vies, Annu. Rev. Biochem. 60 (1991) 321.

[10] M.P. Allen, C.J. Tildesley, Computer Simulation of Liquids, Oxford Science Publications, New York, 1987.

[11] A.A. Komar, A. Kommer, I.A. Krasheninnikov, A.S. Spirin, J. Bio. Chem. 272 (1997) 10646.

[12] C.C. Chow, C. Chow, V. Raghunathan, T.J. Huppert, E.B. Kimball, S. Cavagnero, Biochemistry 42 (2003) 7090.

[13] M.F. Perutz, A.F. Wilkinson, M. Paoli, G.G. Dodson, Annu. Rev. Biophys. Biomol. Struct. 27 (1998) 1.

[14] V. Guallar, A.A. Jarzecki, R.A. Friesner, T.G. Spiro, J. Am. Chem. Soc. 128 (2006) 5427.

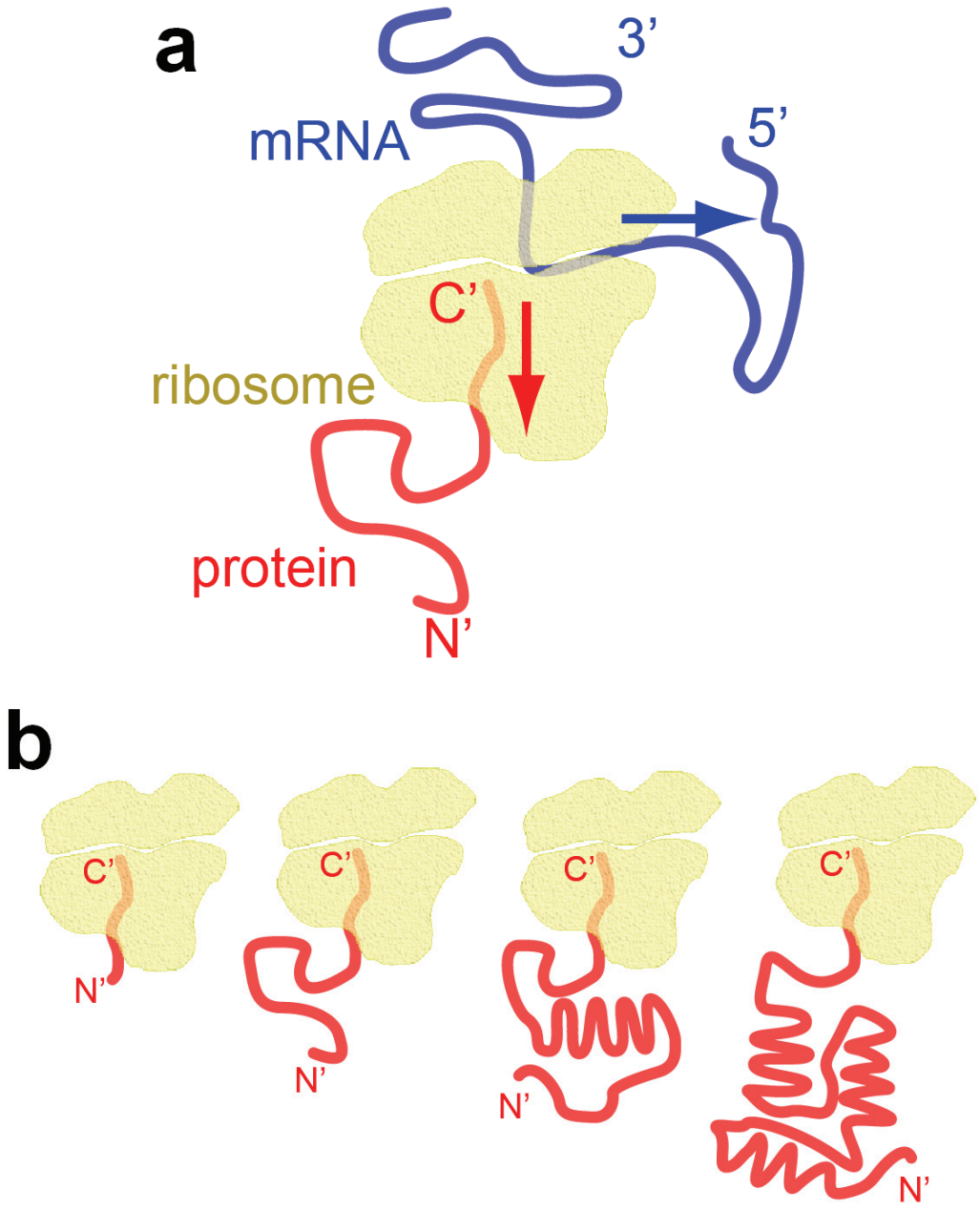

Fig.1.1. (a) Mechanism of the translation. A ribosome reads and decodes genetic codes from messenger RNAs (mRNAs) and synthesizes proteins. (b) Schematic view of the cotranslational folding. During the protein elongation, the nascent peptide is extruded into cytoplasm and starts to fold. It is supposed that the folded structures would change depending on the chain length.

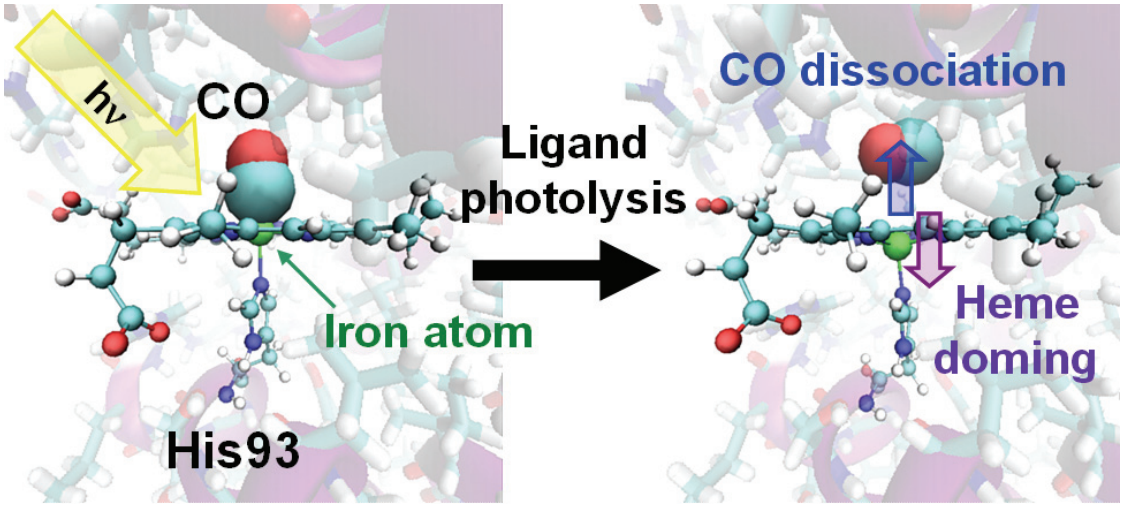

Fig.1.2. Ligand photolysis of MbCO. By absorbing visible light, the CO ligand dissociates from the heme group and, at the same time, the heme changes its structure from planar to domed.

Chapter 2

Theoretical Study on the Stabilities of N-terminal Partial Chains from Apo-myoglobin

2.1. Introduction

A fragment in ribosome-bound N-terminal α-globin (a subunit of hemoglobin), which has amino acid residues longer than 86, is capable of efficient heme binding according to an experimental study of Komar et al. [1]. This indicates that the ribosome-bound nascent chain of 86 amino acid residues acquires such a steric structure that allows its interaction with the heme group. Though this feature suggests the cotranslational folding of the globin, the mechanism is still unknown.

In a study of apo-myoglobin (Mb) N-terminal fragments of 36, 77 and 119 amino acid residues, Chow et al. [2] demonstrated by far-UV circular dichroism (CD) spectra that the shorter fragment enriches β-sheet while the longer one favors α-helix. These folding/misfolding motifs are significantly different from those observed on full-length apo-Mb refolding from a urea unfolded state, where only α-helical conformation is observed [3].

In the present chapter, I report on the molecular dynamics (MD) calculations on the stability of peptides derived from the N-terminal portion of Mb, a well-known protein similar to α-globin (Fig. 2.1). The difference between Mb and α-globin is that Mb folds into 8 α-helices, that are alphabetically named A to H, whereas α-globin folds into 7 α-helices with deletion of the D helix and shortening of the A- and H-helices of Mb. From the analogy of these proteins, it was supposed that Mb also exhibits heme binding like α-globin during elongation and forms native-like partial structures. Thus this study aims to shed light on the structural behavior of nascent Mb chains.

2.2. Computational Methods

MD simulations of four N-terminal partial peptides and a full-length peptide of Mb were performed. The chain lengths were 36, 77, 92, 119, and 153 (full-length Mb). They are denoted by numbers: Mb36, etc. All the MD calculations were carried out with AMBER7 program [4] using the parm99 force field [5]. The SHAKE method was used to restrain the hydrogen-heavy atom bond distances, and the particle mesh Ewald procedure was used to handle long-range electrostatic interactions. The integration time-step was 2 fs. All simulations were performed using periodic boundary conditions. The average temperature and the pressure of the whole system were maintained at 300 K and 1 atm by using the weak-coupling algorithm [6].

The protein structure was retrieved from the PDB, entry 104M, the crystal structure of sperm whale Mb. The initial coordinates for the nascent apo-Mb peptides were taken from the crystal structure with deletion of the heme and the C-terminal amino acid residues. Each was solvated in a rectangular box ($\approx 52\times47\times47$ Å^3), which included ≈ 3000 TIP3P water molecules and neutralized by sodium or chloride ions. The equilibration runs were executed with the structure restraining on the peptide for 200 ps to relax the solvent molecules. Next, the production MD runs were done for 1 ns (Mb36) and 2 ns (others). To adjust the water box size to the structural changes of the chains after the production MD runs, water molecules and counterions were removed and the peptide was re-solvated with water and counterions. Then a series of the equilibration run, the production MD run, and the re-solvation procedure were repeated 10 (Mb36), 5 (Mb77, Mb92, and Mb119) or 3 (Mb153) times. Because of the large structural fluctuations, such a short production run time for Mb36 was needed to update the solvent box size at short intervals. Finally, the total of 10.0 ns (Mb36, Mb77, Mb92, and Mb119) or 6.0 ns (Mb153) unconstrained trajectories were calculated.

The free energies of peptides at each snapshot were estimated by using a method based on the weighted solvent accessible surface area (WSAS) [7]. Then, the free energy of each snapshot G_1 is expressed [8] by

$$G_1 = E_{\mathrm{MM}} + \Delta G_{\mathrm{WSAS}}, \tag{2.1}$$

where E_{MM} is the internal energy of the solute peptide and ΔG_{WSAS} is the solvation free energy. The latter is evaluated by

$$\Delta G_{\mathrm{WSAS}} = \sum_{i=1}^{m} \sum_{j=1}^{n_i} w_i S_j, \tag{2.2}$$

where m, n_i and w_i are the numbers of atom types, atoms with type i and its weight, respectively, and S_j is the solvent accessible surface area (SASA) of atom j. SASA calculations were done using the MSMS program [9].

All structures of chains in this work were drawn with VMD [10].

2.3. Results and Discussion

2.3.1. Instability of the Short Chains

The structural stability of the crystal structure of Mb was evaluated by plotting the root–mean-square deviation (RMSD) from the starting structure and the radius of gyration, R_g, for each chain in Fig. 2.2 as functions of time.

The small RMSD values of Mb153 and Mb119 imply that most of the crystal structure is retained. In contrast, the large values of other short chains, Mb92, Mb77 and Mb36, show that the initial structures substantially collapse within 10 ns. Especially for Mb36, it shows a quick and large increase followed by strong fluctuations.

