学 位 論 文

Experimental and simulation studies on folding of goat
α-lactalbumin

（ヤギα-ラクトアルブミンのフォールディングの実験とシミュレーション
による研究）

平成11年12月博士（理学）申請

東京大学大学院理学系研究科
物理学専攻

依 田 隆 夫

Experimental and simulation studies on folding of goat α-lactalbumin

by
Takao Yoda

Department of Physics, School of Science, University of Tokyo,
Hongo, Bunkyo-ku, Tokyo, 113-0033, Japan

February, 2000

Table of contents

Introduction

Protein folding

The protein is a functional biopolymer that plays essential roles in various biological phenomena. The specific tertiary structure of a protein molecule is indispensable for the biological function. Such specific three dimensional structure that brings about the biological function of the protein is called the native structure. Various conditions (e.g. high concentration of a strong denaturant such as urea or GdnHCl, high temperature, extreme pH that is far from the physiological one) are known to disrupt the native stereoregular structure of the protein and lead the protein molecule to an unfolded polypeptide.

In early 1960's, Anfinsen and his co-workers showed by their *in vitro* experiments that fully unfolded ribonucrease A, which was reduced and unfolded by addition of a reductant (2-mercaptoethanol) a strong denaturant (urea), refolded to the native structure and recovered the full enzymatic activity when the reductant and the denaturant were removed. Based on this observation, they have proposed that the amino-acid sequence of a protein contains all the information required for folding the unfolded polypeptide into the native structure of the protein (the Anfinsen dogma). Many small proteins are known to reversibly unfold as observed in ribonucrease A. The protein folding *in vivo* often requires molecular chaperones, and this apparently seems contradicting the Anfinsen dogma. However, it is generally accepted that the presence of the molecular chaperones does not break the generality of the Anfinsen dogma because the function of the molecular chaperons is to prevent the aggregation of the protein under a cellular condition with

high protein concentration and because the molecular chaperones themselves do not convey any structural information required for the folding.

The Anfinsen dogma indicates that we can define the thermodynamic states, the native and unfolded states, of a protein. When the protein is folded and has the specific native structure it is in the native state, and the protein is unfolded, it is in the unfolded state.

The folding reactions of certain small proteins with less than about 100 amino acid residues have been shown to be represented by a two-state mechanism, in which only the native and the unfolded states are involved. However, larger proteins fold through intermediate states. Such intermediates can be detected by kinetic refolding experiments and occasionally also observed under mildly unfolding conditions in equilibrium unfolding experiments. Thus, it is important for fully understanding the mechanism of protein folding to investigate the intermediates of folding. Various stopped-flow techniques including stopped-flow circular dichroism (CD) and stopped-flow hydrogen-exchange pulse labeling combined with two-dimensional NMR spectroscopy, have successfully been used for detecting and characterizing the transient kinetic folding intermediates.

The molten globule (MG) state is a typical folding intermediate observed in many globular proteins (Arai & Kuwajima, 2000). The common structural characteristics of the molten globule sate are (1) the presence of a pronounced amount of secondary structure, (2) the absence of most of the specific tertiary structure produced by tight packing of side chains, (3) compactness of a protein molecule with a radius of gyration 10 to 30 % larger than that of the native state, and (4) the presence of a loosely packed hydrophobic core that increases the

2

hydrophobic surface accessible to solvent. The MG folding intermediate has been observed in most of the proteins with more than 100 amino acid residues whose folding reaction has been well characterized. The degree of the structural organization and the stability of the MG state have both been found to be remarkably dependent on protein species. Therefore, there is a range of species possible within the basic MG properties of secondary structure, collapse and hydrophobic core. The classic concept of little or no persistent tertiary structure still holds frequently but is not a universal requirement of the molten globules. Thus MG is one typical characteristics widely observed in the intermediates of protein folding.

After the analogy of the transition-state theory of absolute chemical reaction rate (Glasstone et al., 1941), the transition state of protein folding is the state of maximum free energy that may exist along the reaction pathway between the native and the unfolded states. For proteins that show the MG folding intermediate, the transition state usually exists between the native and the intermediate state.

Therefore, characterization of the transition state is also an important issue for understanding the protein folding mechanism. As shown by Fersht and his co-workers, combination of the protein engineering methods and the kinetic folding/unfolding experiments provides us a way to study the structure of the transition state of protein folding. The ϕ-value analysis that is carried out by comparison of the folding/unfolding rates between the wild type and various mutants of a protein allows us to map out the transition state structure (see materials and methods). The transition state of folding of chimotripsin inhibitor 2 (CI2), which is a small protein with 64 amino acid residues and rapidly folds by the two-state mechanism, and barnase, which contains 110

3

amino acid residues and folds by a multi-state mechanism, was characterized by this method. For many other globular proteins, the transition state of folding has not been well characterized. However it must be kept in mind that the structure of the transition state is inferred rather than measured directly.

Computer simulation methods have been effectively applied for studying the protein folding problem. Among various simulation approaches, the molecular dynamics simulations of a protein molecule with solvent water molecules is particularly useful for elucidating the atomic details of the unfolding process.

In principle, the molecular dynamics simulation of refolding might be possible if the available forcefield parameters are accurate enough. However, with the available methodology and computing power, such a simulation would require unrealistic computing time. With the teraflop supercomputers that are on the horizon, a folding simulation for a 100-residue protein where the experimental transition to the folded state takes place in about 1 sec would take 10^3 years. Therefore we shall use, at present, the molecular dynamics simulation at high temperature to investigate the unfolding process of a protein. The use of the high temperature (usually > 400 K) accelerates the unfolding rate, which results in a reasonable simulation time. If the unfolding process simulated is regarded as the inverse of folding, such a simulation will give us abundant information about protein folding.

To investigate the effectiveness of such a simulation approach, comparison of the real folding/unfolding in solution and the unfolding in the computer simulation of a typical model protein is required.

α-lactalbumin

4

α-lactalbumin (LA) is a globular protein, containing 123 amino acid residues and having a molecular weight of 14,200. It is in the whey of mammalian milk, and acts as a regulatory subunit of a lactose synthase system (Hill & Brew, 1975; McKenzie & White, 1991). The crystal structures of LA from various species (authentic baboon, human, goat and guinea-pig LA and recombinant bovine and goat LA) were determined (Acharya et al., 1991; Acharya et al., 1989; Chaudhuri et al., 1999; Pike et al., 1996), and those are highly homologous to the three-dimensional structure of c-type lysozyme. The structure of LA contains two domains, the α-domain and the β-domain. The α-domain contains three or four α-helices, depending on the species, and a 3_{10} helix, while the β-domain contains an anti-parallel β-sheet and a 3_{10} helix. LA is a calcium binding metalloprotein (Hiraoka et al., 1980) and the calcium-binding loop is located near the boundary between the two domains.

LA is often chosen as a model for protein folding study. A major reason for this is that this protein exhibits a molten globule (MG) folding intermediate stably populated in equilibrium under various mildly denaturing conditions, including those at acid pH and at an intermediate concentration of a strong denaturant (e.g., guanidine hydrochloride (GdnHCl)) at neutral pH, and several groups have reported studies of the MG state of LA (Kuwajima, 1996). The MG state of the protein is characterized by formation of native-like secondary structural elements studied by circular dichroism (Kuwajima et al., 1976) and hydrogen-exchange NMR measurements (Baum et al., 1989; Alexandrescu et al., 1993; Chyan et al., 1993; Schulman et al., 1995; Forge et al., 1999), non-cooperative unfolding transition studied by diferential scanning calorimetry (Yutani et al., 1992), non-native

5

clustering of aromatic side chains in the α-domain of the molecule studied by NMR (Alexandrescu et al., 1993) and hydrophobic stabilization of the native-like fold of the α-domain studied by local effective concentration of thiol groups in the disulfide-bridge formation (Song et al., 1998; Wu & Kim, 1998). In short, the MG state of LA has native-like secondary structure and native-like tertiary fold in the α-domain with compact size of the whole molecule (Kataoka et al., 1997) but without the native specific tertiary side-chain packing.

In kinetic refolding reactions of LA, a burst-phase intermediate accumulates within a typical dead time of a stopped-flow apparatus, and this burst-phase intermediate is known to be identical with the MG state observed at equilibrium (Ikeguchi et al., 1986a; Arai & Kuwajima, 1996). Thus, in the kinetic folding, the LA molecule acquires compactness and other partially native-like characteristics of the MG state mentioned above rapidly, and the succeeding stage of folding to acquire the native tertiary side-chain packing is a much slower process.

Here I focus on the role of a hydrophobic core of the α-domain in the stability and folding of goat LA. I have studied experimentally the native-state stability of goat LA mutants, in which the mutations are introduced into residues located at the hydrophobic core on the B-helix, as well as the kinetic folding and unfolding of the wild-type protein and the two mutants (T29I and A30I+T33I). A question I address in this thesis is how the residues at the hydrophobic core of the molecule behave at the final stage of the folding when the molecule acquires the native structure with the hydrophobic tight packing of side chains. This question is important in understanding the folding mechanism of LA because these residues are located on the hydrophobic part of the B-helix that is known to be a stable helical element in the MG state.

6

The experimental results of the present study have, however, shown that in the transition state of the folding of LA, the region around the Thr29 side chain does not acquire the native interactions that are perturbed by the amino acid substitution of isoleucine for threonine. To obtain more detailed interpretations of the final stage of the folding reaction of LA, I have studied trajectories of the LA molecule obtained by molecular dynamics simulations at denaturing temperatures (400 K and 600 K). I have first analyzed the conformational changes observed in the region around Thr29 side chain as well as the N-terminal region for which our previous study (Chaudhuri et al., 1999) have shown experimentally the absence of the native interactions in the transition state of folding. I show that structural disruptions observed in the above regions in the simulations are qualitatively consistent to the experimental data. Loss of the side-chain contacts around Thr29 with penetration of water molecules and unfolding of the N-terminal region with loss of hydrogen bonds have been observed, and they structurally interpret the absence of the native interactions in the transition state observed experimentally. I have also analyzed structural changes in other regions of the molecule along the 400-K trajectory and the six different 600-K MD trajectories. While the molecule undergoes structural change in different manners, the above features at the initial stage of the unfolding are reproducible to some extent among the different trajectories studied.

Materials and Methods

Materials

The expression system of goat α-lactalbumin was previously described (Uchiyama et al., 1995). The recombinant goat α-lactalbumin and its mutants were expressed as inclusion bodies. The expressed proteins were purified as follows. First, the inclusion bodies were solubilized by 20 mM Tris, 5 mM EDTA, 0.1 M DTT and 8 M urea at pH 8.5. The solution was then applied to a DEAE Sepharose FF column equilibrated by 20 mM Tris, 1 mM EDTA, 5 mM DTT and 7 M urea, pH 8.5, and then proteins were eluted by linear gradient of NaCl concentration from 0 to 0.3 M. The oxidative refolding of goat α-lactalbumin was performed by dialysis against a refolding buffer (10 mM Tris, 50 mM NaCl, 1 mM CaCl2, 2.5 mM GSH and 0.5 mM GSSG, pH 8.5) at 4 ℃. The progress of the oxidative refolding was monitored by reversed phase HPLC using a μBONDASPHERE 5μ C4 300 Å column. After stopping the oxidative refolding by decreasing the pH of the solution to pH 5.5 by adding acetic acid, the precipitate was removed by centrifugation, and the protein was purified by ammonium sulfate salting out. Finally, the misfolded proteins were removed by hydrophobic interaction chromatography under the condition that was previously described (Lindahl & Vogel, 1984; Uchiyama et al., 1995). The purified proteins were dialyzed against water, and lyophilized.

GdnHCl used for measurements and urea used for protein purification were specially prepared reagent grade for biochemical use from Nakarai Tesque, Inc. All other chemicals were guaranteed reagent grade. Concentrations of recombinant goat α-lactalbumin were determined spectrophotometrically using an extinction coefficient of $E^{1\%}_1$

$_{cm}$ = 20.1 (Kuwajima *et al.*, 1980). The concentration of GdnHCl was determined from the refractive index at 589 nm with an Atago 3T refractometer (Pace, 1986).

Equilibrium CD measurements

Equilibrium CD measurements were carried out in a Jasco J-720 spectropolarimeter, using optical cuvette with 1-mm and 1-cm path lengths for the peptide region of wave length and the aromatic region of wave length respectively. The temperature of the cuvette was controlled by circulating water at 25 ℃. The equilibrium unfolding transition was measured by CD in 50 mM sodium cacodylate, 50 mM NaCl and 1 mM CaCl$_2$ (pH 7.0) with different concentrations of GdnHCl. The concentration of the protein was 0.33-0.42 mg/ml.

Stopped-flow CD measurements of refolding and unfolding

Kinetic refolding and unfolding measurements were carried out using a stopped-flow apparatus attached to the Jasco J-720 spectropolarimeter (Kuwajima *et al.*, 1987). A cell with 1- or 4-mm optical pathlength was used. The all solutions contained 50 mM sodium cacodylate, 50 mM NaCl, and 1 mM CaCl$_2$. The protein solutions before stopped-flow mixing also contained 1.0 or 5.5 M GdnHCl for unfolding or refolding respectively. The concentration of goat LA in these solutions was 0.8-1.5 mg/ml. The reactions were initiated by mixing the protein solution to with unfolding/refolding buffer containing 50 mM sodium cacodylate, 50 mM NaCl, 1 mM CaCl$_2$ and an appropriate concentration of GdnHCl, with the mixing ratio of 1:10. All kinetic measurements were

9

carried out at pH 7.0 and 25 ℃ controlled with circulating water.

Figure 2 shows a block diagram of our stopped-flow CD system and a diagrammatic representation of the stopped-flow mixer and the observation cell. The structure enclosed by a broken line in the block diagram is a Jasco J-720 CD spectropolarimeter. The stopped-flow mixing apparatus is composed of a mixing driver controlled by a three-way magnetic valve, drive syringes, reservoirs, a mixer, and an observation cell. Two solutions in the driving syringes were mixed in the mixer to initiate the refolding/unfolding reaction, and the time dependent CD signal of the solution in the observation cell was monitored.