For Mb153, R_g is stable around 15.4 Å during the simulation. This agrees well with that in the MD calculation of intact apo-Mb [11]. Those of Mb119, Mb92, and Mb77 show the same tendency: They first decrease to 14.3 Å for Mb119, 13.5 Å for Mb92 and 12.8 Å for Mb77 and then fluctuate around the values. On the contrary, for Mb36, the R_g fluctuations are remarkable from the onset.

As a general trend, the stability of the crystal structure decreases with a decrease in the number of residues; this implies that the C-terminal portion stabilizes the crystal structure of the rest of the N-terminal portion.

2.3.2. Stabilities of Secondary Structures

A secondary structure analysis was performed to investigate the stability of helices of the nascent chains. The criteria for the determination were derived from the paper by Kabsch and Sander [12]. The secondary structures are plotted in Fig. 2.3 as a function of time.

Mb36 shows that two helices A and B are almost fully unfolded (Fig. 2.3(a)). For Mb77 (Fig. 2.3(b)), A and B-helix show gradual unfolding, while other helices are well folded during the simulation. For Mb92 and Mb119 (Figs. 2.3(c) and 2.3(d)), all the helices except for A-helix are appreciably stable. For apo-Mb, Mb153 (Fig. 2.3(e)), all the helices are altogether stable, as expected.

The ratios of the 4-turn and 3-turn helix structures in simulations for the last 2 ns of each chain and the corresponding helix ratios of CD spectra data are listed in Table 2.1. As shown in Fig. 2.3 and Table 2.1, a longer peptide is more helical in its secondary structure. However, the current MD calculations for short partial Mb chains, Mb36 and Mb77, exhibit no β-sheet structure, whereas the CD spectra [2] indicated rich β-sheet structures. This is because β-sheet structures are formed

between two chains in experiments, but the current simulations are limited to a single chain. This difference being taken into account, obtained results agree well with the experimental features [2].

2.3.3. Tertiary Structures of Partial Chains

Tertiary structures are analyzed in Fig. 2.4, where snapshots of the nascent chains are visualized at 0, 2, 5, and 10 ns in ribbon drawings from the N-terminal (blue) to the C-terminal (red). The tertiary structural behaviors of Mb36 and other chains are found to be significantly different. Previous MD simulations of short peptides in water [13] demonstrated that peptides could form turn-like structures, which rapidly disappeared and reformed in a nanosecond time scale. Similarly, the shortest peptide Mb36 was also observed to show large fluctuations of secondary and tertiary structures (Fig. 2.4 (a)). On the other hand, in the simulations of longer chains (Fig. 2.4 (b)-(d)) α-helix structures were retained except for A-helix. This instability observed for A-helix seems to contradict with the previous MD calculations for solvated isolated Mb helices, where only F-helix unfolds within 1 ns at 298 K while A-, G-, and H-helices are most stable [14]. However, this inconsistency can be attributed to the absence of A-helix contacts with G and H-helix, which are missing in these partial chains; this causes formation of A-helix interactions with other helices, leading to the uncoiling of A-helix.

2.3.4. Contributions of Free Energy to the Dynamics of Partial Chains

Table 2.2 lists the results of calculations of the free energy, G_1 in Eq. (2.1), averaged over the first and last 1 ns for each simulation. All of them exhibit

increases in the internal energy E_{MM} and, in contrast, decreases in the solvation free energy ΔG_{WSAS}. These variations can be attributed to helix uncoiling, inducing the breakage of hydrogen bonds, which contribute to an increase in E_{MM}. Thus, several amide and carbonyl groups of the main chain, which are hydrophilic and negatively contribute to the increase in the solvation free energy, are newly exposed to the solvent.

As for the free energy, its numerical values vary for each partial chain. While G_1 increases for both Mb153 and Mb36, its initial value is retained for Mb77, and is lowered for longer chains, Mb119 and Mb92. Since a native structure should have a minimum value in the real free energy, a small increase in G_1 for Mb153 ($\approx 0.5\%$) falls within the fluctuation in enthalpy $\approx \delta H = (k_{\mathrm{B}} T^2 C_p)^{1/2} =$ 28.3 kcal/mol (with C_p assumed to be 4.5 kcal/mol/K [15]) and is deemed negligible. However, these initial structures of other partial chains are likely to differ from their most stable or "native" structures; hence, they must in principle undergo certain conformational changes to lower their free energies toward their "native" structures.

For partial chains, especially for smaller ones, one should take explicit account of the contribution of the conformational entropy ΔS_{conf} associated with the peptide structural fluctuations in addition to G_1, which represents the calculated free energy of each snapshot. Then the real free energy change, ΔG, is expressed by

$$\Delta G = \Delta G_1 - T\Delta S_{\mathrm{conf}} . \tag{2.3}$$

In fact, the unfolding of helices enhances the conformational fluctuations and then increases the number of conformational states, so that the resultant positive ΔS_{conf} should lower the total free energy of the partial chains. Thus, for Mb77, Mb92, and

Mb119, all the real free energy changes including the contributions of ΔS_{conf}, would surely exhibit the clear decreases.

For Mb36, it is further expected that its large structural fluctuation should induce such a large ΔS_{conf} as to adequately compensate for the increase in G_1 (cf. Table 2.2). It is concluded that the structural changes are driven by the lowering of the real free energy.

2.3.5 Cotranslational Folding Pathway of N-terminal Partial Mb Chains

In this section I discuss the folding pathway of nascent Mb chains during elongation. First of all, the posttranslational folding pathway of full-length Mb is briefly described. It was reported that the refolding process of denatured apo-Mb passes through an intermediate state which is very similar to the molten globule state formed at pH ~4 and low salt concentrations. Multi-dimensional NMR spectroscopy of this intermediate revealed that the AGH[B] hydrophobic core, which is composed of A-, G-, H- and a part of the B-helix, is formed and, at the same time, somewhat helical structures exist in the region spanning from the B- to E-helix (BCDE complex) [16,17]. This indicates that the folding process of full-length Mb begins with the formation of the AGH[B] hydrophobic core and then that of the BCDE complex follows.

In Fig. 2.5(a), a native contact map of the Mb crystal structure is shown. Here, the word “native contact" is used as the residue-residue contact (the closest heavy atom-heavy atom distance is less than 4 Å) in the native, or crystal, structure of full-length Mb. To avoid trivial contacts formed between adjacent residues, I excluded such contacts that the two residues are sequential within 5 residues. One can recognize that there are the native contact clusters corresponding to the AGH[B]

hydrophobic core and BCDE complex.

I now turn to the discussion of the cotranslational folding pathway. In Fig. 2.5(b), the numbers of native contacts formed in the N- and C-terminal partial Mb chains are plotted against the chain length ,both showing, of course, monotonically increases as the chain elongates. Considering the native contact formation, at the first stage, it is understood that short chains whose lengths are up to about 40 have very few native contacts, suggesting the structural instability as shown in the present Mb36 calculation. Next, during the elongation to about 80, native contacts located in the region corresponding to the BCDE complex are gradually formed, accompanying the structural stabilization as in Mb77. Then, during the F-helix synthesis, the number of native contacts become almost constant because F-helix has native contacts only with H-helix which is not synthesized yet. Finally, in response to the emergence of G- and H-helix, the formation of the AGH[B] hydrophobic core proceeds and thus the principal folding process completes. Accordingly, there is a major difference between posttranslational and cotranslational folding pathway in the order of the formation of the AGH[B] hydrophobic core and the BCDE complex: the former occurs first in the posttranslational folding but the latter does in the cotranslational one.

Experimentally, it was revealed that the nascent Hb α-subunit chain is capable of an efficient heme binding when it elongates to 86 amino acid residues in spite of the absence of the proximal histidine His87, with which the heme binds [1]. They conjectured that the heme binding is brought about mainly by the formation of the hydrophobic contacts among the heme and the surrounding residues. Including the proximal and distal histidines, there are 18 residues contacting the heme in the full-length α-globin. In case of the length of 86 residues, 11 out of the 18 residues exist and are deemed to be enough to stabilize the heme. If the hypothesis could be

adopted, it would be expected that the heme affinity of Mb92 should be high because, in Mb92, there are 10 residues out of those 16 residues that have native contacts with the heme (this ratio is similar to that of the partial α-globin chain). It can also be speculated that, even without the proximal histidine His93, the heme would bind to the distal histidine His64 before the synthesis of the proximal one.

In Fig. 2.5(b), it is also suggested that the amino-acid sequence of Mb has an advantageous property for the cotranslational folding since those native contacts appear earlier in N-terminal chain than in C-terminal one. For example, when each partial chain length elongates up to the length of 80 residues, N- and C-terminal chains have 92 and 70 native contacts, respectively. Such a propensity that N-terminal chain forms native contacts more easily is consistent with the statistical analysis of compilations of native protein structures [18]: residues tend to interact with the previously synthesized part of the molecule.

2.4. Concluding Remarks

The stabilities of partial chains in apo-Mb have been investigated by MD simulations. The helices longer than 77 residues are stable in the 10 ns simulation, with the exception of A-helix instability. A free energy analysis has revealed that the helical disorder occurs because the gain of solvation and conformational free energies overbalances the loss of internal energy. By analyzing the native contacts, it was estimated that the cotranslational folding pathway of Mb would be different from the posttranslational one.

Nowadays, the importance of cotranslational folding is attracting attention [19,20]. For example, certain nascent proteins exhibit ligand binding [1,21,22] and cotranslational closing of disulfide bonds [23,24]. The efficient heme binding of

nascent α-globin chains [1] is an example of cotranslational folding. The present results support the mechanism that nascent apo-Mb chains start formation of native-like α-helix structures after elongation of at least 77 amino acid residues and shed light on the cotranslational folding process of apo-Mb.

References

[1] A.A. Komar, A. Kommer, I.A. Krasheninnikov, A.S. Spirin, J. Bio. Chem. 272 (1997) 10646.

[2] C.C. Chow, C. Chow, V. Raghunathan, T.J. Huppert, E.B. Kimball, S. Cavagnero, Biochemistry 42 (2003) 7090.

[3] P.A. Jennings, P.E. Wright, Science 262 (1993) 892.

[4] D.A. Case, D.A. Pearlman, J.W. Caldwell, T.E. Cheatham III, J. Wang, W.S. Ross, C.L. Simmerling, T.A. Darden, K.M. Merz, R.V. Stanton, A.L. Cheng, J.J. Vincent, M. Crowley, V. Tsui, H. Gohlke, R.J. Radmer, Y. Duan, J. Pitera, I. Massova, G.L. Seibel, U.C. Singh, P.K. Weiner, P.A. Kollman, AMBER7, University of California, San Francisco, 2002.

[5] J. Wang, P. Cieplak, P.A. Kollman, J. Comp. Chem. 21 (2000) 1049.

[6] H.J.C. Berendsen, J.P.M. Postma, W.F. van Gunsteren, A. DiNola, J.R. Haak, J. Chem. Phys. 81 (1984) 3684.

[7] J. Wang, W. Wang, S. Huo, M. Lee, P.A. Kollman, J. Phys. Chem. B 105 (2001) 5055.

[8] M.R. Lee, Y. Duan, P.A. Kollman, Proteins: Struct. Funct. Genet. 39 (2000) 309.

[9] M.F. Sanner, A.J. Olson, J. Spehner, Biopolymers 38 (1996) 305.

[10] W. Humphrey, A. Dalke, K. Schulten, J. Molec. Graphics, 14 (1996) 33.

[11] A. Onufriev, D.A. Case, D. Bashfold, J. Mol. Biol. 325 (2003) 555.

[12] W. Kabsch, C. Sander, Biopolymers 22 (1983) 2577.

[13] D.J. Tobias, J.E. Mertz, C.L. Brooks III, Biochemistry 30 (1991) 6054.