The ϕ-value analysis of kinetic refolding/unfolding data

After the analogue of the transition state theory for chemical reactions, the activation free energy of folding/unfolding is related to the rate constant that can be derived from the kinetic experimental data of refolding/unfolding as

$$k = a \exp(-\Delta G^{\ddagger}/RT)$$

where k is the microscopic rate constant, $\Delta G^{\ddagger}$ is the activation free energy, R is the gas constant, T is the absolute temperature, and a is a constant. Thus, the change in the activation free energy caused by an amino acid substitution, $\Delta\Delta G^{\ddagger}$ is given by

$$\Delta\Delta G^{\ddagger} = -RT \ln(k_{mutant} / k_{WT})$$

10

where k_{mutant} and k_{WT} are the rate constants for the mutant and the wild type, respectively.

To obtain the structural information about the folding transition state, the ϕ-value analysis is usually adopted. In this method, the protein engineering technique and the equilibrium and the kinetic measurements of folding/unfolding are combined. Although most of the proteins that the ϕ-value analysis has been applied are small proteins which fold by the two-state reaction, this method has also be applied to proteins that fold through an intermediate state. Here I define ϕ_U as

$$\phi_U = \Delta\Delta G^{\ddagger}_U \; / \; \Delta\Delta G_U$$

where $\Delta\Delta G^{\ddagger}_U$ is the change in the unfolding activation free energy caused by the amino acid substitution, and $\Delta\Delta G_U$ is the change in the free energy of equilibrium unfolding caused by the amino acid substitution.

Figure 3 shows the free energy diagrams for the two extreme cases ($\phi_U = 1$ and $\phi_U = 0$). The free energy of the unfolded state of the wild type and that of the mutant are superposed in the figure. A ϕ_U value of 1 means that the free energy of the transition state is perturbed on mutation by the same amount as the unfolded state is perturbed. A ϕ_U value of 0 means that the free energy of the transition state is perturbed on mutation by the same amount as the native state is perturbed. Thus, it is inferred that a ϕ_U value of 1 means that the structure is unfolded at the mutation site as much as it is in the unfolded state, and a ϕ_U value of 0 means that the structure is folded at the mutation site as much as it is in the native state.

Because the $\Delta\Delta G_U$ value is most accurately evaluated at the transition midpoint concentration of GdnHCl, I performed the ϕ-value

analysis at the midpoint GdnHCl concentration of the equilibrium unfolding transition of wild-type goat LA.

In the two-state model, the apparent rate constant directly observed in the stopped-flow measurements is the sum of the refolding rate constant and the unfolding rate constant. For most globular proteins logarithms of both the refolding and unfolding rate constants are approximately linearly related to the denaturant concentration. Because the kinetic unfolding rate constant is much larger than the refolding rate constant in the unfolded region at high concentrations of GdnHCl, the apparent rate constant is equivalent to the unfolding rate constant under the condition. The logarithmic unfolding rate constant (or the activation free energy of unfolding) at the transition midpoint concentration of GdnHCl can be evaluated using the extrapolating linearly the logarithmic apparent rate constant in the unfolded region to the transition midpoint.

For the folding reactions represented by the three-state model,

$$\begin{array}{ccc} K & & k_f \\ U & \rightleftarrows MG \rightleftarrows & N \\ & & k_u \end{array}$$
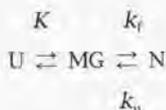
where K is the equilibrium constant between U and MG, and k_f and k_u are the folding and the unfolding rate constants, respectively. In this model, the pre-equiribria between U and MG is achieved quickly after initiation of the folding reaction. Thus, the apparent rate constant (k_{app}) is related to K, k_f and k_u as

$$k_{app} = k_u + k_f \left(K/(1+K) \right).$$

Because $k_f \ll k_u$ and $K > 0$ in the unfolded region, k_{app} again provides a good approximation of k_u at high concentrations of GdnHCl. Thus, the ϕ_U value can be estimated in the same manner.

Molecular dynamics simulations

For both of 400-K and 600-K simulations, a crystal structure of wild-type recombinant goat LA at 2.0 Å resolution with water molecules of crystallization was used as the initial structure (Chaudhuri et al., 1999). Because the crystal structure does not have the coordinates for the C-terminal three residues, these three residues were added to the structure used for the simulations. Because the pH of the crystallization was pH 6, Asp and Glu residues were made negatively charged, and Arg and Lys were made positively charged. Although determining the protonation of the three histidines was not simple, these three histidine residues were made positively charged. The N-terminal amino and C-terminal carboxyl groups were positively charged and negatively charged respectively. The resulting LA molecule contains 1,972 atoms including the hydrogen atoms and the calcium ion.

The molecular dynamics simulation at 400 K was carried out using the COSMOS90 software with an AMBER all-atom force field by the PPPC method without Coulomb-interaction cut-off (Cornell et al., 1995; Saito, 1992; Saito, 1994). The simulation procedure was similar to that previously described (Saito, 1999). The protein molecule and the crystal water molecules were immersed within a sphere of a radius of 34 Å with water molecules. SPC water molecules were used and the total number of the water molecules in the system was 4,737. An

equilibration for 600 ps at 300 K was done. For the first 10 ps of the equilibration, position constraining of the heavy atoms in the protein molecule and the crystal water oxygen atoms was applied. The temperature of the system was then increased to 400 K, and a 6-ns simulation was carried out at 400 K. The radius of the spherical system was increased from 34 Å to 36 Å simultaneously with the elevation of the temperature. The bound calcium ion was not removed during the simulation. The time step was 0.5 fs for both the equilibration and 400-K simulations. The coordinates of the protein molecule and those of oxygen atoms of water molecules were saved every 1000 steps.

Molecular dynamics simulations at 600 K were carried out using the AMBER5.0.1 program package with the AMBER all-atom force field (Case et al., 1997; Cornell et al., 1995). The protein molecule and the crystal water molecules were put within a periodic-boundary box with solvent water molecules. The size of the rectangular periodic-boundary box was determined so that the boundaries are distant more than 10 Å from any solute atom. TIP3P (Jorgensen et al., 1983) water molecules were used and the total number of the water molecules was 7,502. A twin-range non-bond energy cutoff (van Gunsteren & Berendsen, 1990) (12 Å and 16 Å) method was used here. After a conjugate gradient energy minimization of 500 steps, a 420-ps NPT molecular dynamics simulation was done at 298 K and 1 atm. The temperature and the pressure were controlled by coupling with external bath with coupling constants of 0.5 ps and 1.0 ps respectively (Berendsen et al., 1984). The six molecular dynamics simulations at 600 K were started from the 300, 320, 340, 360, 400 and 420 ps conformations of the 298-K simulation. The temperature was gradually increased to 600 K, taking about 3 ps, with fixed volume of the periodic boundary box. After that, NVT

molecular dynamics simulations at 600 K were performed for 150, 210 or 250 ps. The time step of the simulation was 1 fs, and SHAKE (Ryckaert *et al.*, 1977) was used (tolerance = 10^{-4}). The list of the pairs for the non-bond energy calculation was updated every 10 and 5 steps in the 298-K and 600-K simulations respectively. The coordinates of all atoms in the system were saved every 500 steps.

I estimated the contacts between residues in the protein molecule. Two residues were defined as to be contacted if any two non-hydrogen atoms of the residues are within 4.5 Å. An inter-residue contact was defined as to be native if the two residue involved in the contact were contacted in more than the half of the 400 conformations in the final 200 ps of the 300-K equilibration. In Fig. 9, 15, and 16, only the native inter-residue contacts whose participants are distant by more than 5 residues from each other in the primary sequence of goat LA are shown.

Accessible surface area of the protein molecule was calculated using NACCESS (Hubbard & Thornton, 1993). The secondary structural elements were evaluated using DSSP (Kabsch & Sander, 1983). All the other analysises of the trajectories of both of 400-K and 600-K simulations were performed with programs written in FORTRAN77 or FORTRAN90.

15

Results

Equilibrium unfolding transitions of wild-type and mutant goat LA at neutral pH

To investigate the role of a hydrophobic core in the α-domain of goat LA, I carried out equilibrium unfolding experiments of wild-type, four single mutant (T29I, A30T, A30I and T33I), and one double mutant (A30I+T33I) goat LA (Uchiyama et al., 1995). All of these mutation sites (Thr29, Ala30 and Thr33) are within the B-helix. The B-helix is located at the center of the α-domain of the protein (Fig.1).

Figure 4 shows equilibrium unfolding transition curves of the wild-type and mutant proteins. The unfolding transitions were induced by a denaturant, guanidine hydrochloride (GdnHCl), at pH 7.0 in the presence of 1 mM Ca^{2+} and followed by the CD ellipticities at 222 nm and 270 nm. The transition curves are represented by the apparent fractional extents of unfolding, which have been obtained from the observed ellipticity values and are normalized with respect to the values linearly extrapolated from the pre- and the post-transition regions (see below), as a function of GdnHCl concentration. The transition curves measured at the two wavelengths are coincident with each other for all the proteins except those that contain the T33I mutation. The coincidence of the transition curves allows us to analyze the transition by a two-state model, in which only the native and the unfolded states are populated in the transition region.

In the mutant proteins in which Ile is substituted for Thr33, the unfolding transition measured at 270 nm occurs at a slightly lower GdnHCl concentration than the transition measured at 222 nm. This disagreement of the transition curves indicates accumulation of an

unfolding intermediate (the MG state) at an intermediate concentration (~2.4 M for T33I and 3.0 M for A30I + T33I) of GdnHCl (Ikeguchi *et al.*, 1986b; Kuwajima et al., 1976). The T33I mutation has stabilized the MG state (Uchiyama et al., 1995) and destabilized the native state, which has led to the accumulation of the intermediate. However, because the disagreement between the two transition curves at 270 nm and 222 nm is only marginal, I have also employed the two-state model for the analysis of the unfolding transitions of these mutant and used the transition curves measured at 270 nm for the analysis.

The unfolding transition curves measured by the near UV CD (270 nm) for the wild-type and mutant proteins are compared with each other in Fig. 4g. By the use of these transition curves, I carried out thermodynamic analysis of the unfolding transitions on the basis of the two-state model:

$$N \overset{K_U}{\rightleftarrows} U \qquad (1)$$

where N and U denote the native and the unfolded states, respectively, and K_U is the unfolding equilibrium constant, which is related to the unfolding free energy (ΔG_U) as

$$\Delta G_U = -RT \ln K_U \qquad (2)$$

where R is the gas constant and T is the absolute temperature. The ΔG_U is known to linearly depend on the concentration (c) of the denaturant, so that

17   (3)

$$\Delta G_U = m(c_M - c)$$

where c_M is the midpoint concentration of the denaturant, and m represent dependence of ΔG_U on c and characterizes the cooperativity of the transition. I have also assumed that the ellipticities of the native and the unfolded states, θ_N and θ_U, respectively, are both linear functions of c as

$$\theta_N = \theta_1 + \theta_2 c$$
$$\theta_U = \theta_3 + \theta_4 c \tag{4}$$

where θ_1, θ_2, θ_3 and θ_4 are constants. Thus, we finally obtain the observed ellipticity, θ_{obs} as a function of c, as

$$\theta_{obs}(c) = \frac{\theta_1 + \theta_2 c + (\theta_3 + \theta_4 c)\exp\{-m(c_M - c)/RT\}}{1 + \exp\{-m(c_M - c)/RT\}} \tag{5}$$

I performed non-linear least squares fittings of the observed data to this equation. In wild-type goat LA and a mutant T29I, data points at low denaturant concentrations (0 M for WT, less than 0.5 M for T29I) have been excluded from the fitting analysis because there is a small but sharp increase of the ellipticity, which clearly does not fit with equation (5), at these GdnHCl concentrations. The best-fit values of c_M and m thus obtained are summarized in Table 1. The ΔG_u values at zero concentration of GdnHCl ($\Delta G_{U H2O}$) have been obtained by extrapolation of ΔG_U to $c = 0$, and these $\Delta G_{U H2O}$ values are also shown in Table 1.

The difference in ΔG_U between each mutant and the wild-type protein ($\Delta\Delta G_U$) have been calculated by the equation,

18

$$\Delta\Delta G_U = m_{(mutant)}(c_{M\,(mutant)} - c_{M\,(WT)}) \tag{6}$$

where $c_{M\,(mutant)}$ and $c_{M\,(WT)}$ are the c_M values for the mutant and wild type goat LA. The $\Delta\Delta G_U$ values thus obtained for the mutant proteins are also included in Table 1.

From Table 1, the T29I mutation stabilizes the N state of goat LA by 3.5 kcal/mol at the transition midpoint GdnHCl concentration of WT (2.6 M). This stability increase is roughly equivalent to that expected from the increase in hydrophobicity caused by replacement of Ile for Thr (Sharp *et al.*, 1991). The other mutant proteins with increased hydrophobicity, except for T33I, are also stabilized, but the T33I and A30T mutations slightly destabilize the protein.

The changes in stability caused by the amino acid substitutions other than T29I are, however, not remarkable. The change in stability caused by the A30I mutation is much smaller than that expected from the change in hydrophobicity (Table 1). The N state of the T33I mutant is less stable than that of the wild-type protein, and that of the T30I+T33I double mutant is again less stable than that of the T30I single mutant. Thus, the T33I mutation, which increases the hydrophobicity of the amino acid residue, appears destabilizing the protein by the order of 0.1 kcal/mol.

Kinetic refolding and unfolding measurements of goat LA and mutants

To investigate the effects of the mutations within the B-helix on the folding of goat LA, I carried out kinetic folding and unfolding

experiments of the wild-type and the two mutants (T29I and T30I + T33I) of goat LA (Fig. 5). The refolding and the unfolding reactions were induced by concentration jumps of GdnHCl from 5.5 M and 1.0 M, respectively, to the indicated concentrations by the stopped-flow method, and monitored by the far UV CD at 225 nm.