[14] J.D. Hirst, C.L. Brooks III, Biochemistry 34 (1995) 7614.

[15] B. Hallerbach, H.–J. Hinz, Biophys. Chem. 76 (1999) 219.

[16] D. Eliezer, J. Yao, H.J. Dyson, P.E. Wright, Nature Struct. Biol. 5 (1998) 148.

[17] D. Eliezer, J. Chung, H.J. Dyson, P.E. Wright, Biochemistry 39 (2000) 2894.

[18] N. Alexandrov, Protein Sci. 2 (1993) 1989.

[19] V.A. Kolb, Mol. Biol. 35 (2001) 584.

[20] G. Kramer, V. Ramachandiran, B. Hardesty, Int. J. Biochem. Cell Biol. 33 (2001) 541.

[21] J.E. Mullet, P.G. Klein, R.R. Klein, Proc .Natl. Acad. Sci. USA 87 (1990) 4038.

[22] M.F. Mouat, Int. J. Biochem. Cell Biol. 32 (2000) 327.

[23] L.W. Bergman, W.M. Kuehl, J. Biol. Chem. 254 (1979) 8869.

[24] W. Chen, J. Helenius, I. Braakman, A. Helenius, Proc. Natl. Acad. Sci. USA 92 (1995) 6229.

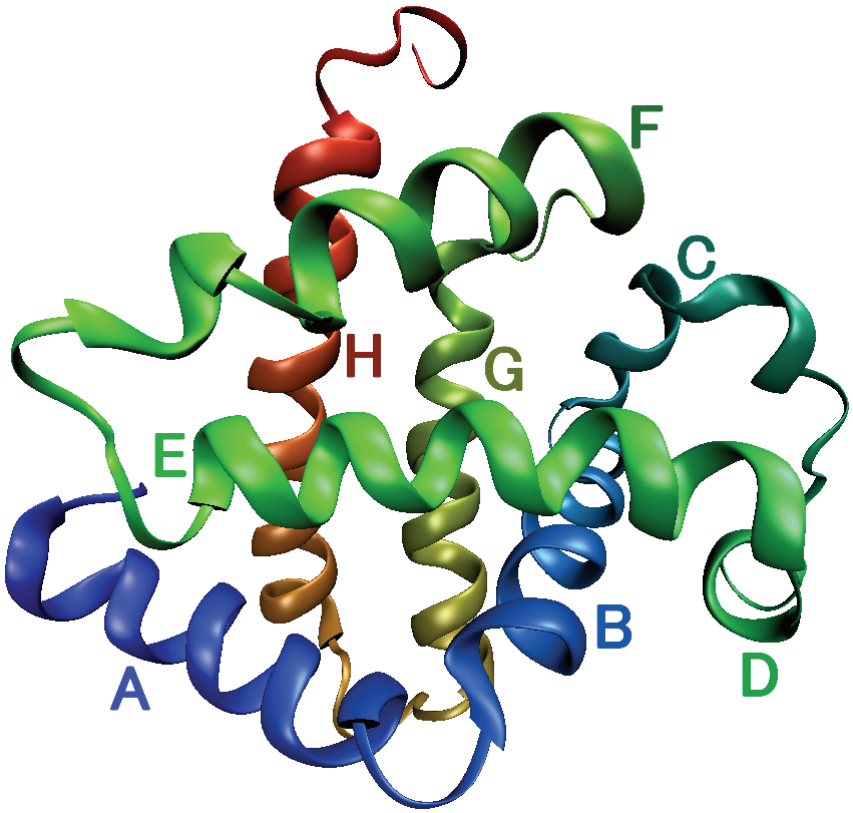

Fig.2.1. Ribbon presentation of the X-ray crystallographic structure of Mb consisting of eight helices A to H (PDB entry: 104M).

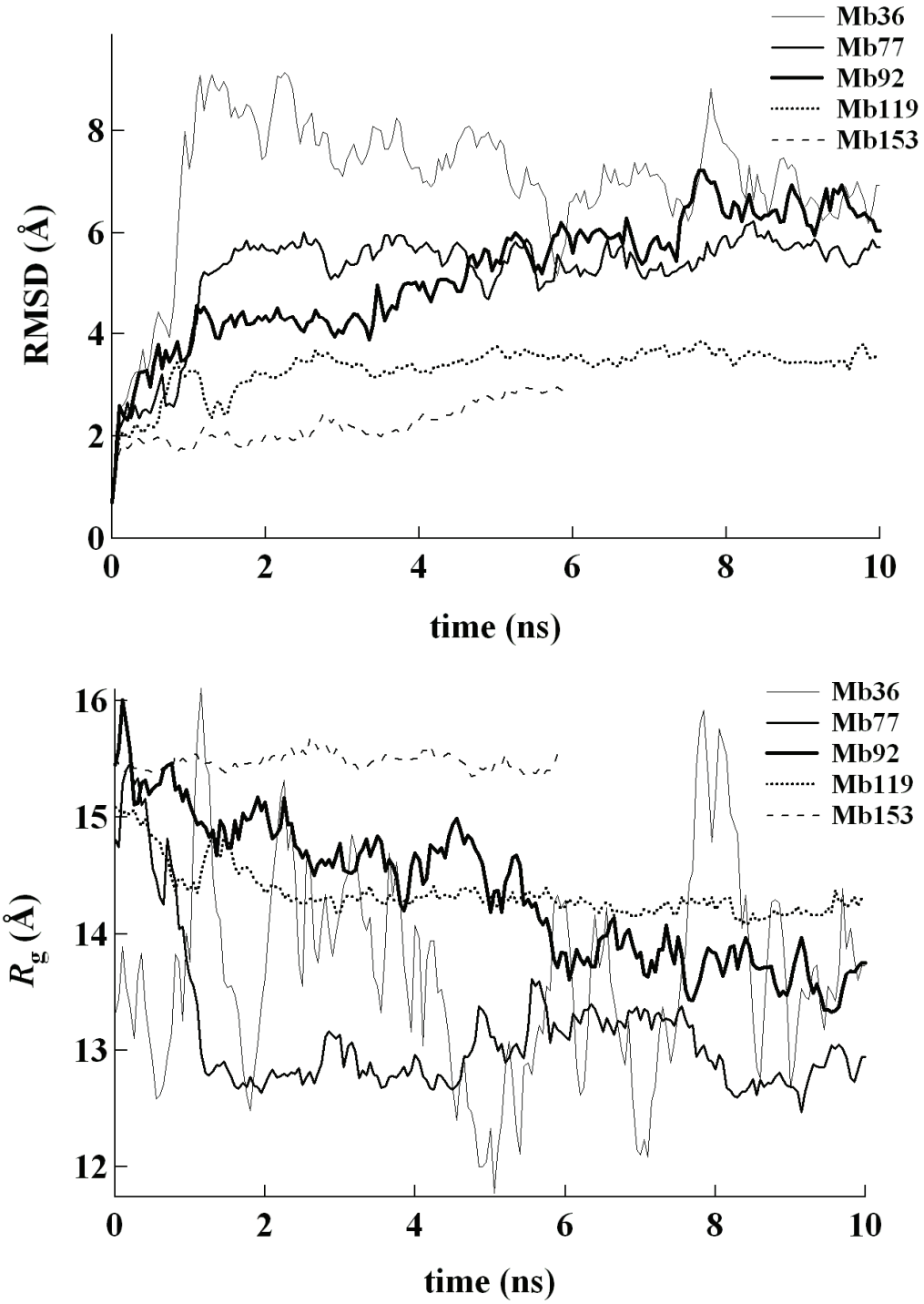

Fig.2.2. (a) RMSDs of all atoms with reference to the crystal structure of each partial Mb chain during MD simulations for 10 ns. (b) Radius of gyrations of each partial Mb chain during MD simulations for 10 ns.

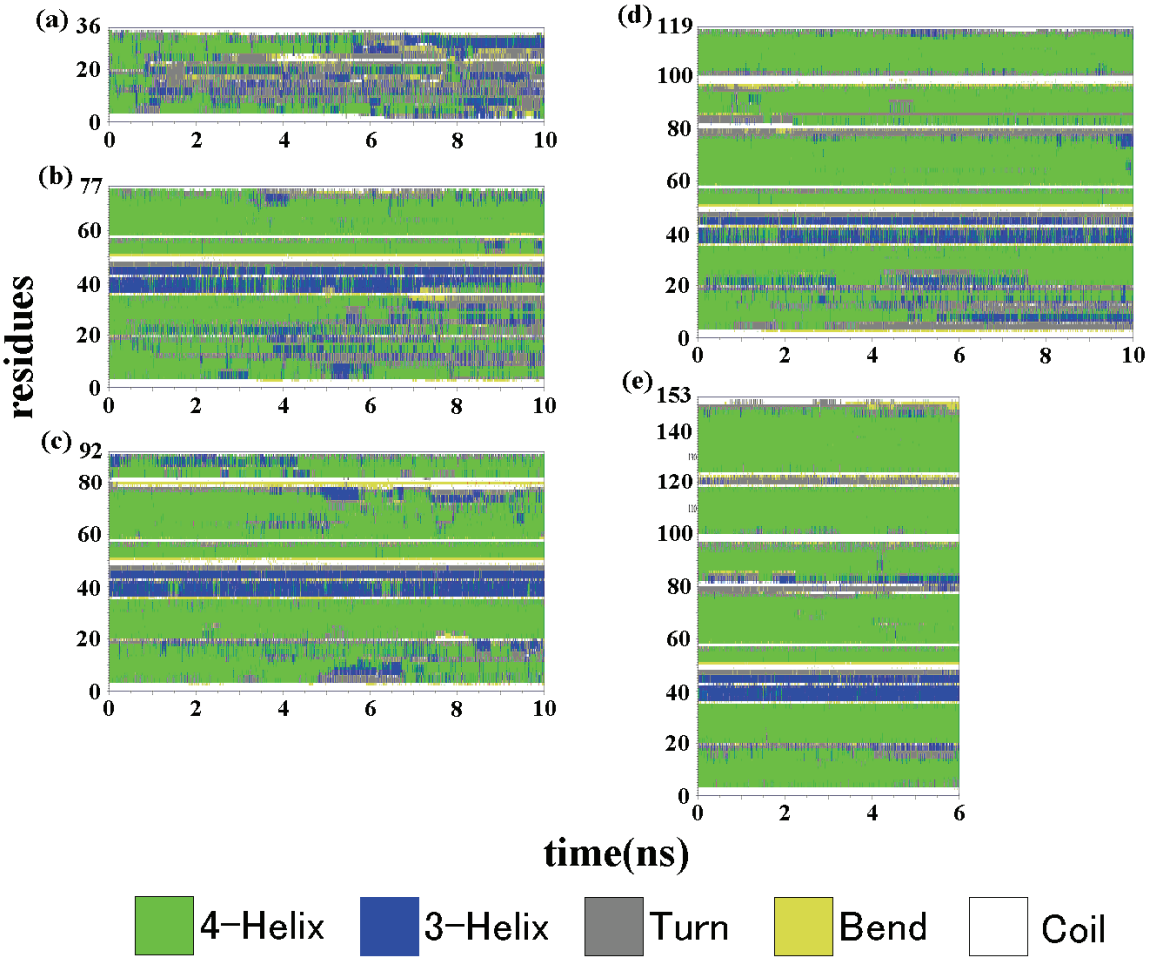

Fig.2.3. Secondary structures as a function of time: (a) Mb36, (b) Mb77, (c) Mb92, (d) Mb119, and (e) Mb153. Colors refer to: green: 4-Helix, blue: 3-Helix, gray: Turn, yellow: Bend, and white: are Coil.