The time course of the observed refolding and unfolding reactions was analyzed by non-linear least squares fitting to the sum of exponentials,

$$|\theta| = A_0 + \sum A_i \exp(-k_i t) \tag{7}$$

where $[\theta]$ is the observed ellipticity at time t, and k_i and A_i are the apparent rate constant and the amplitude, respectively, for the ith kinetic phase.

The kinetic refolding reactions of the wild-type and the two mutant proteins involve up to three exponential phases. The refolding reactions of the proteins at low denaturant concentrations (< 1.0 M) have three phases, while one or both of the two slower phases disappear at a GdnHCl concentration near C_M. The observed unfolding reactions were well fitted to the single-exponential function. Among the three refolding phases, the phase having the largest rate constant (the first phase) is the one whose rate constant is most strongly influenced by the amino acid substitutions. The first phase has the largest amplitude of the three phases (more than 80 % of the observed ellipticity change for the wild type and T29I, as shown in Table 2), and its rate constant remarkably decreases with increasing the GdnHCl concentration (Fig. 6), indicating that this phase reflects a real folding reaction. The observed amplitude

20

of refolding of the double-mutant (A30I+T33I) is remarkably smaller than that of the wild type. This is due to the increased stability of the MG state caused by the mutation (Uchiyama *et al.*, 1995). The slowest phase (the third phase), however, shows little denaturant concentration dependence, suggesting that this phase may be due to an isomerization phenomenon such as cis-trans isomerization of either of the two proline residues in goat LA. On the other hand, the rate constant of the second phase slightly dependent on the GdnHCl concentration. In this thesis, I focus mainly on the first phase for the purpose of investigating the changes in the energetics of folding caused by the amino acid substitutions (T29I and A30I + T33I).

Effect of amino acid substitutions to kinetic constants

The T29I substitution significantly affects only the unfolding kinetics (Fig. 6). The unfolding rate constant of this mutant is more than 40 times smaller than that of the wild-type protein, while this substitution leads to only a small increase (1.5-fold) in the refolding rate constant. This indicates that the protein molecule acquires little native interaction around the side chain of the mutation site in the transition state of folding (see Discussion).

The T30I + T33I double mutation does not affect the refolding nor the unfolding rate constants remarkably. This is as expected since the stabilization of the N state by the double mutation is only marginal and similar in extent to the stabilization of the MG state (Uchiyama et al., 1995) brought about by the same mutation.

Apparent kinetic refolding constants of slow phases are not affected largely by these two amino acid substitutions.

Molecular dynamics simulations of goat LA at high temperatures

I carried out molecular dynamics simulations of the protein under unfolding conditions (Kazmirski & Daggett, 1998; Li & Daggett, 1998; Tsai *et al.*, 1999; Fulton *et al.*, 1999). The purpose of my high-temperature simulation study is (1) examining whether or not the simulation method provides the results consistent with those of experiments, and (2) investigating the structural changes upon unfolding of goat LA at an atomic level of resolution. I particularly focus on the changes in the structure near the Thr29 side chain and of other portions of the α-domain of goat LA, and the results will be compared with the experimentally observed refolding and unfolding kinetics described above. Because the practical simulation time is much shorter than the time required for the simulations under usual unfolding conditions at room temperature, I employed high temperature conditions to accelerate the unfolding reaction, and denaturant molecules (Tirado-Rives *et al.*, 1997; Caflisch & Karplus, 1999) were not included in the system of simulation. To make the simulation as realistic as possible, a relatively low temperature, 400 K, was chosen as an unfolding temperature. Because it was not possible to perform more than one 400-K simulations, I additionally performed 600-K simulations. The initial structure of the protein was taken from the X-ray coordinates of the recombinant goat LA molecule (Chaudhuri et al., 1999) with the N-terminal methionine in both the 400-K and 600-K simulations.

Six-ns molecular dynamics simulation of goat LA at 400 K

The system of the 400-K simulation contained a protein

molecule and 4,737 water molecules. The system was first equilibrated for 600 ps at 300 K, and the temperature was elevated up to 400 K. The molecular dynamics simulation was then carried out for 6 ns at 400 K.

Over-all structure

Figure 7 shows time-dependent changes of the radius of gyration (R_g) of the LA molecule during the simulation at 300 K and 400 K. The R_g value was calculated by the following equation:

$$R_g = \sqrt{\frac{\sum_i m_i r_i^2}{\sum_i m_i}} \tag{8}$$

where m_i denotes the mass of atom i and r_i denotes the position vector of atom i from the center of mass. The R_g value was calculated using the coordinates of the 1972 atoms in the LA molecule.

The R_g value slightly increased with time and reached 15.2 Å at the final step of the 400-K simulation, and this value was about 4 % larger than that observed in the 300-K equilibration (14.55 Å; mean value over 400 snapshots in the final 200 ps of the 300-K equilibration). The time dependent R_g value calculated using only residues 6-120 is also shown in Fig. 7 and the final value is close to that in the 300-K simulation. Therefore the increase of R_g observed in the 400-K simulation is mainly ascribable to the structural change near N-terminus (see below). Figure 8 shows time-dependent changes of the rmsd value of the structure during the 400-K simulation relative to the crystallographic structure of recombinant goat LA. The rmsd was calculated using the coordinates of the main-chain C^α, N and C of residues 0-120 in the protein

molecule. The rmsd value became 4.6 Å at 6 ns after elevating the temperature to 400 K.

Both the R_g and the rmsd values thus show that the protein molecule after the 400-K simulation is perturbed only to a small extent and may correspond to a structure at the beginning of the unfolding reaction. It seems that the rmsd value does not necessarily decrease when the R_g value decreases. This indicates that decrease of the R_g does not directly mean 'renaturation' of the molecule in the high-temperature simulations. During the period of 2.7-3.2 ns after the elevation of the temperature (time = 3.3-3.8 ns in Fig. 7 and 8), for example, the R_g value decreased and the backbone rmsd value increased. This suggests structural disruption without expansion of the size of the molecule occurred in the simulation. Thus there are rearrangements of inter-residue contacts in the 400-K trajectory.

Secondary structure in the α-domain

Three major α-helices (A-, B- and C- helices) and the C-terminal 3_{10} helix found in the α-domain of the crystal structure of goat LA (Pike *et al.*, 1996; Chaudhuri et al., 1999) were stable during the equilibration. After elevating the temperature, parts of these secondary structure elements became less stable. The N-terminal end of the C-helix became broken at 1 ns after elevating the temperature to 400 K. The backbone conformation of the N-terminal end of the A-helix changed and became the conformation of a 3_{10} helix. The C-terminal 3_{10} helix was occasionally broken during the 400-K simulation.

Conformational behavior around the Thr29 side chain.

24

To characterize the structural changes occurring around the side chain of Thr29 during the unfolding simulation, I examined the native contacts between Thr29 side chain and the surrounding residues in the conformations produced along the MD trajectories. Two residues were defined as to be contacted if any two non-hydrogen atoms of the residues are within 4.5 Å. The crystal structure of recombinant goat LA shows that the side chain of Thr29 has non-local contacts with the side chains of Ile101, Trp104, Leu105 and Cys111. These native inter-residue contacts of Thr29 were stable until 800 ps after elevating the temperature up to 400 K, but after 800 ps the native inter-residue contacts with Ile101 and Leu105 were lost (Fig. 9a). The Thr29-Cys111 contact was stable except for a period of from 2.6 to 4.0 ns after elevating the temperature (time = 3.2-4.6 ns in Fig. 9a). This contact is expected to be stable because of the presence of the disulfide bridge between Cys28 and Cys111. The inter-residue contact between Thr29 and Trp104 was also kept stable in the 400-K simulation.

Figure 10 compares a snapshot structure at 3.3 ns after elevating the temperature to 400 K with the crystal structure, showing the whole protein molecule (Fig. 10a) and local structure around Thr29 (Fig. 10b). It is clearly seen that in the high temperature structure, the side chains of Ile101 and Leu105, which fill the space near the Thr29 side chain in the Crystal structure, is apart from the residue Thr29, and that water molecules fill the empty space. The final rmsd value calculated from the coordinates of non-hydrogen atoms of the five residues was about 2 Å (Fig. 11a).

The disruption of the local structure around Thr29 in the 400-K simulation was accompanied by an increase of the solvent accessible

surface area of the Thr29 side chain from 1.1 $Å^2$ (mean value for the 400 snapshots in the final 200 ps of the equilibration at 300 K) to a value larger than 10 $Å^2$. The number of water molecules that lies within 3.3 Å of the side-chain oxygen atom of Thr29 was increased from 0-2 to 0-4 by the structural disruption (Fig. 9b). Interestingly, the extent of exposure of the Thr29 side chain to the solvent water did not continuously increases but had a maximum at 3.3 ns after elevating the temperature (time = 3.9 ns in Fig. 9b) along the 6-ns trajectory.

Changes in native packing in the aromatic clusters

I also examined the packing properties of two well-known aromatic clusters, the aromatic cluster around Trp104 (aromatic cluster II, ACII) and the aromatic cluster containing Trp118 (aromatic cluster I, ACI), of LA (Acharya *et al.*, 1989; Acharya *et al.*, 1991; Pike et al., 1996). The time dependence of the native contacts (Fig. 9a) shows that the native inter-residue contacts of ACII were remarkably more stable than those around Thr29 were, while the inter-residue contacts between Tyr103 and others sometimes got lost. The rmsd calculations using the coordinates of the atoms involved in the aromatic clusters (Fig. 11) also showed that the structure and packing properties of ACII were more stably native-like than those of the Thr29 region were during the 400-K simulation. The rmsd value of ACII remained 0.3 - 1.4 Å for 5 ns after elevating the temperature to 400 K with an exception of a period from 2.3-2.6 ns (time = 2.9-3.2 ns in Fig. 11). The native like characteristics of ACI located between the B-helix and the C-terminal 3_{10} helix was also preserved longer than the structure around the Thr29 side chain (Fig. 11).

N terminal region

The most drastic conformational change was observed in the N-terminal region in the 6.0-ns trajectory of the 400-K simulation, and it occurred at an early stage of the simulation. The difference of the backbone conformation in the N-terminal region between the simulated structure and the crystal structure was already obvious at 3.3 ns after elevating the temperature to 400 K (Fig. 10a). A number of the native non-local inter-residue contacts between the N-terminal (Met0 - Lys5) region and Gly35 - Ala40 region were lost.

As described above, the size of the molecule was remarkably native-like during the 400-K simulation. When a portion of the molecule had lost the native inter-residue contacts, generally it got new non-native inter-residue contacts. Contrary to this behavior of most region of the molecule, the N-terminal region did not acquire other non-native contacts, and became free from the side chain packing and hydrogen bonding to other parts of the molecule after the loss of the native contacts.

Other parts of the molecule

Residue contact maps are usually used to search for the native inter-residue contacts lost and new non-native contacts formed during the simulations. Figure 12 shows residue contact maps for the 300-K and 400-K simulations based on the final 400 snapshots of the simulation at respective temperature. Figure 9a shows time dependence of the inter-residue contacts between the A-helix and the B-helix. Most of the

native contacts between the A-helix and other helices in the α-domain began to be lost in the 400-K simulation. These native contacts, except for the contacts between Phe9 and neighbors, became lost in the first 3 ns of the simulation at 400 K, soon after the conformational change in the N-terminal region was observed.

Another short-term but remarkable conformational change was also observed near the boundary between the α- and β- domains in the first 1.0 ns after elevating the temperature. This conformational change was accompanied by the loss of the three inter-domain inter-residue contacts Trp26-Phe53, Thr33-Phe53 and Leu85-Lys13. The distance between the calcium binding loop and the loop region between the B-helix and the strand 1 was increased, and the distance between the C^α of Gln39 and the C^α of Leu81 (= 4.6 Å in the crystal structure) became ~ 10 Å. Increase of the side-chain accessible surface area for two buried residues (Leu52 and Phe53) near the inter-domain boundary was also observed. In spite of this remarkable conformational change, dissociation of the bound Ca^{2+} was not observed. Trp60 also locates near the inter-domain boundary, having non-local inter-residue contacts with amino acid residues in C-helix in the native crystal structure. As shown in Fig. 9a, some of these native contacts were also unstable at 400 K. As described above, ACII located near the inter-domain boundary was one of the most robust parts during the simulation. Therefore, robustness seems to be heterogeneous in the inter-domain boundary region.

600 K simulations of goat LA

For the purpose of examining the reproducibility of the behavior observed in the 400-K simulation, I further performed simulations at a

higher temperature (600 K). The starting structure of the goat LA molecule was the same as that in the 400-K simulation. The system was equilibrated for 300 ps at 298 K, and the temperature was increased up to 600 K. I carried out a 150-ps (run600-1), a 210-ps (run600-6) and four 250-ps (run600-2, 3, 4 and 5) simulations at 600 K.

Also in these 600-K simulations, the R_g values were not largely increased (Fig. 13). The larger R_g value in run600-3 is attributable to the conformational disruption of N-terminal region (see below). The backbone rmsd values reached 1.7 – 2.9 Å and 3.0 – 3.8 Å in 150 ps and 250 ps, respectively, after elevating the temperature to 600 K (Fig. 14).

The qualitative tendency of the stability of the major secondary structural elements in the α-domain of goat LA in the 600-K simulations was similar to that in the 400-K simulation. However, the C-terminal 3_{10} helix, in particular, seemed to be less stable in the 600 K simulations than that in the 400-K simulation.

Inter-residue contacts monitored along the simulation trajectories showed that the structural change of the protein molecule took place in different manners (Lazaridis & Karplus, 1997; Tsai et al., 1999) among these six simulations (Fig. 16), but some features of the structural change similar to those observed in the 400- K simulation were also observed in the 600-K simulations. The protein underwent large conformational fluctuation in the domain boundary region near the calcium binding site and the loop between the B-helix and a β-strand (strand 1) in these simulations, as observed in the 400-K simulation. The native inter-residue contacts between the N-terminus and the loop between the B-helix and strand 1 were disrupted in three (run600-3, run600-4 and run600-5) of the six trajectories as observed in the 400-K simulation (Fig. 15). Also in the other three, loss of the native non-local inter-residue

29

contacts near Glu1 or Gln2 was observed.