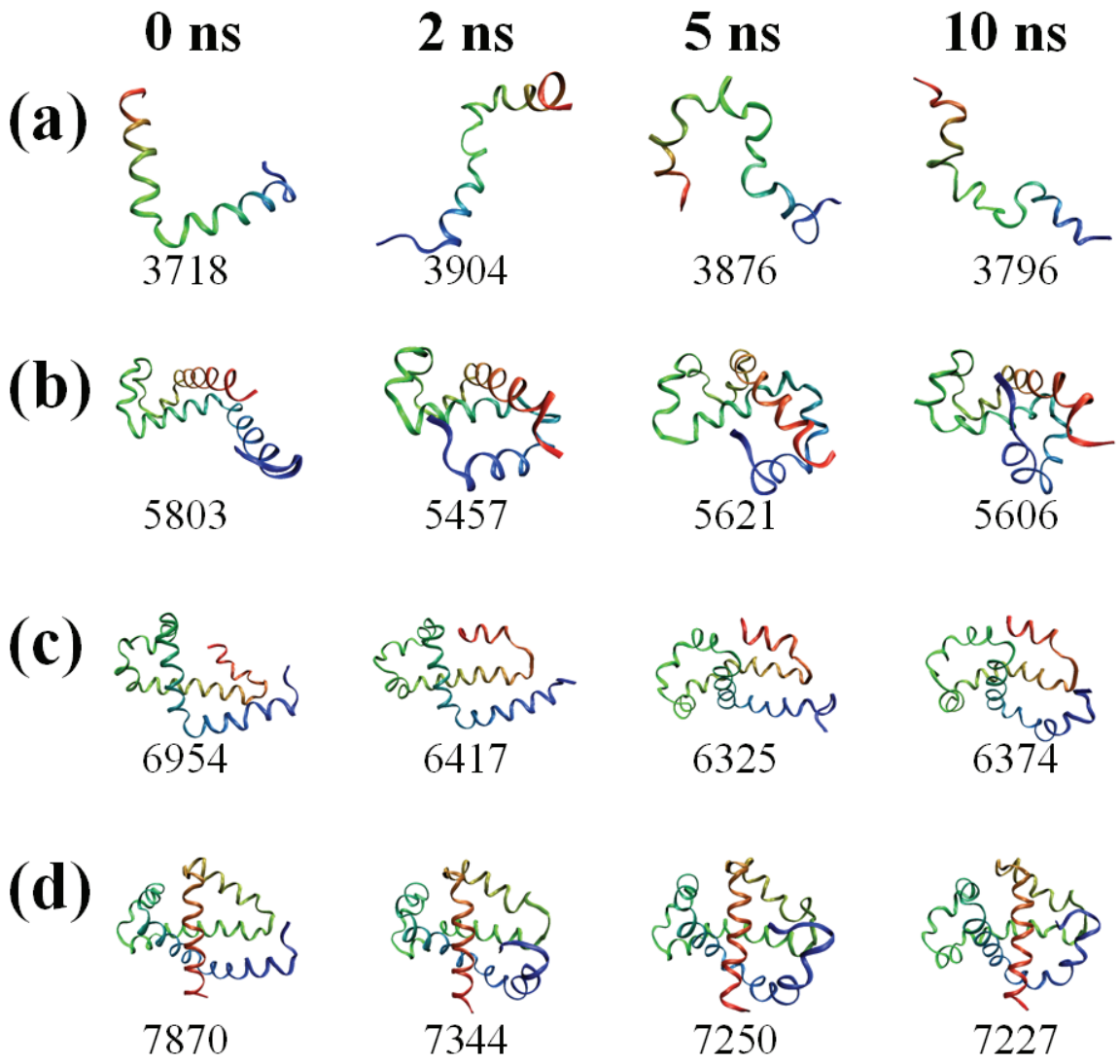

Fig.2.4. Snapshots of: (a) Mb36, (b) Mb77, (c) Mb92, and (d) Mb119 at 0, 2, 5, and 10 ns. Values of solvent accessible surface area (SASA) in $Å^2$ are listed below the structures.

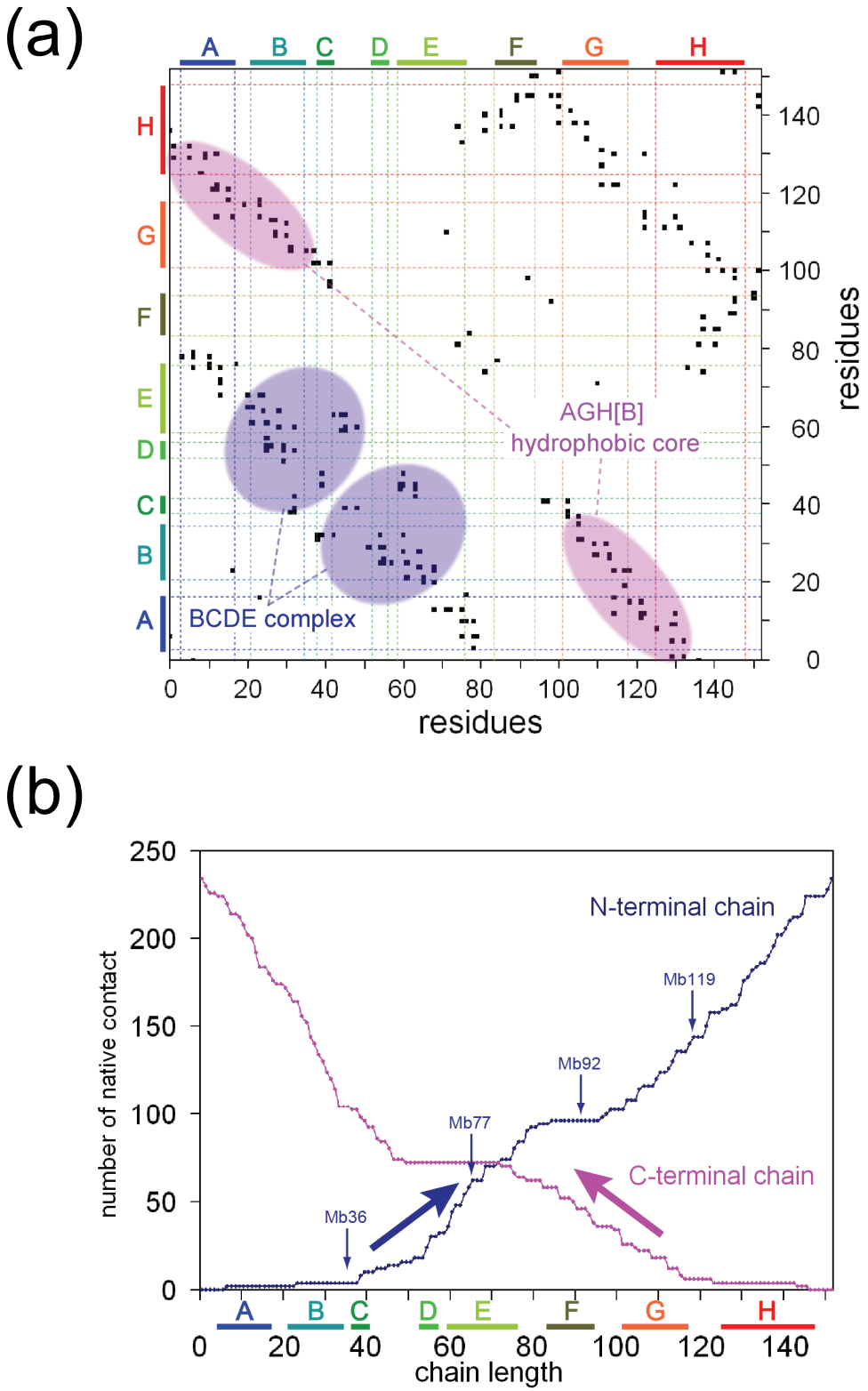

Fig.2.5. (a) Native contact map of the Mb crystal structure. (b) Number of the native contacts of N- and C-terminal partial Mb chains plotted against the chain length.

Table 2.1

Estimated helix (4-turn and 3-turn) ratios of amino acid residues in α-helix [a]

Peptide	MD [b]	Far-UV CD [c]
Mb36	0.38	0.04
Mb77	0.57	0.16
Mb92	0.69	—
Mb119	0.68	0.30
Mb153	0.74	0.50

[a] All amino acid residues in the last 2 ns, MD simulations and Far-UV CD data for α-helix.

[b] Present study.

[c] Ref. [2].

Table 2.2

Estimated free energies of peptide chains (in kcal/mol)

Peptide		First 1 ns	Last 1 ns	Change
Mb36	E_{MM} [a]	-226	-184	42
	ΔG_{WSAS} [b]	-605	-623	-18
	G_1 [c]	-831	-807	24
Mb77	E_{MM}	-783	-715	68
	ΔG_{WSAS}	-1288	-1356	-68
	G_1	-2071	-2071	0
Mb92	E_{MM}	-955	-943	12
	ΔG_{WSAS}	-1551	-1575	-24
	G_1	-2506	-2518	-12
Mb119	E_{MM}	-824	-818	6
	ΔG_{WSAS}	-1853	-1905	-52
	G_1	-2677	-2723	-46
Mb153	E_{MM}	-1323	-1260	63
	ΔG_{WSAS}	-2323	-2370	-47
	G_1	-3646	-3630	16

[a] Sum of internal energy.

[b] Solvation free energy.

[c] Free energy of each snapshot.

Chapter 3

Anisotropic Structural Relaxation and its Correlation with the Excess Energy Diffusion in the Incipient Process of Photo-dissociated MbCO: High Resolution Analysis via Perturbation Ensemble Method

3.1. Introduction

Even if the equilibrium structure of a protein might be precisely understood, you often need further investigation to clarify the molecular mechanisms for protein functions at the atomic level. Such an example is *allostery* in the oxygen binding of hemoglobin (Hb) that occurs in the low affinity deoxy state (T state) or in the high affinity ligated state (R state) [1]. So far, as a first step to understand the Hb allostery, many experimental and theoretical studies have been carried out for a ligand photolysis process of myoglobin (Mb), which has a structure very similar to a Hb subunit.

The photolysis of carbonmonoxy myoglobin (MbCO) is one of the most thoroughly studied phenomena because of its advantageous properties such as few side reactions. Such as structural changes and vibrational energy relaxation of the heme group and the CO ligand escape process after the photolysis were investigated by the picosecond time-resolved resonance Raman spectroscopy [2,3], transient grating (TG) technique [4,5], molecular dynamics (MD) simulations [6-10] and so on. In particular, the most remarkable one is the direct observation of the structural changes driven by the photolysis with time-resolved X-ray crystallography [11]. Thus, one can literally "*watch*" the CO escape and concomitant structural changes mainly occurring in the vicinity of the heme group from 100 ps to several μs after the photolysis.

In the present chapter, the structural change of MbCO immediately after the photolysis (less than 100 ps time regime) was elucidated by the perturbation ensemble (PE) method. Its use makes it possible to make a high accuracy analysis and to detect such subtle changes that are almost completely obscured by the large fluctuation

in a single MD trajectory. Obtained results show clearly the incipient globin dynamics after the photolysis that undergoes faster than the time resolution of the time-resolved X-ray crystallography [11]. This should fill the gap of the experiments. In the following section 3.2, computational model and method are explained. In section 3.3, the results and discussion are presented. Finally, in section 3.4, concluding remarks are provided.

3.2. Computational Model and Methods

3.2.1. Molecular Dynamics Simulations

All MD simulations were carried out with AMBER7 program using the parm99 force field [12]. For the heme and the heme-bound CO ligand parameters, I used the bond, angle, and dihedral parameters developed by Henry *et al.* [9a] and the electrostatic and van der Waals parameters determined by Giammona [13]. The atomic charges of the dissociated CO molecule are calibrated to mimic its dipole moment 0.11 D [14], i.e., +0.021*e* and -0.021*e* for the carbon and oxygen atom, respectively. The SHAKE method was used to constrain the hydrogen-heavy atom bond distances, and the particle mesh Ewald procedure was used to handle long-range electrostatic interactions. The integration time-step was 2 fs. All the simulations were performed using periodic boundary conditions.

The MbCO structure obtained by a neutron diffraction analysis was retrieved from the PDB (code 2MB5). It was solvated with 2986 TIP3P water molecules and 9 chloride counter ions yielding a periodic box size ~ 51×47×47 $Å^3$. Next, after a 400 ps equilibration NPT run at ambient condition (300 K, 1 atm), a 600 ps NPT run was performed. By saving snapshots every 1 ps interval, 600 snapshots were obtained. Then, from each of the 600 snapshots, with and without the preparation of

the photolyzed state described in the following section, a couple of perturbed MD (PMD) and unperturbed MD (UMD) simulation were performed for 100 ps in the NVE ensemble.

To simulate the photolysis of CO ligand molecule by its absorption of one photon of wavelength 353 nm (81 kcal/mol), the force field parameters of the heme group were changed from those of the ground state ("Liganded" in ref [9a]) to those of the excited state ("Unliganded hard" in ref [9a]). In the ground state, the heme is six-coordinated, planar, and CO ligated. On the other hand, in the excited state, it is five-coordinated, domed, and in the deoxy form. In addition, as described in my previous paper [6], I introduced a repulsive potential function between the iron atom and the ligand molecule and deposited the excess kinetic energy 54 kcal/mol into 24 atoms of the porphin ring by adding 24 randomly-generated excess velocity vectors to the original velocity ones. It should be noted here that the 24 random velocity vectors were generated to conserve the same total linear and angular momenta of the 24 atoms as before the photon absorption. The remaining energy of the photon (27 kcal/mol) was, therefore, assumed to be used to change the electronic state of the heme.

3.2.2. Perturbation Ensemble Method

Here, the procedure of the PE method is explained. This is a very accurate method to estimate physical quantity changes induced by a perturbation. To start with, many pairs of PMD and UMD simulation are executed with and without the perturbation, respectively. Note that all the initial atomic positions and velocities for a couple of PMD and UMD calculation are identical except for those in the perturbed region (i.e., the atomic velocities of the 24 heme atoms were modified). Next, N_{traj}

pairs of numerical values of the physical quantity A are calculated from the PMD and UMD simulations.

$$\left(A^{\mathrm{PMD},i}, A^{\mathrm{UMD},i} \right) i = 1, 2, \cdots, N_{\mathrm{traj}} \tag{3.1}$$

where N_{traj} is the number of trajectories and $A^{\mathrm{PMD},i}$ and $A^{\mathrm{UMD},i}$ are those values of the physical quantity A calculated through the ith PMD and UMD trajectories, respectively. Then, the variation of A between the couple of PMD and UMD simulation for the trajectory number i, is calculated:

$$\delta A^i = A^{\mathrm{PMD},i} - A^{\mathrm{UMD},i}. \tag{3.2}$$

Finally, the variation is ensemble-averaged over the N_{traj} trajectories:

$$\left\langle \delta A \right\rangle = \frac{1}{N_{\mathrm{traj}}} \sum_{i=1}^{N_{\mathrm{traj}}} \delta A^i = \frac{1}{N_{\mathrm{traj}}} \sum_{i=1}^{N_{\mathrm{traj}}} \left(A^{\mathrm{PMD},i} - A^{\mathrm{UMD},i} \right). \tag{3.3}$$

In addition, the time change of A at time t with respect to that at the initial time is defined as

$$\Delta A^i(t) = A^i(t) - A^i(0) \tag{3.4}$$

and the ensemble-averaged time change and its perturbative variation are thus calculated as

$$\left\langle \Delta A(t) \right\rangle = \frac{1}{N_{\mathrm{traj}}} \sum_{i=1}^{N_{\mathrm{traj}}} \Delta A^i(t) = \frac{1}{N_{\mathrm{traj}}} \sum_{i=1}^{N_{\mathrm{traj}}} \left(A^i(t) - A^i(0) \right) \tag{3.5}$$

and

$$\begin{aligned} \left\langle \delta\Delta A(t) \right\rangle &= \frac{1}{N_{\mathrm{traj}}} \sum_{i=1}^{N_{\mathrm{traj}}} \delta\Delta A^i(t) \\ &= \frac{1}{N_{\mathrm{traj}}} \sum_{i=1}^{N_{\mathrm{traj}}} \left\{ \left(A^{\mathrm{PMD},i}(t) - A^{\mathrm{PMD},i}(0) \right) - \left(A^{\mathrm{UMD},i}(t) - A^{\mathrm{UMD},i}(0) \right) \right\}. \end{aligned} \tag{3.6}$$

3.2.3. Radius of Gyration and its Components

To detect the anisotropic structural change after the photolysis, the total radius of gyration R_g and its three Cartesian components, R_{gx}, R_{gy}, and R_{gz}, are calculated as

$$R_g(t) = \sqrt{\frac{\sum_i m_i \left|\mathbf{r}_i(t) - \mathbf{r}_c(t)\right|^2}{\sum_i m_i}}, \tag{3.7}$$

and

$$R_{ga}(t) = \sqrt{\frac{\sum_i m_i \left(a_i(t) - a_c(t)\right)^2}{\sum_i m_i}} \qquad (a = x, y, z) \tag{3.8}$$

where $\mathbf{r}_i = (x_i, y_i, z_i)$ and $\mathbf{r}_c = (x_c, y_c, z_c)$ are the Cartesian coordinates of the *i*th atom and the center of mass (COM) of the Mb, respectively. Here, I used the same coordinate system as Sagnella and Straub used [8a]: the *xy*-plane is taken to coincide with the heme plane defined by the Fe atom and the two N atoms of the pyrrole heme rings A and B, and the *z*-axis is perpendicular to it. The axes are shown in Fig. 3.1. For each MD trajectory analysis, the coordinate system was fixed at that defined at each initial structure.

3.3. Results and Discussion

3.3.1. High Accuracy of the PE Method

In Fig. 3.2, plotted are the R_g change of a single PMD trajectory and those changes ensemble-averaged over 5, 25, 125, and 600 PMD trajectories. It can be understood that the fluctuation was significantly reduced by increasing the number of trajectories for the average. By averaging the 600 values, the R_g was obtained with

small fluctuation and subtle changes can be detected within the accuracy of 0.03 Å, which were almost completely obscured in a single PMD value due to its fluctuation larger than 0.1 Å. The ensemble-averages, standard deviations (SD), and two-sided 95% confidence intervals (95% CI) of the R_g that were calculated from the initial (t = 0 ps) structures of 600 PMD trajectories are shown in Table 3.1. It was quantitatively shown that the R_g has a 0.1 Å SD, or fluctuation, and the ensemble-averaging reduces the fluctuation drastically down to less than 0.01 Å and achieves the high accuracy data. Thus, it follows from what has been said that the ensemble-averaging over many trajectories makes it possible to analyze the very subtle changes that would be completely obscured in the values calculated from only a single or a few trajectories

Here, in order to investigate the anisotropic structural change after the photolysis, three Cartesian components of the radius of gyration, R_{gx}, R_{gy}, and R_{gz} were calculated by averaging over 600 PMD trajectories (Fig. 3.3(a)). Although they seem to exhibit just small fluctuations and different behaviors of each component are clearly indicated, those calculated from the 600 UMD trajectories also exhibit varying behaviors even without the photolysis: Mb expands in the x-direction and contracts in the y- and z-directions (Fig. 3.3(b)). It is estimated that this might be due to an ongoing structural change that proceeds continuously for a long time (several hundreds picoseconds or more) from the initial structures.

A perturbative treatment was, therefore, used additionally to cancel out the changes of the UMD trajectories. That is to say, I first calculate differences from each pair of PMD and UMD trajectory (i.e., a physical quantity in a UMD trajectory is subtracted from that in a corresponding PMD trajectory) and then ensemble-average some physical quantities over the trajectories. In the early days on the simulation study of heme heating effects, Henry et al. executed also the similar approach, but on a

much more modest scale (8 trajectories), to know the effect of photoexcitation with respect to Mb and Cyt-c [9b]. I call hereafter the present extensive version of the combined method of an ensemble-average with a perturbation procedure as the *perturbation ensemble* (*PE*) *method*. R_g changes with respect to the initial structure were calculated by the PE method using eq. (3.6) (Fig. 3.4). Since both the averaging and the perturbation treatment significantly reduce the high- and low-frequency fluctuations, respectively, the changes in Fig. 3.4 reflect only structural deviations triggered by the photolysis. This demonstrates clearly the high accuracy of the PE method.

Let me now assess the margin of error of the R_gs calculated by the PE method. Both changes with their 95% CIs at 1 ps and 100 ps after the photolysis are summarized in Table 3.2. The reduction of fluctuations is quantitatively clear. At the moment of the excitation (t = 0 ps), all the changes and 95% CIs are, of course, zero because the initial structures of the pairs of PMDs and UMDs are exactly identical. As time advances, the structural deviations between PMD and UMD continue to increase concurrently with a monotonic widening of the 95% CIs. At 1 ps, the CIs are very narrow, ±0.002 Å. Even at 100 ps, i.e., at the time of the widest CIs in the current 100 ps long calculations, the averaging over the 600 trajectories provides adequately accurate R_gs, whose maximum 95% CI is ±0.02 Å and in the same order as those of the structural change that I now focus on.

3.3.2. Anisotropy of the Incipient Structural Evolution Following CO Photodissociation

As I have seen in the previous section, the R_g changes, which show adequately small fluctuations (maximum 95% CI is ±0.02 Å), were obtained under only the

influence of the photolysis (Fig. 3.4). Comparing the three component values, therefore, the following results were obtained: the behaviors of in-plane (R_{gx} and R_{gy}) and out-of-plane (R_{gz}) components are specifically different. While those in the former show contractions after the short-time small expansions within 1 ps (peak at about 0.7 ps), that in the latter obviously rises during the simulation time with a rapid increase up to 0.04 Å immediately after the photolysis. This anisotropic structural change, which mainly occurs in the *z*-direction, was induced by two mechanisms: one is the displacement of the heme iron atom induced by the heme conformational change from the planer form to the domed one and the other is the photolyzed CO ligand collision with the distal pocket residues. It should be noted that, among and during 100 ps PMD simulations, most of the dissociated COs appeared in the primary docking site except for such a few trajectories that the COs migrated to the Xe4 site [10]. These push the proximal histidine or distal residues outward along the *z*-direction and result in the *z*-direction expansion. In fact, the structure ensemble-averaged over the PMD trajectories provides such clear displacements of residues in the heme vicinity as were measured by the time-resolved X-ray crystallography [11].

It was found that most of the obtained anisotropic changes undergo within 1 ps, followed by relatively small changes. This agrees with the results of TG spectroscopy by Goodno *et al*. [4]. They also demonstrated the asymmetric conformational relaxation in the protein within 500 fs whose out-of-plane (*z*-direction) expansion is larger than the in-plane (*x*- and *y*-directions) changes. However, their analysis could not definitively determine whether the in-plane changes should be expansion or contraction because the observed signals were the sum of the contributions from the globin strain and the solvent heating. On the other hand, since the current calculations include only the contribution from the structural change of Mb and reveal the contraction in the *x*- and *y*-directions, the present observation is

considered an affirmative evidence that Mb partially should show a physical property of an elastic medium. That is to say, if Mb expands along one direction, it contracts at the same time along the perpendicular directions to keep its atomic number density constant.