I also observed the structural change around the Thr29 in four of the six trajectories. In these simulations, loss of the inter-residue contacts of Thr29-Ile101 and Thr29–Leu105 were observed at 600 K (Fig. 15). In these 600-K simulations, the inter-residue contacts between Thr29 and Cys111 were not very stable as observed in the 400-K simulation. The number of water molecules that lies within 3.3 Å of the side-chain oxygen atom of Thr29 was, again, increased from 0-1 to 0-4 by the structural disruption (Fig. 15). A short-term water penetration into the Thr29 portion and a small disruption of the substructure around Thr29 were observed, but the structure around the Thr29 became again native-like, in one (run600-3) of the 600-K trajectories. This may be due to the fluctuation much larger at 600 K than at 400 K.

Native inter-residue contacts in ACII were stably preserved with the exception of contacts between Tyr103 and others, as observed in the 400-K simulation. On the other hand, the native structure of ACI was not stably preserved in the 600 K trajectories. In all of the six simulations at 600 K, the native inter-residue contact between His32 and Trp118 was lost and rmsd values calculated using heavy atoms in ACI remarkably increased after elevating temperature to 600 K (Fig. 17).

Discussion

I have studied the kinetic unfolding and refolding reactions of wild-type and mutant goat LA experimentally by stopped-flow CD measurements as well as the unfolding process performed in a computer by molecular dynamics simulations at high temperatures (400 and 600 K). The molecular dynamics simulations can describe structure and energetics of the protein molecule during the unfolding into atomic details, so that it may provide an effective approach that is complementary to experiments and can offer physical interpretations of the experimental results of protein folding. Here I compare the experimental results of folding with the structural properties of goat LA revealed by the simulations, and show that both are consistent with each other.

I have mainly focused on the folding and unfolding behavior of the α-domain, especially around the B-helix, of goat LA. The α-domain of the protein is known to be native-like in the MG state, and the B-helix is stable in that state. Thus, I have first studied experimentally the folding and unfolding behavior of the B-helix mutants (T29I, A30T, A30I, T33I, A30T + T33I). Particularly, the T29I mutation stabilized the native state by as much as 3.5 kcal/mol, and this mutational stabilization has been utilized for investigating the transition-state structure of the protein. Our previous studies have also shown that the presence of the extra N-terminal methionine residue in recombinant LA remarkably destabilize the protein by as much as 3.5 kcal/mol at 0 M denaturant (GdnHCl) (Chaudhuri et al., 1999). My present and our previous results of the kinetic refolding and unfolding measurements have shown that these stability changes are almost entirely due to the changes in the unfolding

rate constant, indicating that the native structures around Thr29 and the N-terminus has not yet been organized in the transition state of folding. These results are thus consistent with the observations in the molecular dynamics simulations in which the structural changes around Thr29 and the N-terminus occur at early stages of unfolding. In the following, I describe more details about the observations made by the experiments and the simulations and discuss the consistency between them and possible limitations of the present comparative approach of experiments and simulations.

Structure around Thr29 and the N-terminus of goat LA in the native state

The side chain of Thr29 has non-local inter-residue contacts with the side chains of Ile101, Trp104, Leu105 and Cys111 in the X-ray crystallographic structure of goat LA (Pike et al., 1996; Chaudhuri et al., 1999). These contacting residues shield the Thr29 residue completely from the solvent water. The Thr29 residue is thus located within the hydrophobic core of the α-domain, and the side chain hydroxyl group of Thr29 is hydrogen bonded to the main chain oxygen of Glu25.

The N-terminal methionine of recombinant goat LA is located on the surface of the molecule, but its movement is restricted within the rigid structure of the N-terminal region. The N-terminal methionine has non-local inter-residue contacts with the side chains of Asp37 and Gln39, and the N-terminal amino group is hydrogen-bonded with the side chain oxygen of Gln39. There is another hydrogen bond between the main-chain amide group of Glu1 and the carboxyl group of Asp37. The rigid structure of the extra N-terminal methionine and a conformational entropy effect in the unfolded state was proposed to be responsible to the

32

destabilization of the recombinant protein (Chaudhuri et al., 1999).

Structure around Thr29 and the N-terminus of goat LA in the transition state

The structure around the mutation site in the folding transition state can be investigated by the ϕ-value analysis (Serrano et al., 1992; Itzhaki et al., 1995). I thus calculated the apparent ϕ_U value for the T29I mutation. First, the logarithmic apparent unfolding rate constants, ln k_{app_WT} and ln k_{app_T29I}, for the wild-type and mutant proteins, respectively, were extrapolated linearly from the unfolded region of the GdnHCl concentration (higher than 3.4 M or 4.5 M for the wild type and T29I, respectively) to 2.6 M GdnHCl that is the transition midpoint of the wild-type protein. I used these two extrapolated k_{app} values as the approximation of the microscopic rate constants. Then, the ϕ_U is given by

$$\phi_U = RT \ln(k_{app\_WT} / k_{app\_T29I}) / \Delta\Delta G_U$$

where R is the gas constant, T is the absolute temperature, $\Delta\Delta G_U$ is the difference in the unfolding free energy change between the T29I mutant and the wild-type protein. When we use the ln k_{app_WT}, ln k_{app_T29I}, and $\Delta\Delta G_U$ values at 2.6 M, ϕ_U is estimated at 1.1, which is close to 1, indicating that the interaction and the structure around the side chain of Thr29 in the native state has not yet been organized in the transition state of folding.

As to the effect of the N-terminal methionine residue, the refolding and unfolding rate constants were investigated only at limited concentrations of GdnHCl, 0.5 M and 5.5 M GdnHCl for refolding and

unfolding, respectively (Chaudhuri et al., 1999), so that we could not evaluate the ϕ_U quantitatively. However, only the unfolding rate constant is affected by the presence of the extra methionine residue, and the structure around the methionine residue has not yet been native-like in the transition state of folding. The rigid structure of the N-terminal methionine is thus not organized in the transition state.

Molecular dynamics simulations at high temperature

The present molecular dynamics simulation results have shown that the structure around Thr29 is a brittle part in the α-domain. The native non-local contacts T29-Ile101 and T29-L105 were less stable than most of the contacts around Trp26 (B-helix / C-helix interface) and contacts involved in ACI (B-helix / 3_{10} helix interface), in the 400-K simulation. The similar tendency was observed in the 600-K simulations, while Trp118-His32 contact involved in ACI was less stable than those around Thr29 were in some 600-K trajectories. The loss of the two native non-local inter-residue contacts of Thr29 with Ile101 and Leu105 took place earlier (the main-chain rmsd ~ 2.0 Å) than the loss of the native packing of ACII (aromatic cluster locating at the domain boundary) that was found to be a robust part in the molecule, in the 400-K simulation. Similar trends were also observed in the 600-K simulations, in which not only the two inter-residue contacts, Thr29-Ile101 and Thr29-Leu105, but also the contact between Thr29 and Cys111 was lost. The observation of the similar trends in the 400-K and 600-K simulations, where not only the temperature but also other simulation conditions were different from each other (see materials and methods), indicates the universality of the present observations.

34

The present simulation results have also shown that the structural change around Thr29 is accompanied by penetration of water molecules into the space near the Thr29 side-chain oxygen at a stage of the main-chain rmsd of 2.0-2.5 Å in both the 400-K and 600-K simulations. A short-lived penetration of water molecules near Thr29 side chain was also observed at earlier stage (main chain rmsd < 1.5 Å) in one of the 600-K molecular dynamics trajectories .

Because the difference in not only the intra-molecular interaction in the native state but also the protein-water interaction in the unfolded state between the wild type and the mutant should contribute to the stability change caused by the mutation, the structural change observed in the high-temperature trajectories is qualitatively consistent to the experimental results presented here. Thus the loss of the inter-residue contacts Thr29-Ile101 and Thr29-Leu105 accompanied by the penetration of the water molecules into the neighbor is a probable conformational behavior of the Thr29 part of the molecule at the initial stage of the unfolding of goat LA.

The molecular dynamics simulation results have shown that the structure of the N-terminal methionine is disrupted early in unfolding. Similar local unfolding in the N-terminal region was also observed in three of the six 600-K simulations.

The above observations in the molecular dynamics simulations, the structural change around Thr29 accompanied by penetration of water molecules, and the structural disruption of the N-terminal region are thus consistent with the results of the experiments by the stopped-flow CD method. The consistency between the experimental and simulation results indicate that the molecular dynamics simulation at high temperature is a useful approach to provide structural

interpretations of the experimental results of protein folding, and the observation of the behavior of other parts of LA in the high-temperature simulation is also likely to provide meaningful information about the initial stage of the unfolding. For example, further analysis of the unfolding simulation trajectories would teach us what is the trigger of the unfolding of LA. And , longer simulations would provide a model of the transition state of the folding of LA.

Possible limitations

Molecular dynamics simulations at high temperatures give an insight into the structural behavior that occurs at the initial stage of protein unfolding. I have assumed that the unfolding path is the reversal of the folding path for the purpose of applying this approach to a study of protein folding, and the initial stage of unfolding observed by the simulation may correspond to the final stage of folding. Although the validity of this assumption has not yet been established, previous molecular dynamics simulation studies of several different globular proteins strongly suggest that this approach is effective. The consistency between my unfolding simulations and the folding experiments described above also supports the validity of the above assumption.

Furthermore, I have also assumed that the unfolding behavior at high temperature used in the simulations is essentially the same as the behavior observed in the GdnHCl induced unfolding experiments. Although the validity of this assumption has also not been established there is evidence that the assumption may well hold. The m value, which gives the denaturant concentration dependence of the free energy

change of the denaturant induced unfolding (between the native state and the denaturant induced unfolded state), is known to be well correlated with the heat capacity change (ΔC_p), which gives the temperature dependence of the enthalpy change of the thermal unfolding (between the native and the thermal unfolded state), and both the parameters are proportional to the change in the accessible surface area of the protein molecule on unfolding (Myers et al., 1995). Similarly, the stability of a protein against the thermal unfolding is well correlated with the stability of the protein against the denaturant-induced unfolding. These suggest that both the unfolding transitions may occur by essentially the same molecular mechanism and that the effect of hydration is most important for both the unfolding transitions.

It should also be noted that the unfolding behavior observed in the simulations is the behavior of a single protein molecule, which may be different from the ensemble-averaged behavior of a large number of protein molecules (~ nmol/ml in my CD study) observed in the experiments. Here I carried out the six unfolding simulations from different initial conformations at 600 K and observed structural behavior of the LA molecule at the initial stage of the unfolding. While the number of the trajectories is still much smaller than the number of the molecules in the real solution in experiments, I found some characteristics of the unfolding behavior with moderate reproducibility, including the structural change around Thr29 and the local unfolding in the N-terminal region described above.

It should be noted that the conformational properties described here can not necessarily refer to those of the transition state, while the data of the kinetic measurements of refolding and unfolding of mutant proteins give the information about the energetics in the transition state.

In addition, because the molecule reached neither the MG state nor the fully unfolded state, it is not clear whether what I observed in the simulations are the real unfolding or a dead-end structural fluctuation. I believe, however, that what I observed is an unfolding behavior. Judging from the time-dependent main-chain rmsd and the time-dependent non-local inter-residue contacts, "refolding" from a dead-end partially unfolded state did not occur in the simulation time in both of the 400-K and 600-K trajectories.

Other Molecular Dynamics Studies of LA

Murphy et al. reported the changes in inter-residue contacts and hydrogen bondings in the guinea pig LA molecule without Ca^{2+} at a low pH studied by the molecular dynamics simulations (Murphy et al., 1998). According to the contact map that they reported from their simulations, the non-local interactions between the N terminus and the region of residues 35-39 were at least partially maintained. The difference between their and my results may be due to the differences in the conditions of the simulations, including temperature, pH, the length of the simulation and the presence or absence of bound Ca^{2+}. They also mentioned a structural change in the domain boundary. My observation of the fluctuation near the boundary may be similar to their results.

Smith et al. reported a conformational change in the C-terminal region accompanied by the disruption of the C-terminal 3_{10} helix and the D-helix studied by the molecular dynamics simulations of human LA at acidic pH (Smith et al., 1999). This conformational change is consistent with the increase of the rmsd of AC118 observed in the present molecular

dynamics simulations at high temperature. However, the decrease in the radius of gyration to a value smaller than 14 Å reported by them was not observed in the present simulations. This may be due to the difference in the protonation of the charged groups (pH) or in the temperatures.

Conclusion

The experimental and simulation study of folding/unfolding of goat LA has been described. The results of kinetic refolding and unfolding by the stopped-flow method showed that the hydrophobic parts around Thr29 side chain acquire the native interactions at the later stage of the folding of LA. Molecular dynamics simulations at high temperatures successfully provided detailed structural interpretations of the behavior at the initial stage of the unfolding of goat LA. This suggests that the other observation of unfolding behavior in the simulation trajectories would provide meaningful information about folding reaction of LA.

I also consider that the study of folding reaction of LA should provide insight into the folding of multi-domain globular proteins with disulfide bridges. So, experimentally examining the behavior of the inter-domain region observed in the simulation trajectories should be a next subject to investigate.

Summary

Folding reaction of goat α-lactalbumin has been studied by stopped-flow circular dichroism method and molecular dynamics simulation.

Effect of four single mutations and a double mutation on the stability of the protein in a native condition and on the kinetic refolding/unfolding rate was studied experimentally. The mutations were introduced into residues located at the hydrophobic core in the α-domain of the molecule. Here I show that an amino acid substitution (T29I) increases the stability of the native state of goat α-lactalbumin against the guanidine hydrochloride-induced unfolding at neutral pH by 3.5 kcal/mol. Kinetic refolding and unfolding of the wild-type and the mutant goat α-lactalbumin measured by stopped-flow circular dichroism method have shown that the stabilization is due to the decrease of the unfolding rate. This indicates that the local interactions which are perturbed by the T29I mutation are not native-like at all in the transition state of the folding reaction of goat α-lactalbumin. Thus the local structure around the Thr29 side chain should be constructed at a last stage of the folding reaction.

To investigate the local structural change around the Thr29 side chain in atomic resolution, high-temperature (at 400 K and 600 K) molecular dynamics simulations were performed. Although the experimental condition and the simulation condition were different, I initially assumed such simulations provide a representative subset of the unfolding pathways of the protein molecule. To prevent the influence of the truncation of the long-range electrostatic interaction, the 400-K and the 600-K simulations were carried out without the truncation of non-

bond interaction and with twin-range cutoff method, respectively.

The protein molecule did not largely unfold during the simulation time. The protein molecule also did not reach the molten globule state. Thus the simulation results only provide the information of the initial stage about the unfolding path. The Thr29 part of the molecule, however, experienced a structural disruption with the loss of inter-residue contacts and the water molecule penetration in the 400-K simulation and four of the six 600-K simulations. Another structural disruption that is consistent with the previous experimental result was also observed in the N-terminal region.

Results of the observations of some other conformational changes are also described, and the applicability of such methodology is discussed.

Acknowledgements

References

Alexandrescu, A. T., Evans, P. A., Pitkeathly, M., Baum, J. & Dobson, C. M. (1993). Structure and dynamics of the acid-denatured molten globule state of α-lactalbumin: a two-dimensional NMR study. Biochemistry 32(7), 1707-18.

Arai, M. & Kuwajima, K. (1996). Rapid formation of a molten globule intermediate in refolding ofα- lactalbumin. Fold Des 1(4), 275-87.

Arai, M. & Kuwajima, K. (2000). Role of the molten globule state in protein folding. Adv. Protein chem. 53, in press

Baum, J., Dobson, C. M., Evans, P. A. & Hanley, C. (1989). Characterization of a partly folded protein by NMR methods: studies on the molten globule state of guinea pigα-lactalbumin. Biochemistry 28(1), 7-13.

Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A. & Haak, J. R. (1984). Molecular dynamics with coupling to an external bath. J. Chem. Phys. 81(8), 3684-3690.

Caflisch, A. & Karplus, M. (1999). Structural details of urea binding to barnase: a molecular dynamics analysis. Structure Fold Des 7(5), 477-88.

Case, D. A., Pearlman, D. A., Caldwell, J. W., Cheatham III, T. E., Ross, W. S., Simmerling, C. L., Darden, T. A., Merz, K. M., Stanton, R. V., Cheng, A. L., Vincent, J. J., Crowley, M., Ferguson, D. M., Radmer, R. J., Seibel, G. L., Singh, U. C., K., W. P. & Kollman, P. A. (1997). AMBER 5. University of California, San Francisco.

Chaudhuri, T. K., Horii, K., Yoda, T., Arai, M., Nagata, S., Terada, T. P., Uchiyama, H., Ikura, T., Tsumoto, K., Kataoka, H., Matsushima, M., Kuwajima, K. & Kumagai, I. (1999). Effect of the extra N-terminal methionine residue on the stability and folding of recombinant α-lactalbumin expressed in Escherichia coli. J Mol Biol 285(3), 1179-94.

Chyan, C. L., Wormald, C., Dobson, C. M., Evans, P. A. & Baum, J. (1993). Structure and stability of the molten globule state of guinea-pig α- lactalbumin: a hydrogen exchange study. Biochemistry 32(21), 5681-91.

Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M., Ferguson, D. M., Spellmeyer, D. C., Fox, T., Caldwell, J. W. & Kollman, P. A. (1995). A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. J. Am. Chem. Soc. 117, 5179-5197.

Glasstone, S., Laidler, K. J., Eyring, H., (1941). The Theory of RATE PORCESSES, The Kientics of Chemical Reactions, Viscosity, Diffusion and Electrochemical Phenomena. McGRAW-HILL., New York

Forge, V., Wijesinha, R. T., Balbach, J., Brew, K., Robinson, C. V., Redfield, C. & Dobson, C. M. (1999). Rapid collapse and slow structural reorganisation during the refolding of bovine α-lactalbumin. J Mol Biol 288(4), 673-88.

Fulton, K. F., Main, E. R., Daggett, V. & Jackson, S. E. (1999). Mapping the interactions present in the transition state for unfolding/folding of FKBP12. J Mol Biol 291(2), 445-61.

Hill, R. L. & Brew, K. (1975). Lactose synthetase. Adv. Enzymol. Relat. Areas Mol. Biol. 43, 411-490.

Hiraoka, Y., Segawa, T., Kuwajima, K., Sugai, S. & Murai, N. (1980). α-Lactalbumin: a calcium metalloprotein. Biochem. Biophys. Res. Commun. 95(3), 1098-104.

Hubbard, S. J. & Thornton, J. M. (1993). 'NACCESS', Computer program. Department of biochemistry and molecular biology, University College, London.

Ikeguchi, M., Kuwajima, K., Mitani, M. & Sugai, S. (1986a). Evidence for identity between the equilibrium unfolding intermediate and a transient folding intermediate: a comparative study of the folding reactions of α-lactalbumin and lysozyme. Biochemistry

25(22), 6965-72.

Ikeguchi, M., Kuwajima, K. & Sugai, S. (1986b). Ca2+-induced alteration in the unfolding behavior of α-lactalbumin. J Biochem (Tokyo) 99(4), 1191-201.

Itzhaki, L. S., Otzen, D. E. & Fersht, A. R. (1995). The structure of the transition state for folding of chymotrypsin inhibitor 2 analysed by protein engineering methods: evidence for a nucleation-condensation mechanism for protein folding. J Mol Biol 254(2), 260-88.

Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. (1983). Comparison of simple potential functions for simulating liquid water. J. Chem. Phys. 79(2), 926-935.

Kabsch, W. & Sander, C. (1983). Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers 22(12), 2577-2637.

Kataoka, M., Kuwajima, K., Tokunaga, F. & Goto, Y. (1997). Structural characterization of the molten globule of α-lactalbumin by solution X-ray scattering. Protein Sci 6(2), 422-30.

Kazmirski, S. L. & Daggett, V. (1998). Simulations of the structural and dynamical properties of denatured proteins: the "molten coil" state of bovine pancreatic trypsin inhibitor. J Mol Biol 277(2), 487-506.

Kuwajima, K. (1996). The molten globule state of α-lactalbumin. Faseb J 10(1), 102-9.

Kuwajima, K., Nitta, K. & Sugai, S. (1980). Intramolecular perturbation of tryptophans induced by the protonation of ionizable groups in goat α-lactalbumin. Biochim Biophys Acta 623(2), 389-401.

Kuwajima, K., Nitta, K., Yoneyama, M. & Sugai, S. (1976). Three-state denaturation of α-lactalbumin by guanidine hydrochloride. J Mol Biol 106(2), 359-73.

Kuwajima, K., Yamaya, H., Miwa, S., Sugai, S. & Nagamura, T. (1987). Rapid formation of secondary structure framework in protein

folding studied by stopped-flow circular dichroism. FEBS Lett. 221, 115-118.

Lazaridis, T. & Karplus, M. (1997). "New view" of protein folding reconciled with the old through multiple unfolding simulations. Science 278(5345), 1928-31.

Li, A. & Daggett, V. (1998). Molecular dynamics simulation of the unfolding of barnase: characterization of the major intermediate. J Mol Biol 275(4), 677-94.

Lindahl, L. & Vogel, H. J. (1984). Metal-ion-dependent hydrophobic-interaction chromatography of α- lactalbumins. Anal Biochem 140(2), 394-402.

McKenzie, H. A. & White, F. H., Jr. (1991). Lysozyme and α-lactalbumin: Structure, function, and interrelationships. Adv. Protein Chem. 41, 173-315.

Myers, J. K., Pace, C. N., and Scholtz, J. M. (1995) Denaturant *m* values and heat capacity changes: Relation o changes in accessible surface areas of protein unfolding. Protein Sci. 4, 2138-2148

Murphy, L. R., Li, N., Baum, J. & Levy, R. M. (1998). Tertiary contacts in α-lactalbumin at pH 7 and pH 2: a molecular dynamics study. J Biomol. Struct. Dyn. 16(2), 355-65.

Pace, C. N. (1986). Determination and analysis of urea and guanidine hydrochloride denaturation curves. Methods Enzymol. 131, 266-80.

Pike, A. C. W., Brew, K. & Acharya, K. R. (1996). Crystal structures of guinea-pig, goat and bovine α-lactalbumin highlight the enhanced conformational flexibility of regions that are significant for its action in lactose synthase. Structure 4(6), 691-703.

Ryckaert, J. P., Ciccotti, G. & Berendsen, H. J. C. (1977). Numerical integration of the cartesian equations of motion of a system with constraints: Molecular dynamics of n-Alkanes. J. Comput. Phys. 23, 327-341.

Saito, M. (1992). Molecular dynamics simulations of proteins in water without the truncation of long-range coulomb interactions. Mol. Simul. 8, 321-333.

Saito, M. (1994). Molecular dynamics simulations of proteins in solution: Artifacts caused by the cutoff approximation. J. Chem. Phys. 101(5), 4055-4061.

Saito, M. (1999). Molecular dynamics model structures for the molten globule state of α-lactalbumin: aromatic residue clusters I and II. Protein Eng. 12(12).

Schulman, B. A., Redfield, C., Peng, Z. Y., Dobson, C. M. & Kim, P. S. (1995). Different subdomains are most protected from hydrogen exchange in the molten globule and native states of human α-lactalbumin. J Mol Biol 253(5), 651-7.

Serrano, L., Matouschek, A. & Fersht, A. R. (1992). The folding of an enzyme. 3. Structure of the transition state for unfolding of barnase analysed by a protein engineering procedure. J. Mol. Biol. 22, 805-818.

Sharp, K. A., Nicholls, A., Friedman, R. & Honig, B. (1991). Extracting hydrophobic free energies from experimental data: Relationship to protein folding and theoretical models. Biochemistry 30, 9686-9697.

Smith, L. J., Dobson, C. M. & van Gunsteren, W. F. (1999). Molecular dynamics simulations of human α-lactalbumin: changes to the structural and dynamical properties of the protein at low pH. Proteins 36(1), 77-86.

Song, J., Bai, P., Luo, L. & Peng, Z. Y. (1998). Contribution of individual residues to formation of the native-like tertiary topology in the α-lactalbumin molten globule. J Mol Biol 280(1), 167-74.

Tirado-Rives, J., Orozco, M. & Jorgensen, W. L. (1997). Molecular dynamics simulations of the unfolding of barnase in water and 8 M aqueous urea. Biochemistry 36(24), 7313-29.

Tsai, J., Levitt, M. & Baker, D. (1999). Hierarchy of structure loss in MD

simulations of src SH3 domain unfolding. J Mol Biol 291(1), 215-25.

Uchiyama, H., Perez-Prat, E. M., Watanabe, K., Kumagai, I. & Kuwajima, K. (1995). Effects of amino acid substitutions in the hydrophobic core of α- lactalbumin on the stability of the molten globule state. Protein Eng 8(11), 1153-61.

Van Gunsteren, W. F. & Berendsen, H. J. C. (1990). Computer simulation of molecular dynamics: Methodology, applications and perspectives. Angew. Chem. Int. Ed. Engl. 29, 992-1023.

Wu, L. C. & Kim, P. S. (1998). A specific hydrophobic core in the α-lactalbumin molten globule. J Mol Biol 280(1), 175-82.

Yutani, K., Ogasahara, K. & Kuwajima, K. (1992). Absence of the thermal transition in apo-α-lactalbumin in the molten globule state. A study by differential scanning microcalorimetry. J Mol Biol 228(2), 347-50.

Table 1

The native-state stability of the wild type and the mutants of goat LA evaluated from the results of equilibrium unfolding experiments.

	C_M		m	$\Delta G_{U\ H2O}$	$\Delta \Delta G_U$	$\Delta \Delta G_{tr}$
WT	2.60	±0.02	2.2±0.1	5.7	0	
T29I	3.93	±0.01	2.6±0.1	10.1	3.5	3.34
A30T	2.44	±0.06	3.1±0.7	7.4	−0.5	0.50
A30I	3.01	±0.05	2.7±0.5	8.1	1.1	3.84
T33I	2.31	±0.05	2.0±0.3	4.6	−0.6	3.34
A30I + T33I	2.85	±0.02	2.1±0.1	6.1	0.5	7.18

C_M is the transition midpoint concentration of GdnHCl, m is a measure of the cooperativity of the transition, $\Delta G_{U\ H2O}$ is the unfolding free energy change at 0 M GdnHCl. $\Delta \Delta G_U$ is the change in the unfolding free energy caused by the mutation calculated at 2.6 M GdnHCl. $\Delta \Delta G_{tr}$ is the value from Sharp *et al.* (1991)

Table 2

The kinetic refolding parameters of goat LA and mutants ([GdnHCl] = 0.5 M, 25 °C, pH = 7.0)

	wild type		T29I		A30I + T33I	
i	A_i (deg cm²/dmol)	k_i (s⁻¹)	A_i (deg cm²/dmol)	k_j (s⁻¹)	A_i (deg cm²/dmol)	k_j (s⁻¹)
1	2250±115 (0.82)	5.7±0.3	2660±51 (0.92)	8.8±0.3	329±67 (0.49)	12±5
2	396 ±120 (0.14)	1.3±0.4	163 ±57 (0.06)	1.6±0.6	277±81 (0.41)	2.6±0.6
3	88 ±13 (0.03)	0.06±0.02	70 ± 9 (0.02)	0.07±0.02	63 ± 8 (0.09)	0.08±0.02

Each A_i represents the observed ellipticity change for the ith phase. Each k_i represents the observed apparent rate constant of the ith phase. Fractional amplitude of each phase is shown in the parenthesis.

Legends to figures

Fig. 1

X-ray structure of recombinant goat LA. The side chains of Thr29, Ala30 and Thr33 are shown as sticks. B-helix (residues 23 - 34) is colored orange. The other parts of alpha domain and beta domain are colored cyan and magenta respectively.

Fig. 2

The block diagram of the stopped-flow CD apparatus attached to the Jasco J720 spectropolarimeter that is used in this study. Principle of the stopped-flow apparatus is described in text.