3.3.3. Energy Flow to Water Solvent through the Anisotropic Structural Change

So far, several experimental and theoretical studies have been devoted to the relaxation process of the heme excess energy or the energy transfer to the water solvent. Lian *et al*. have studied the energy transfer from solute to solvent in the photoexcited MbCO in D_2O solvent by the femtosecond IR spectroscopy [15]. Their results revealed that the heating of solvent occurs in two timescales, i.e. slow and fast components. While the slow one was attributed to a classical diffusion process whose time constant is approximately 20 ps, the fast one was best fitted by a Gaussian rise function with a time constant of 7.5 ps, not the classical diffusion.

In fact, this fast process occurring within 10 ps has been investigated by several researchers. For example, the heating of the water solvent in the vicinity of the heme propionate side-chains was directly calculated by MD simulations [8] and the normal mode analysis revealed that the vibrational modes of the propionate groups are coupled with those of water and the efficient energy transfer to solvent was suggested [6]. In addition, those studies by MD simulations [8b] or by spectroscopy [16] revealed that the deletion of the propionate side chains notably delays the relaxation times. All of these studies focused on the propionate side chains that were exposed to the water solvent and, actually, their important roles in the energy transfer process were recognized.

On the other hand, different from the fast energy dissipation through the exposed propionate side chains, the present anisotropic structural change clarified by the PE method is also considered another pathway through the collective motions of the globin matrix, as Lian *et al*. supposed [15]. Precisely, the displacements of the proximal histidine or the distal residues, triggered by the CO ligand photolysis, excite the low-frequency protein dynamics of the globin and then the fast energy transfer to the water solvent occurs. Now, to estimate the contribution, calculated are those distance changes between the COM of Mb and those of exposed residues located in the perpendicular direction relative to the heme plane (Fig. 3.5). All the residues obviously show positional extensions (most of them occur in the *z*-direction). In fact, there are steep rises in the first 1 ps and then gradual extensions continue up to 20 ps. The results make it clear that the *z*-direction expansion propagates to a number of exposed residues forming the surface of Mb. Considering the fact that most of the expansion took place within the timescale of the fast component (7.5 ps) [15], it is supposed that the displacements of the residues play important roles in the fast energy transfer process by inducing collisional interactions with surrounding water molecules.

Now, the mechanical work done by the *z*-direction expansion within the first 1 ps is thermodynamically estimated. The *isothermal compressibility* takes the form

$$\kappa_T = -\frac{1}{V^{\mathrm{wat}}}\frac{\Delta V}{\Delta p} \tag{3.9}$$

where ΔV and Δp are the volume and pressure change of the water solvent, respectively, and V^{wat} is a volume of the region of water solvent where the expansion of Mb influences directly within 1 ps after the photolysis. Then the work toward the water solvent can be obtained by

$$W = -p\Delta V \approx -\Delta p \Delta V = \frac{\Delta V^2}{\kappa_T V^{\mathrm{wat}}}. \tag{3.10}$$

Here it is approximated as $p = p_0 + \Delta p \approx \Delta p$ (p_0 is the ambient pressure) because the calculated variation is $\Delta p >> p_0 = 1$ atm. In calculating the volume change, it is assumed that Mb might be a sphere of radius $R_g = 15.5$ Å and expands by $\Delta z^{Mb} = 0.1$ Å in the z-direction on both sides to be an ellipsoidal. This yields the change of $\Delta V = -(4\pi R_g^2 \times 2\Delta z^{Mb}) / 3 = -201.3$ Å^3. In addition, V^{wat} is estimated by the velocity of the sound wave in water, 15 Å/ps [14]. Since the R_g peak is at 0.7 ps (Fig. 3.4), the compression wave propagate to the surface of Mb in 0.7 ps and, after the following 0.3 ps, the thickness of the compressed water region in the z-direction from the Mb surface is estimated to be $\Delta z^{wat} = 4.5$ Å. Under these assumptions, V^{wat} is estimated to be $\pi R_g^2 \times 2\Delta z^{wat} = 6792.9$ Å^3. Substituting these values and the isothermal compressibility of TIP3P water ($\kappa_T = 18\times10^{-11}$ Pa^{-1} [17]) into eq 3.10, the work is estimated to be $W = 4.8$ kcal/mol.

You should notice that the estimation is based on some fundamental assumptions that the system is continuum and characterized by the equilibrium physical properties (isothermal compressibility and velocity of sound wave). In fact, it is true that the incipient expansion within 1 ps is, of course, a nonequilibrium phenomenon in the atomic level and, hence, the accuracy of $W = 4.8$ kcal/mol, should be carefully examined. However, considering the fact that the value, 4.8 kcal/mol, is about 6% of the excess energy of 81 kcal/mol that the heme would reserve after the photoexcitation, it can be concluded that the contribution of the supposed work to the fast energy transfer process should not be negligible.

3.4. Concluding Remarks

In this chapter, in order to detect the subtle anisotropic structural change of MbCO triggered by the ligand photolysis, the time changes of R_g and its three Cartesian components, R_{gx}, R_{gy}, and R_{gz}, were calculated for 100 ps by the PE method,

for the first time, in which many pairs of PMD and UMD simulations were executed and the variations between them were ensemble-averaged. This method works very well and provides accurate enough R_g changes free from large fluctuations. The results drawn from the method agree very well with the previous observation of the transient grating experiment indicating the same anisotropic expansion within 500 fs, i.e. Mb largely expands in the perpendicular direction of the heme plane and slightly contracts in the other directions. The good agreement between the experimental results and my computational ones demonstrates the high accuracy of the PE method to detect subtle changes that are completely obscured under their thermal fluctuations. It is, therefore, concluded that Mb undergoes the anisotropic expansion immediately after the CO ligand photolysis.

Moreover, several previous studies have been devoted to the excess energy relaxation from the heme to the surrounding solvent after the photolysis. Consequently, on one hand, an important role of the propionate side chains to the fast energy transfer process has been elucidated. On the other hand, it has been long supposed that the obtained anisotropic structural changes, or collective motions of the globin matrix, also contribute to the fast process. Presently, it was quantitatively shown that its contribution is important and the work done on the solvent by the expansion within 1 ps was thermodynamically estimated to be 4.8 kcal/mol. Although you should notice that the estimation is based on several assumptions, this estimation suggests that its contribution to the fast energy relaxation should be nonnegligible.

References

[1] (a) M.F. Perutz, A.J. Wilkinson, M. Paoli, G.G. Dodson, Annu. Rev. Biophys. Biomol. Struct. 27 (1998) 1; (b) W.A. Eaton, E.R. Henry, J. Hofrichter, A. Mozzarelli, Nat. Struct. Biol. 6 (1999) 351.

[2] S. Franzen, B. Bohn, C. Poyart, J.L. Martin, Biochemistry 34 (1995) 1224.

[3] (a) Y. Mizutani, T. Kitagawa, Science 278 (1997) 443; (b) Y. Mizutani, T. Kitagawa, J. Phys. Chem. B 105 (2001) 10992; (c) Y. Mizutani, T. Kitagawa, Bull. Chem. Soc. Jpn. 75 (2002) 623; (d) A. Sato, Y. Mizutani, Biochemistry 44 (2005) 14709.

[4] G.D. Goodno, V. Astinov, R.J.D. Miller, J. Phys. Chem. A 103 (1999) 10630.

[5] (a) M. Sakakura, S. Yamaguchi, N. Hirota, M. Terazima, J. Am. Chem. Soc. 123 (2001) 4286; (b) Y. Nishihara, M. Sakakura, Y. Kimura M. Terazima, J. Am. Chem. Soc. 126 (2004) 11877.

[6] I. Okazaki, Y. Hara, M. Nagaoka, Chem. Phys. Lett. 337 (2001) 151.

[7] J.E. Straub M. Karplus, Chem. Phys. 158 (1991) 221.

[8] (a) D.E. Sagnella, J.E. Straub, J. Phys. Chem. B 105 (2001) 7057; (b) L. Bu and J.E. Straub, J. Phys. Chem. B 107 (2003) 10634.

[9] (a) E.R. Henry, M. Levitt, W.A. Eaton, Proc. Natl. Acad. Sci. USA 82 (1985) 2034; (b) E.R. Henry, W.A. Eaton, R.M. Hochstrasser, Proc. Natl. Acad. Sci. USA 83 (1986) 8982.

[10] (a) C. Bossa, M. Anselmi, D. Roccatano, A. Amadei, B. Vallone, M. Burunori, A. Di Nola, Biophys. J. 86 (2004) 3855; (b) C. Bossa, A. Amadei, I. Daidone, M. Anselmi, B. Vallone, M. Burunori, A. Di Nola, Biophys. J. 89 (2005) 465.

[11] (a) F. Schotte, M. Lim, T.A. Jackson, A.V. Smirnov, J. Soman, J.S. Olson, G.N. Phillips Jr., M. Wulff, P.A. Anfinrud, Science 300 (2003) 1944; (b) F. Schotte, J. Soman, J.S. Olson, M. Wulff, P.A. Anfinrud, J. Struct. Biol. 147 (2004) 235; (c) G. Hummer, F. Schotte, P.A. Anfinrud, Proc. Natl. Acad. Sci. USA 101 (2004) 15330.

[12] D.A. Case, D.A. Pearlman, J.W. Caldwell, T.E. Cheatham III, J. Wang, W.S. Ross, C.L. Simmerling, T.A. Darden, K.M. Merz, R.V. Stanton, A.L. Chen, J.J. Vincent, M. Crowley, V. Tsui, H. Gohlke, R.J. Radmer, Y. Duan, J. Pitera, I. Massova, G.L. Seibel, U.C. Singh, P.K. Weiner, P.A. Kollman, AMBER 7, University of California, San Fransisco, CA, 2002.

[13] D.A. Giammona, Ph.D. thesis, Univ. of California, Davis, 1984.

[14] *Kagaku-binran* (Handbook of Chemistry), Maruzen, The Chemical Society of Japan, edited, Basic Volume, 4th ed. II, 1993, p27 (in Japanese).

[15] T. Lian, B. Locke, Y. Kholodenko, R.M. Hochstrasser, J. Phys. Chem. 98 (1994) 11648.

[16] X. Ye, A. Demidov, F. Rosca, W. Wang, A. Kumar, D. Ionascu, L. Zhu, D. Barrick, D. Wharton, P.M. Champion, J. Phys. Chem. A 107 (2003) 8156.

[17] W.L. Jorgensen, J. Chandrasekhar, J.D. Madura, R.W. Impey M.L. Klein, J. Chem. Phys. 79 (1983) 926.