Fig. 3

Explanation of the ϕ-value analysis in the case of a two-state transition. U represents the unfolded state. N represents the native state. ‡ represents the transition state of the folding. $\Delta\Delta G_U$ is the change in the unfolding free energy caused by the amino acid substitution. $\Delta G_{U\ddagger}$ and $\Delta G_{N\ddagger}$ are refolding and unfolding activation free energy, respectively.

Fig. 4

a - f: GdnHCl induced equilibrium unfolding transition curves monitored by CD (222 (solid line) and 270 nm (dotted line)), showing fractional existence of unfolded state as a function of GdnHCl concentration for wild type (a), T29I (b), A30T (c), A30I (d), T33I (e) and A30I + T33I (f). g: GdnHCl induced unfolding transition curves of wild type (thick solid line) and mutants (T29I: thin solid line, A30T: thick dotted line, A30I: thin dotted line, T33I: thin dashed line, A30I + T33I: thick dashed line) of goat LA monitored by CD (270 nm).

Fig. 5

Time courses of kinetic refolding (a) and unfolding (b) reactions of wild type (thick solid line) and mutants (T29I: dashed line, A30I + T33I: thin solid line) of goat LA monitored by CD (225 nm).

Fig. 6

Apparent kinetic constants of refolding and unfolding of wild type (filled triangle) and mutants (T29I: cross, A30I + T33I: open triangle) of goat LA (a). The denaturant (GdnHCl) concentration jumps from 1.0 M (for unfolding) or 5.5 M (for refolding) were done by stopped-flow apparatus. The kinetic refolding and unfolding were monitored by CD (225 nm) at neutral pH and 25 ℃. The apparent refolding and unfolding rate constants of wild type and T29I are compared in (b) and (c), respectively.

Fig. 7

The radius of gyration (R_g) of the goat LA molecule during the 300 K and 400 K simulations. It was calculated using the coordinates of the all atoms in the LA molecule (solid line) or atoms in residues 6-123 (dotted line). The temperature was elevated at time = 600 ps.

Fig. 8

The main-chain rmsd for the 300-K and 400-K simulations. The rmsd was calculated using coordinates of main-chain N, C and C^α of residues 0-120 (solid line) or 6-120 (dotted line). The temperature of the system was elevated at time = 600 ps. Inset shows the value in equilibration at 300 K. Positional constraint of heavy atoms to the X-ray structure was applied during the first 10 ps (see materials and methods).

Fig. 9

a: Time dependent change of the non-local inter-residue native contacts near N-terminus and in hydrophobic portions of goat LA during the 300-K and 400-K simulations, showing non-local contacts of N-terminus (Met0-Asp37, Met0-Gln39, Met0-Ala40, Glu1-Tyr37, Glu1-Asp37, Glu1-Thr38, Glu1-Gln39, Gln2-Gly35, Gln2-Tyr36, Gln2-Asp37, Leu3-Tyr36 and Leu3-Thr38) A, non-local contacts of Thr29 (Thr29-Ile101, Thr29-Leu105, Thr29-Cys111 and Thr29-Trp104) B, contacts between the A-and B-helices (Lys5-Val27, Val8-Val27, Val8-Ala30, Phe9-Leu23, Phe9-Pro24 and Phe9-Val27) C, non-local contacts in aromatic cluster I (Phe31-Trp118 and His32-Trp118) D, non-local contacts in aromatic cluster II (Phe53-Tyr103, Ile55-Tyr103, Ile95-Tyr103, Ile55-Ile95, Phe53-Trp104 and Ile95-Trp104) E, non-local contacts of Trp26 (Trp26-Leu12, Trp26-Leu15, Trp26-Tyr18, Trp26-Val21, Trp26-Leu96, Trp26-Ile101, Trp26-Trp104, Trp26-Phe53) F and non-local contacts of Trp60 (Trp60-Cys91, Trp60-Lys94, Trp60-Ile95, Trp60-Lys98, Trp60-Val99 and Trp60-Tyr103) G. The figure shows whether the native contact was preserved in more than 50 % of conformations within every ten conformations (for 5 ps, filled), or not (not filled). b: The number of water molecule that is distant from the side-chain oxygen of Thr29 by less than 3.3 Å in 300-K and 400-K simulations.

Fig. 10

The X-ray structure of goat LA and a snapshot structure at 3.3 ns after elevating the temperature to 400 K. Ribbon structures of the whole LA molecule (a-c) and detail structure around Thr29 (d-e). The backbone of the molecule is colored gray and yellow for the X-ray structure (a-left, b, and d) and the 3.3-ns snapshot structure (a-right, c, and e), respectively. d-e: Three residues, Ile101, Leu105 and Trp104, contacting with Thr29 are shown colored orange, green and violet respectively. The three water oxygen atoms near the side chain oxygen

of Thr29 are also shown in (e), colored blue.

Fig. 11

The rmsd values calculated for substructures in the α-domain for the equilibration at 300-K and 400-K simulations. Around Thr29 (Thr29, Ile101, Trp104, Leu105, Cys111) (a), aromatic cluster I (Phe31, His32, Gln117, Trp118) (b), aromatic cluster II (Phe53, Ile55, Ile95, Tyr103, Trp104) (c), around Trp26 (Trp26, Leu15, Val21, Leu96, Trp104) (d), the interface between the A- and B-helices (Val8, Phe9, Leu23, Val27) (e). The rmsd was calculated using coordinates of all non-hydrogen atoms of the corresponding residues.

Fig. 12

Inter-residue contact maps for the equilibration at 300 K (a) and 400-K simulation (b). a: An inter-residue contact was defined as native (■) if it was maintained in more than 50 % of 400 conformations for the final 200 ps of 300-K simulation. b: An inter-residue contact was defined as present in 400 K if it was maintained in more than 50 % of 400 conformations for the final 200 ps of the 400-K simulation. ■ represents a native inter-residue contact and ◆ represents a non-native contacts that is absent in (a). □ represents a lost native contact.

Fig. 13

The radius of gyration (R_g) of the goat LA molecule during the 600-K simulations. (blue: run600-1, black: run600-2, green: run600-3, red: run600-4, sky blue: run600-5, orange: run600-6). Inset shows the value in equilibration at 298 K.

Fig. 14

The main-chain rmsd for the 600-K simulations. The same colors are

used as Fig.11. The rmsd was calculated using coordinates of main-chain N, C and C$^\alpha$ of residues 0-120. Inset shows the value in equilibration at 298 K.

Fig. 15

a - f: Time dependent change of the non-local inter-residue contacts near N-terminus and in hydrophobic portions of goat LA in the 600-K simulations (a: run600-1, b: run600-2, c: run600-3, d: run600-4, e: run600-5, f: run600-6). The same non-local inter-residue contacts listed in the legend of Fig.7a are shown. g - l: The number of water molecule that is distant from the side-chain oxygen of Thr29 by less than 3.3 Å in the 600-K simulations (g: run600-1, h: run600-2, i: run600-3, j: run600-4, k: run600-5, l: run600-6).

Fig. 16

Time dependent change of the all native non-local inter-residue contacts in the goat LA molecule during the 600-K simulations, showing the contacts within the α-domain (α), the contacts within the β-domain (β), and the contacts between the α- and the β-domains (α/β). Detailed examination of the changes in the all of these inter-residue contacts occurred in the simulations is not the purpose of my study. I show this figure just as the 'fingerprints' of the trajectories.

Fig. 17

The time dependent rmsd value calculated using only the residues that are in the ACI. Coordinates of the all non-hydrogen atoms in the corresponding residues were used. The same colors are used as Fig.11.

Figures

Fig. 1

Fig. 2

Fig. 3

Fig. 4 a

Fig. 4 b

Fig. 4 c

F_{app}

[GdnHCl] (M)

Fig. 4 d

Fig. 4 e

Fig. 4 f

Fig. 4 g

Fig. 5 a

Fig. 5 b
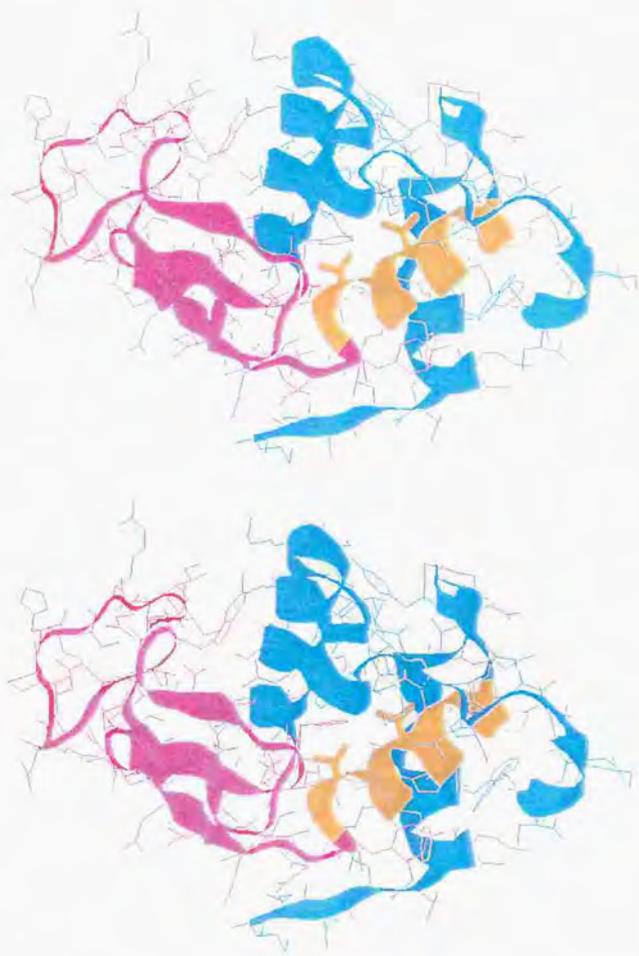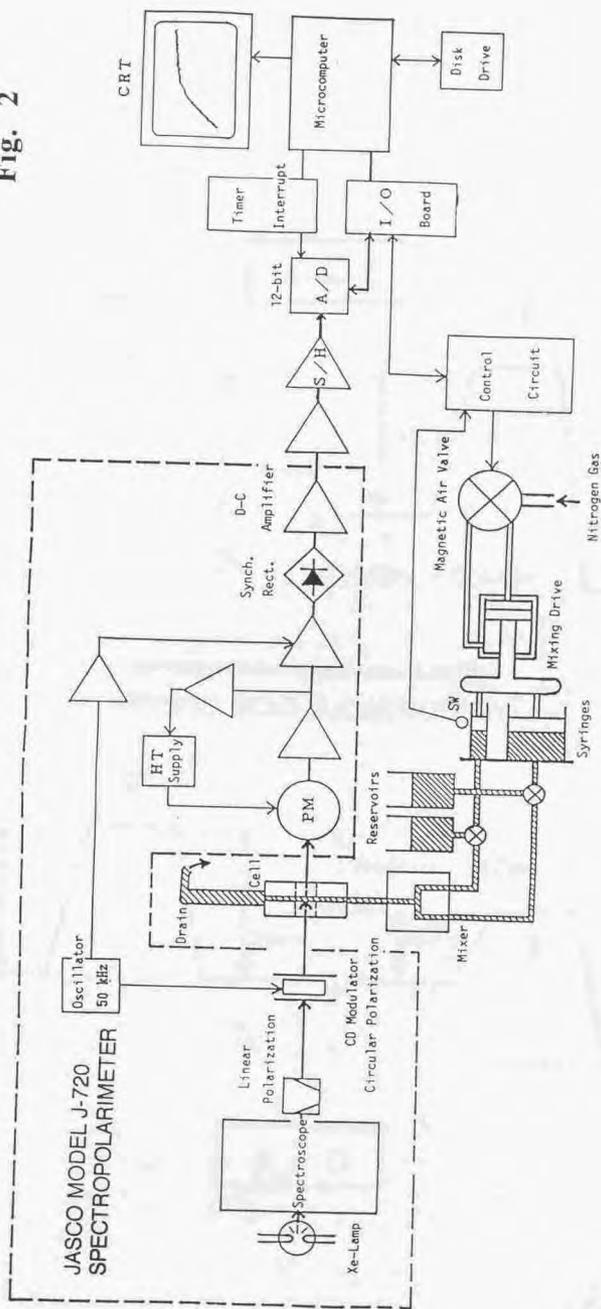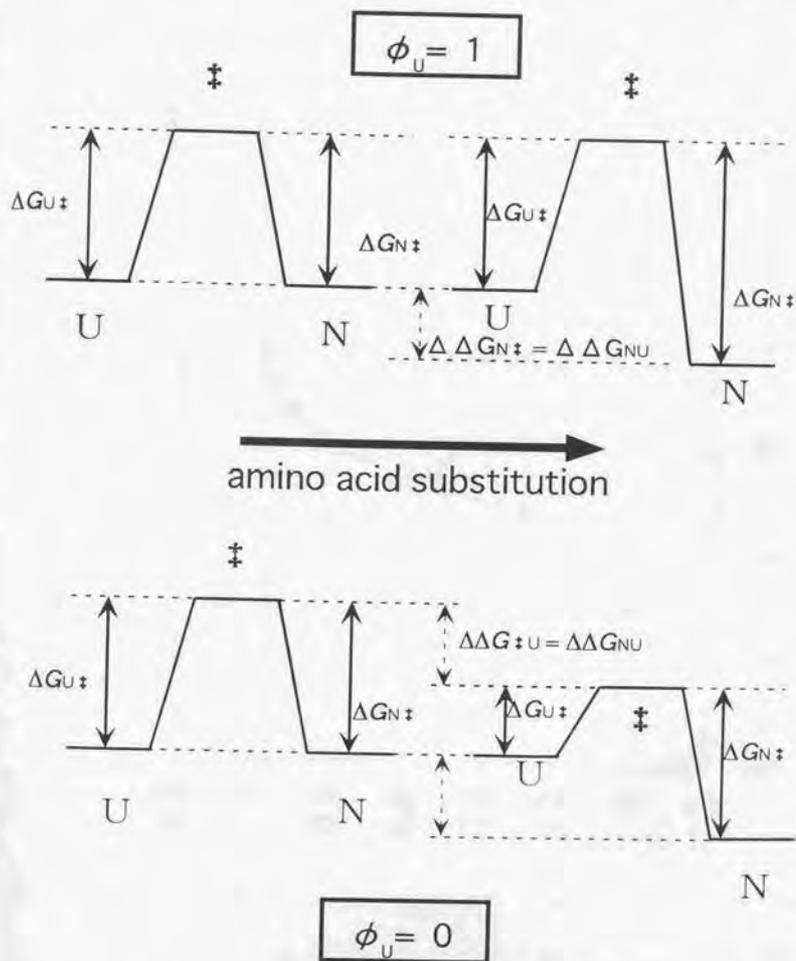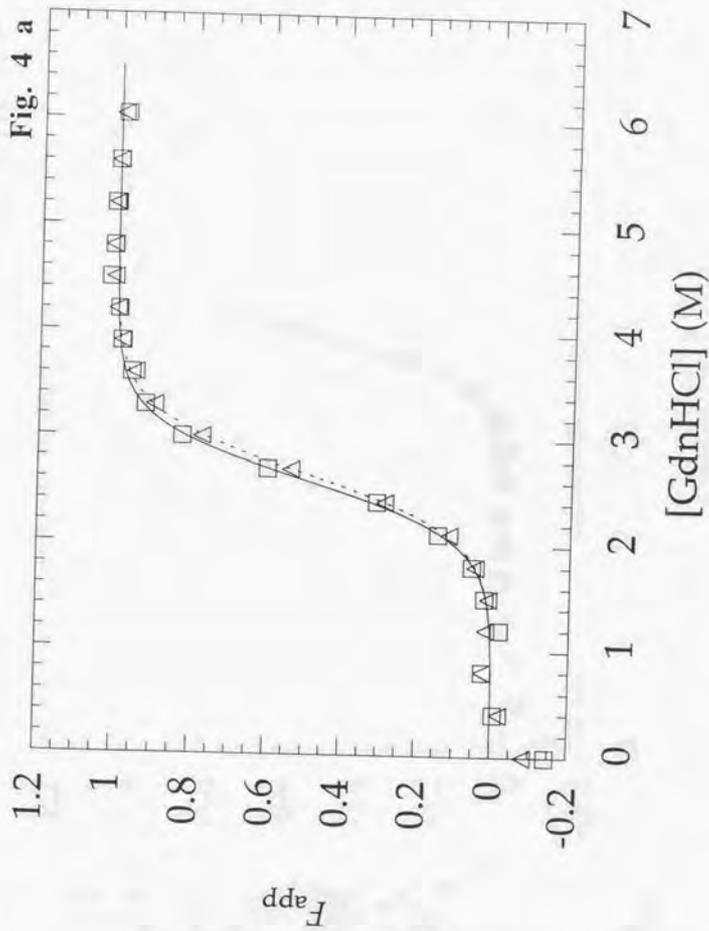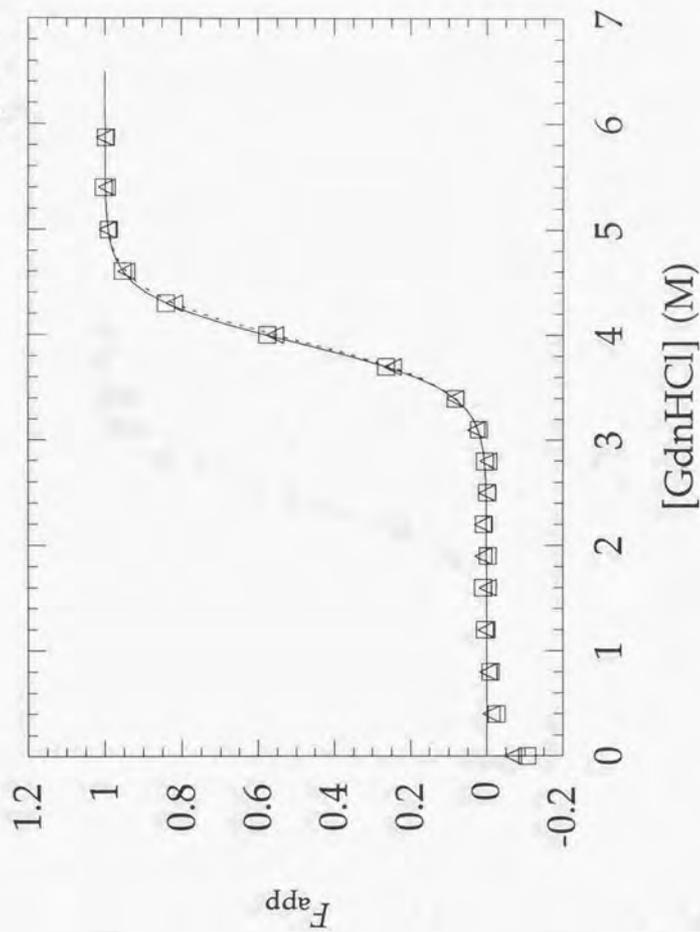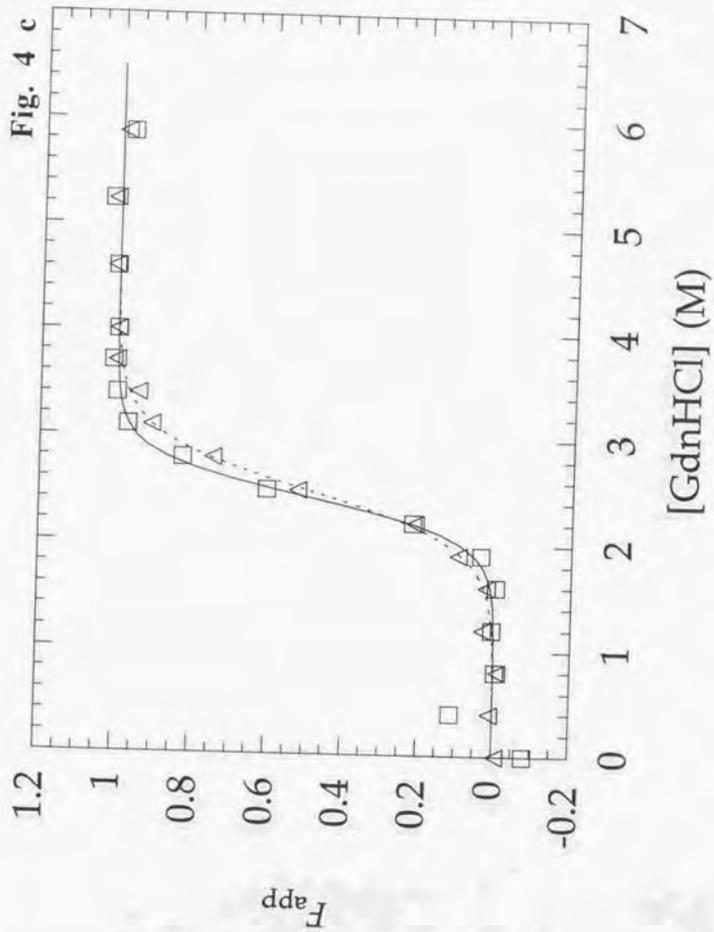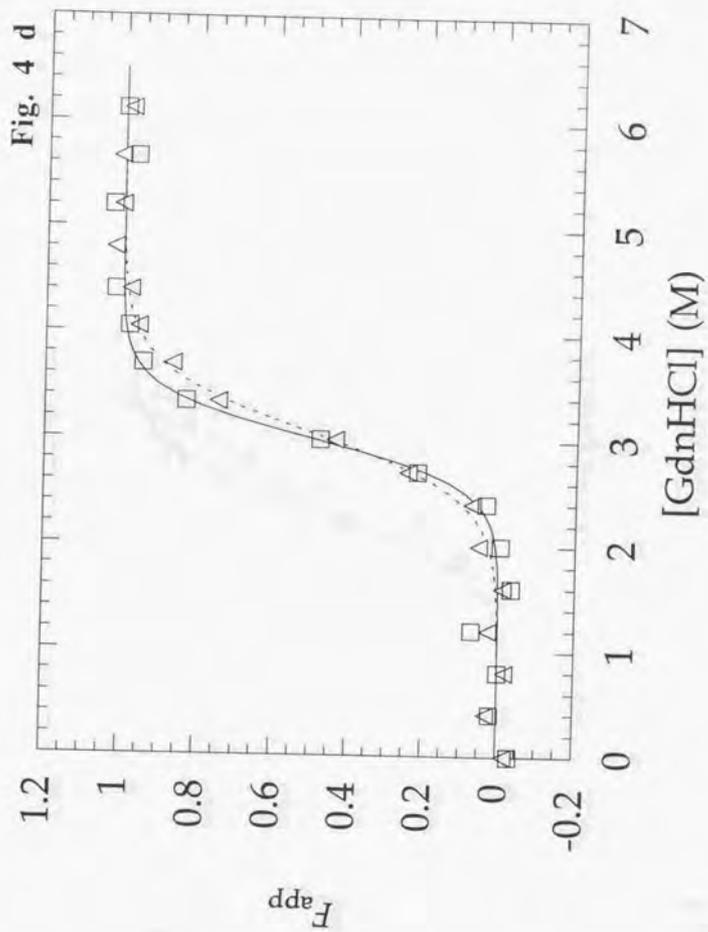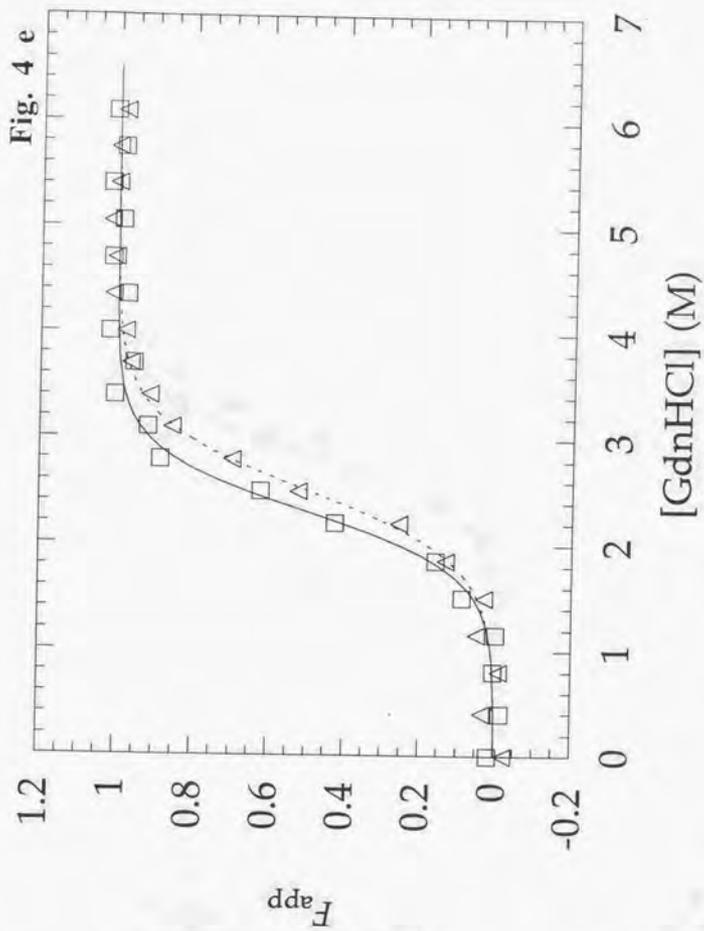