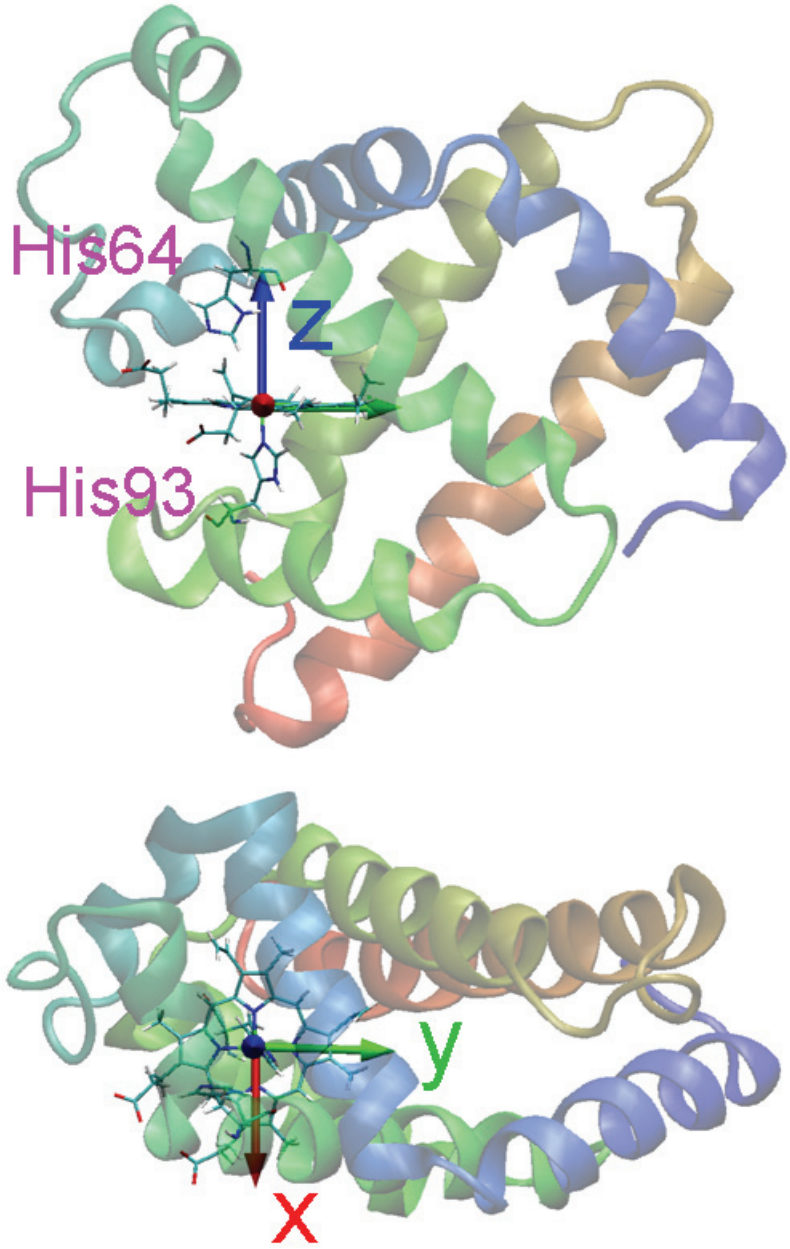

Fig.3.1. The coordinate system used to calculate the R_{gx}, R_{gy} and R_{gz}. The *xy*-plane is nearly on the heme plane and the *z*-axis is perpendicular to it.

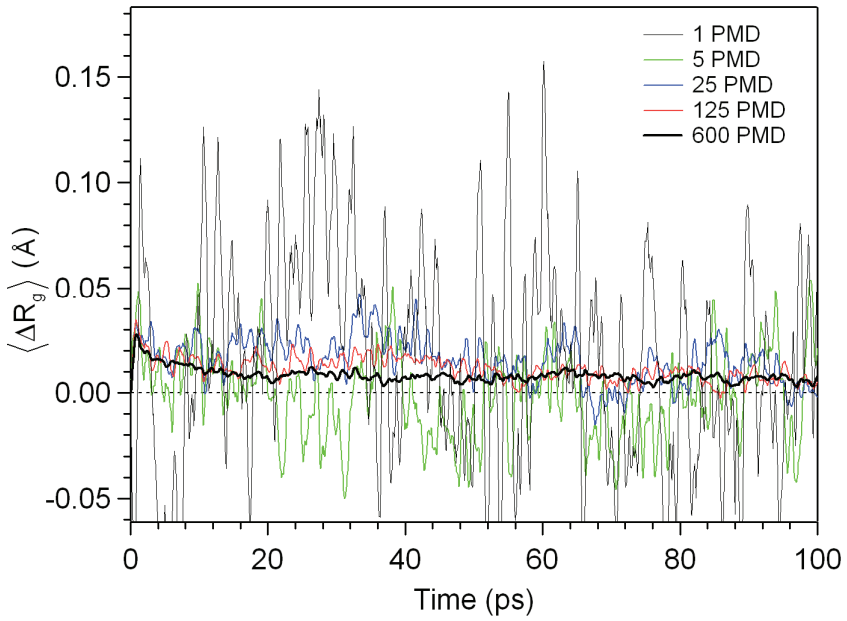

Fig.3.2. R_g changes of a single PMD trajectory (1 PMD) and ensemble-averaged over 5, 25, 125, and 600 PMD trajectories (5 PMD, 25 PMD, 125 PMD, and 600 PMD).

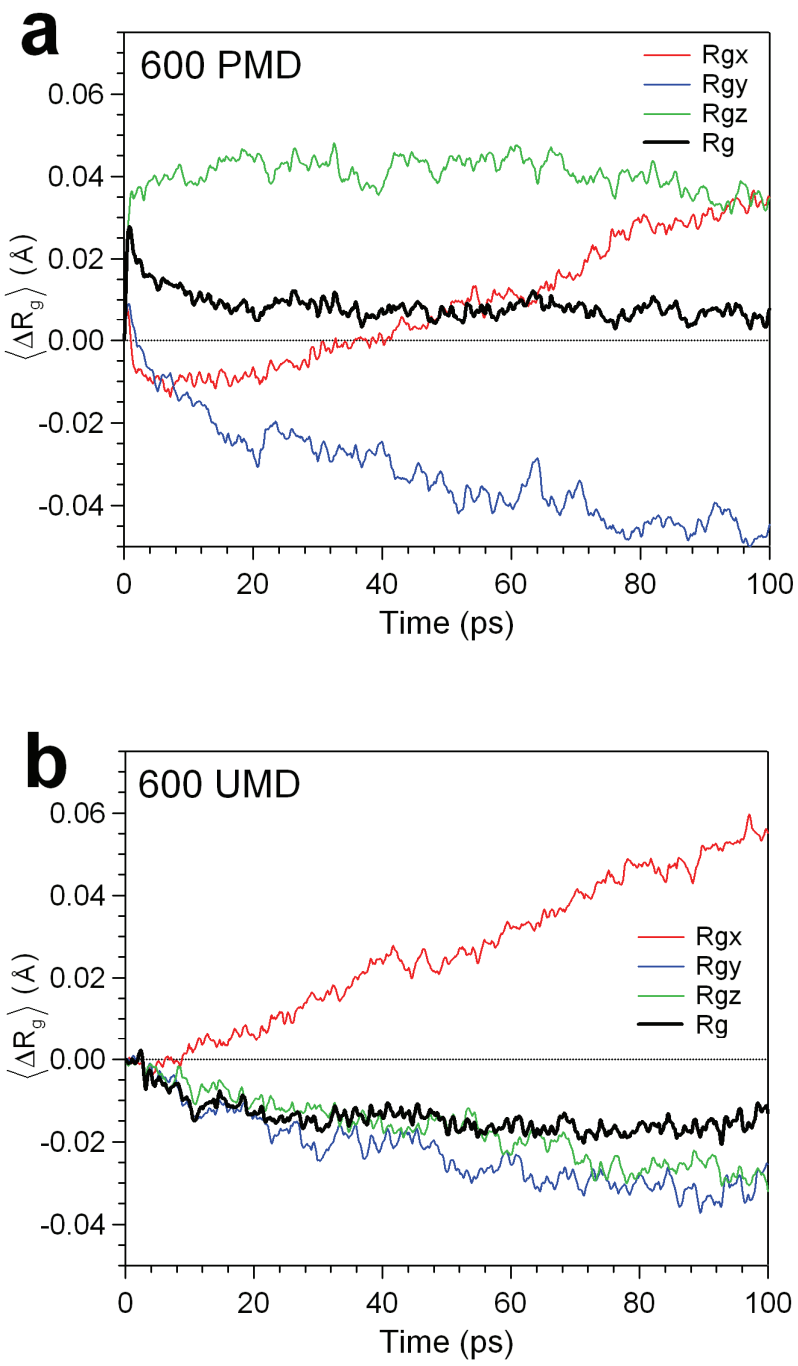

Fig.3.3. R_gs changes ensemble-averaged over (a) 600 PMD and (b) 600 UMD trajectories.

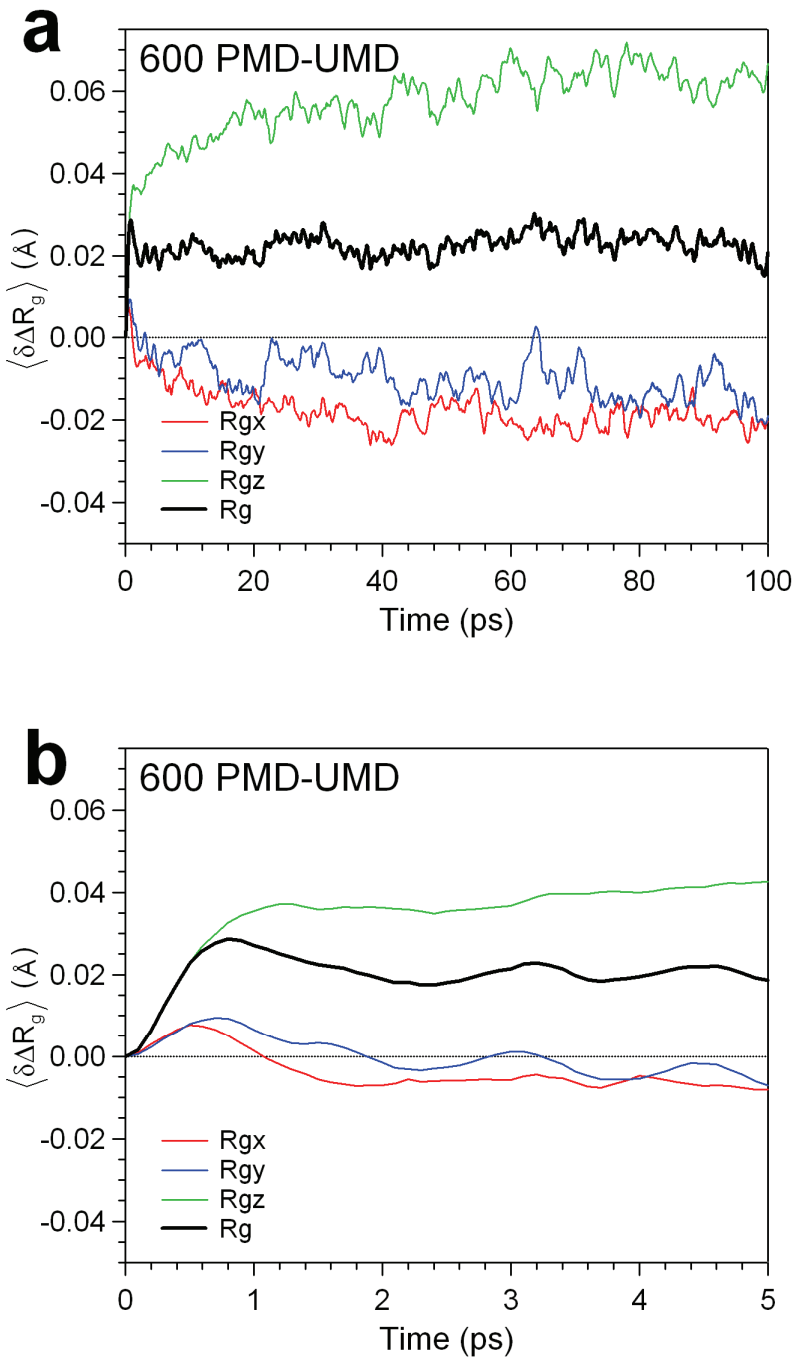

Fig.3.4. R_gs changes calculated by the PE method using the 600 pairs of PMD and UMD trajectories. (a) 0 – 100 ps and (b) 0 – 5 ps.

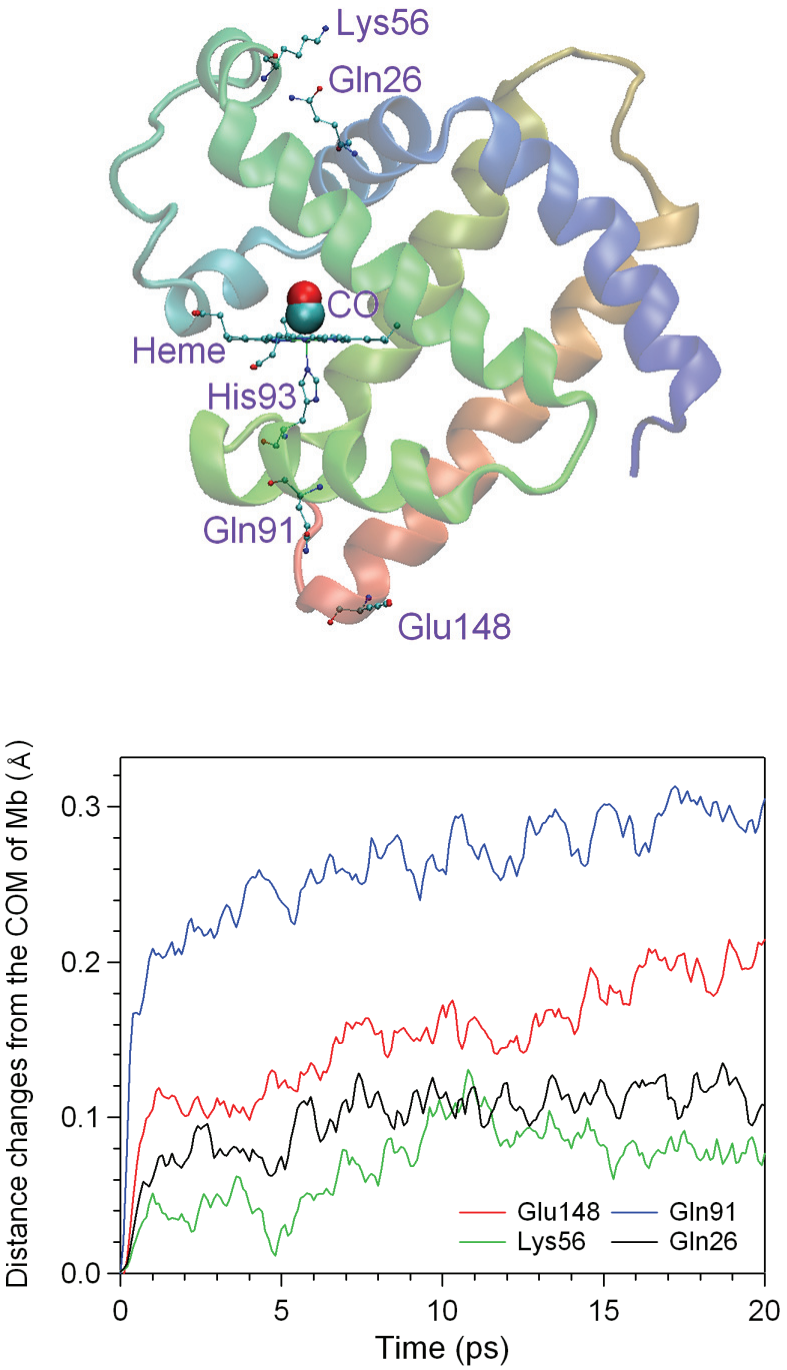

Fig.3.5. Distance changes between the COM of Mb and those of four exposed residues located in the perpendicular direction relative to the heme plane. The changes are calculated by the PE method with 600 pairs of PMD and UMD trajectories.

Table 3.1 Ensemble-averages, standard deviations (SD) and two-sided 95% confidence intervals (95% CI) of the radius of gyration (R_g) and its three Cartesian components (R_{gx}, R_{gy}, and R_{gz}) at t = 0 ps calculated from the 600 PMD trajectories. All the numerical values are in Å.

	R_{gx}	R_{gy}	R_{gz}	R_g
average	6.247	10.313	9.680	15.464
SD	0.129	0.112	0.103	0.056
95% CI	± 0.010	± 0.009	± 0.008	± 0.004

Table 3.2 R_g changes calculated by the PE method from N_{traj} pairs of PMD and UMD trajectories at 1 ps and 100 ps after the photolysis. The 95% CI of each value are also shown. All the numerical values are in Å.

1 ps				
N_{traj}	$\langle \delta\Delta R_{gx} \rangle$	$\langle \delta\Delta R_{gy} \rangle$	$\langle \delta\Delta R_{gz} \rangle$	$\langle \delta\Delta R_{g} \rangle$
5	0.009 ± 0.050	0.014 ± 0.043	0.045 ± 0.021	0.040 ± 0.019
25	0.003 ± 0.012	0.001 ± 0.011	0.042 ± 0.009	0.028 ± 0.008
125	0.000 ± 0.004	0.005 ± 0.004	0.037 ± 0.004	0.026 ± 0.004
600	0.001 ± 0.002	0.007 ± 0.002	0.036 ± 0.002	0.027 ± 0.002
100 ps				
N_{traj}	$\langle \delta\Delta R_{gx} \rangle$	$\langle \delta\Delta R_{gy} \rangle$	$\langle \delta\Delta R_{gz} \rangle$	$\langle \delta\Delta R_{g} \rangle$
5	-0.109 ± 0.231	-0.029 ± 0.125	0.110 ± 0.220	0.022 ± 0.091
25	-0.031 ± 0.089	-0.054 ± 0.062	0.122 ± 0.074	0.028 ± 0.033
125	-0.008 ± 0.044	-0.049 ± 0.040	0.060 ± 0.035	0.016 ± 0.014
600	-0.021 ± 0.017	-0.019 ± 0.015	0.067 ± 0.015	0.021 ± 0.007

Chapter 4

General Conclusion

Proteins govern all the biological mechanisms, and so their functions widely range, e.g., catalytic reaction, signal transduction, mechanical work, and so on. However, because they are very huge and structurally complex molecules, it is difficult to understand their functions from the atomistic point of view. As a first step to elucidate a function of a certain protein, its folded or native structure should be clarified. Although many native structures of proteins have already been clarified experimentally by X-ray crystallography or NMR spectroscopy, there still remain a number of their structures unknown. For such proteins, their structures could be predicted in principle by computational methods, however the accuracy of the current predictions is still limited because of the lack of computational resources. In addition, since in vivo folding processes are usually assisted by ribosomes, chaperones, or cotranslational folding process, the prediction becomes further more difficult. In particular, to shed light on the cotranslational folding processes, I have theoretically analyzed the stabilities of N-terminal partial Mb chains in chapter 2.

Moreover, even if the equilibrium folded structure might be obtained, there would be still an important problem to understand the protein dynamics ruling its function. That is to say, proteins usually respond on external inputs such as the substrate binding, thermal or photo-excitation, and so on and then switch on to function themselves. During the process expressing the function, they do not stay at their stable folded structures but move around them flexibly. Hence, in order to understand the functions from a microscopic point of view, it is indispensable to investigate time-dependent nonequilibrium properties of the functioning proteins. Unfortunately, however, it is impossible to observe these microscopic motions directly by experiments due to the limitations of spatial and time resolutions, and thus the complete understanding is not yet at hand.

On the other hand, it is a methodological advantage of the MD technique that the nonequilibrium trajectories can be directly computed at the atomic level by solving the equations of motion. Therefore, in chapter 3, I have tried to theoretically study the "nonequilibrium" protein dynamics by using MD simulations. The ligand photolysis of MbCO was chosen as a model nonequilibrium system because a large number of experiments have been reported for this phenomenon thus I can verify my computational results.

In the standard procedure, if the system is in equilibrium, it is usual to execute a single or a few long MD simulations to take time-averages to predict physical quantities. However, since the changes triggered by the MbCO ligand photolysis show time-dependent nonequilibrium characteristics and are smaller than thermal fluctuations at room temperature, it is impossible to obtain significant information only from a single MD trajectory. Under the circumstances, the subtle changes of MbCO have been analyzed via the perturbation ensemble (PE) method, in which an extensive number of MD simulations were executed for the time-dependent ensemble averaging. It became possible for the first time to calculate time-dependent changes very accurately driven by the ligand photolysis of MbCO as the perturbation with a statistically-meaningful precision. Since I have obtained successfully the theoretical estimation which is in good agreement with the experimentally measured value, it was confirmed that the procedure of the PE method is valid.

In this thesis, first, in chapter 2, structural stabilities of N-terminal peptides from apoMb were investigated to study the behavior of nascent chains during their elongation processes. Starting from the crystal structure, MD simulations of those chain lengths 36, 77, 92 and 119 chain lengths were executed for 10 ns. As a result, those chains longer than 77 residues are found to keep the α-helical initial structures.

This suggests that a nascent apoMb chain starts to form helical structures after its elongating to the length with at least 77 amino acid residues. Further, by analyzing the residue-residue native contacts, it was investigated at the same time that the cotranslational folding pathway of Mb would be different from the posttranslational one.

In chapter 3, the PE method, in which a lot of pairs of perturbed and unperturbed MD simulations were executed for ensemble average, was applied to elucidate global subtle structural changes after the photodissociation of MbCO. Since the statistically-meaningful theoretical results showed clearly the anisotropic expansion with respect to the heme plane, the applicability of the PE method was examined justifiably.

The application of the PE method is, of course, not limited to the process of MbCO photolysis. For instance, the allosteric effect of Hb, whose four subunits are very similar to Mb, is regulated by the quaternary structural changes that are associated with the ligand binding and dissociation. It is anticipated that, by applying the same procedure as that for Mb, the detailed analysis is possible to shed light on the functionally important structural dynamics. For examples, the PE procedure can also be applied to motor proteins, which convert the chemical energy of ATP into mechanical work, and to membrane proteins which mediate signal transduction by ligand binding or dissociation.

Finally, I would like to emphasize, in this thesis, that the theoretical analyses with MD simulations have clarified successfully the microscopic properties of Mb in nonequilibrium states. It is convinced that further applications of the present procedures should expand the scope of computational and theoretical methodologies toward the mesoscopic world where biomacromolecules should play very important

roles.

Acknowledgments

The present thesis is the summary of my researches in the Graduate School of Information Science, Nagoya University.

I would like to thank many people who together completed this thesis during their supportive cooperation with me. First, I would like to express my deep gratitude to Prof. Masataka Nagaoka who gives me kind guidance and continuous encouragement for instructive discussions and initiation into the research publication. I would like to thank also Prof. Nobuaki Koga, Prof. Motonori Ota, Prof. Masaki Sasai, Prof. Sei-ichiro Tenno, and Prof. Koji Yasuda who instructed me in theories of protein dynamics and electronic states. I thank also Dr. Isseki Yu, Mr. Hiroto Okumura, Ms. Chiharu Iwahashi and other collaborators in Nagaoka's laboratory for their many important advices and discussions. Finally, I would like to thank my parents, brother and sister for their "watching over" fondly and supporting me during my life of research.

This thesis was partially supported by a Grand-in-Aid for the 21st Century COE program 'Frontiers of Computational Science'.

Publication List

1. M. Takayanagi, I. Yu, M. Nagaoka, “Theoretical study on the stabilities of N-terminal partial chains from apo-myoglobin”, Chem. Phys. Lett. 421 (2006) 300.

2. M. Takayanagi, H. Okumura, M. Nagaoka, “Anisotropic Structural Relaxation and Its Correlation with the Excess Energy Diffusion in the Incipient Process of Photodissociated MbCO: High-Resolution Analysis via Ensemble Perturbation Method”, J. Phys. Chem. B 111 (2007) 864.

3. I. Yu, M. Takayanagi, M. Nagaoka, “Intrinsic Alteration in the Partial Molar Volume on the Protein Denaturation: Surficial Kirkwood-Buff Approach”, J. Phys. Chem. B, in revision.

4. M. Nagaoka, I. Yu, M. Takayanagi, "Energy Flow Analysis in Proteins via Ensemble Molecular Dynamics Simulations: Time-Resolved Vibrational Analysis and Surficial Kirkwood-Buff Theory" in D.M. Leitner and J.E. Straub, Eds., "Proteins: Energy, Heat and Signal Flow", CRC, Boca Raton, 2009.