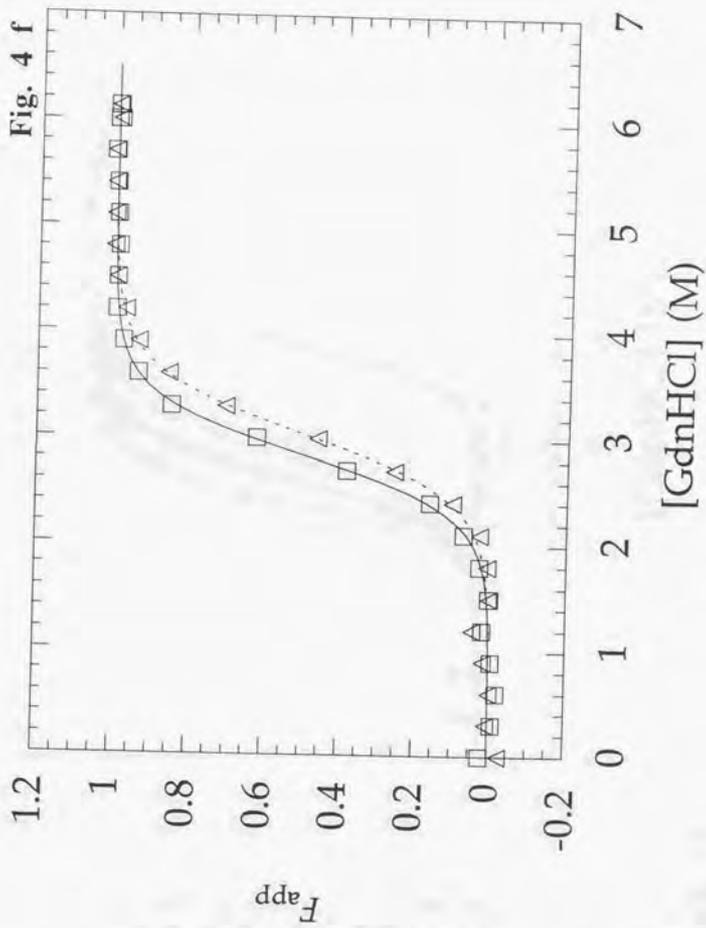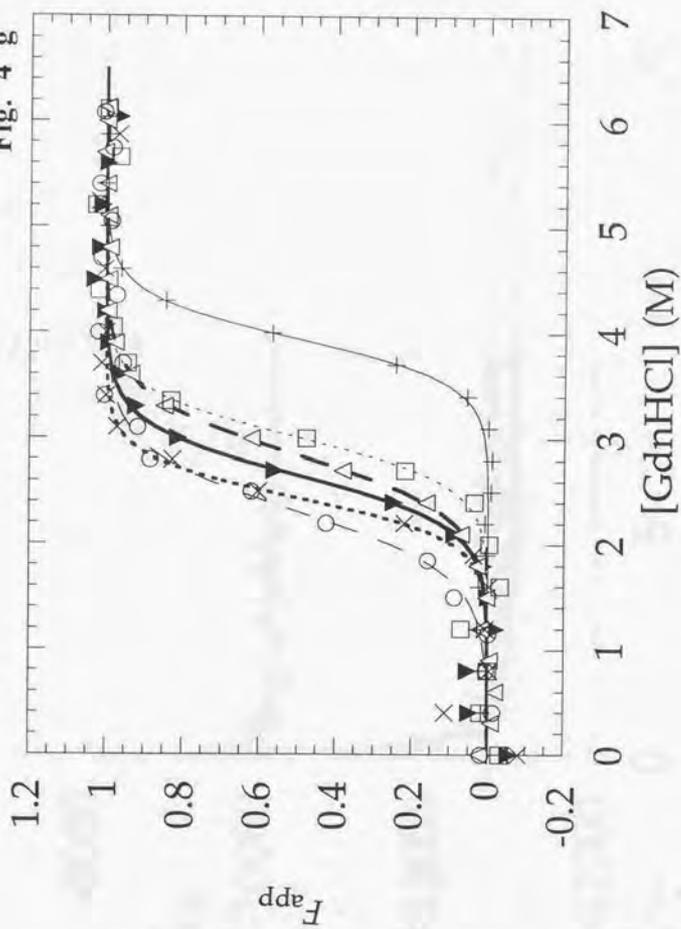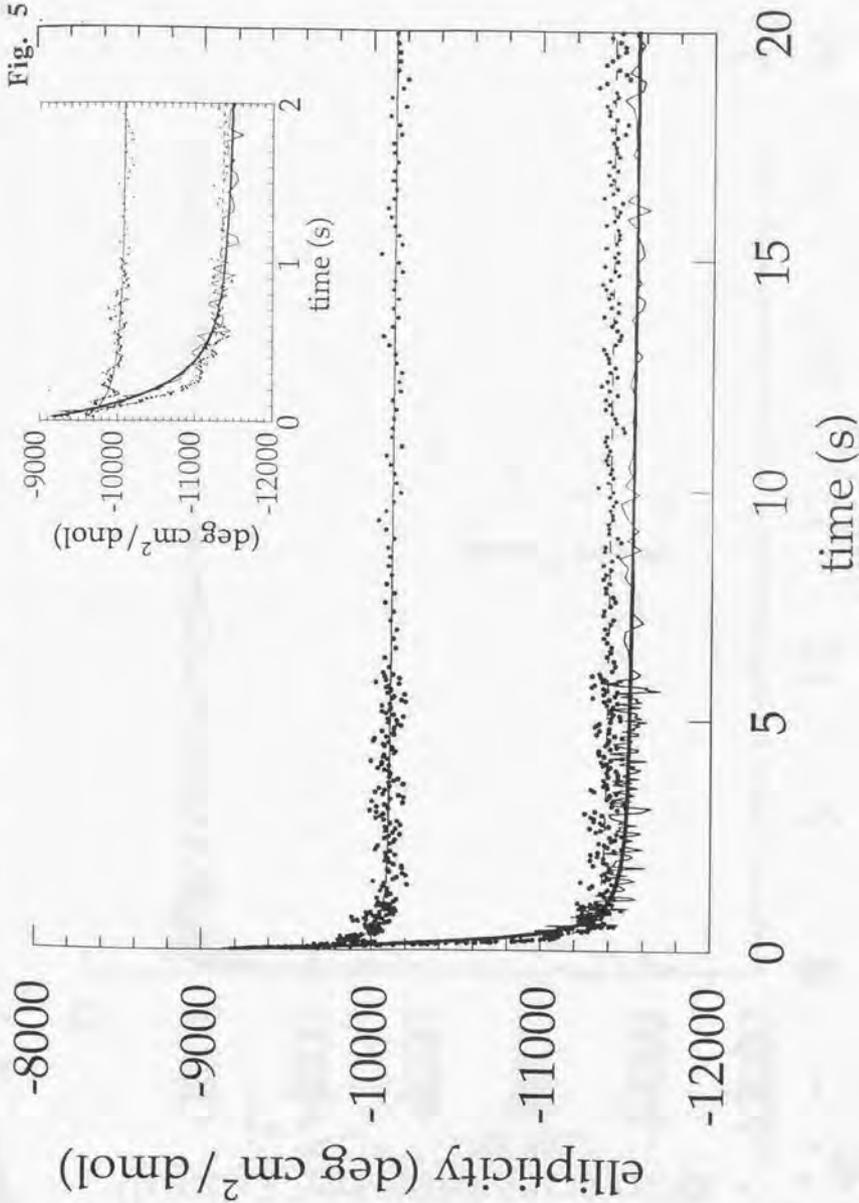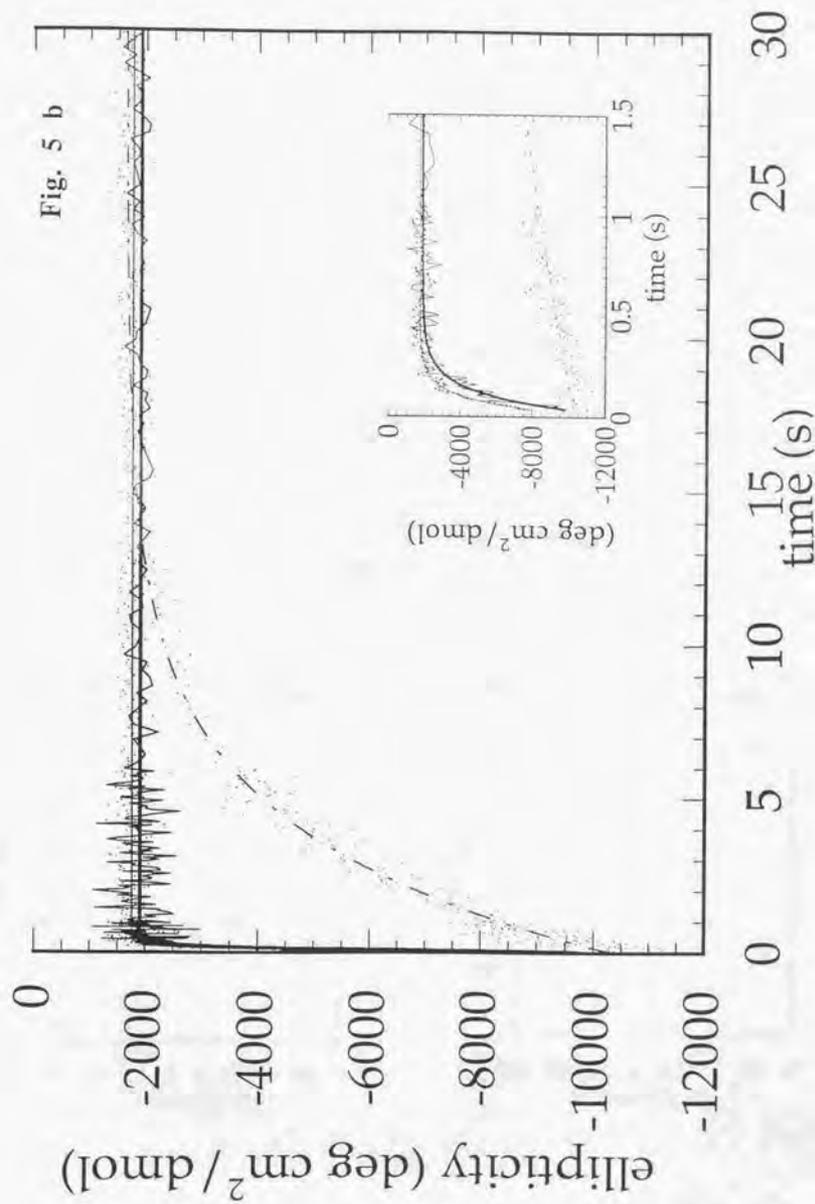
Fig. 6
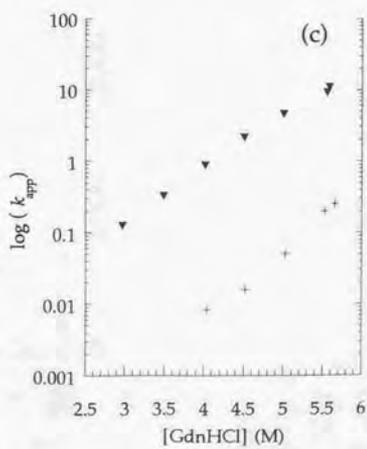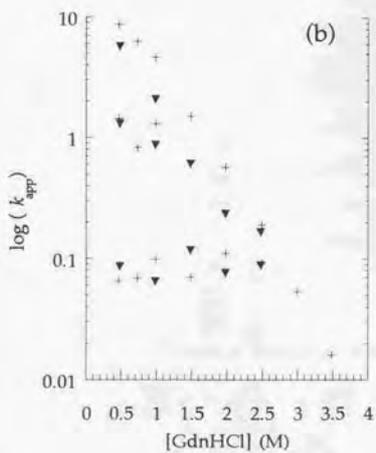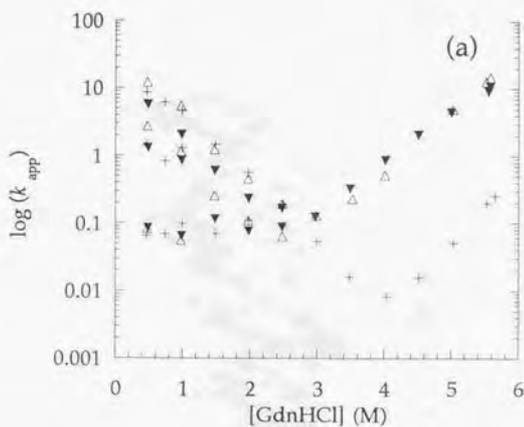
Fig. 7
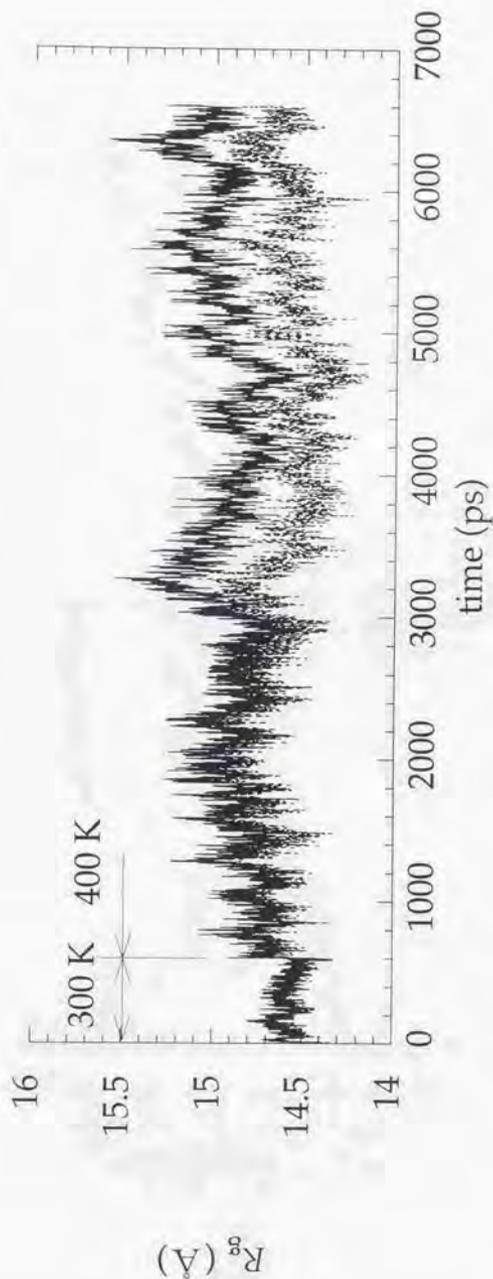
Fig. 8
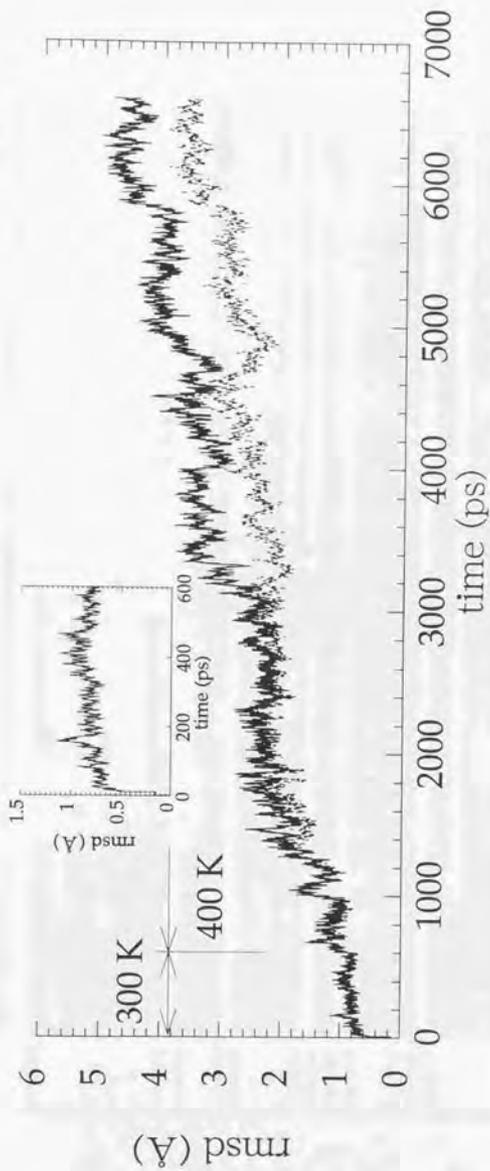
Fig. 9 a
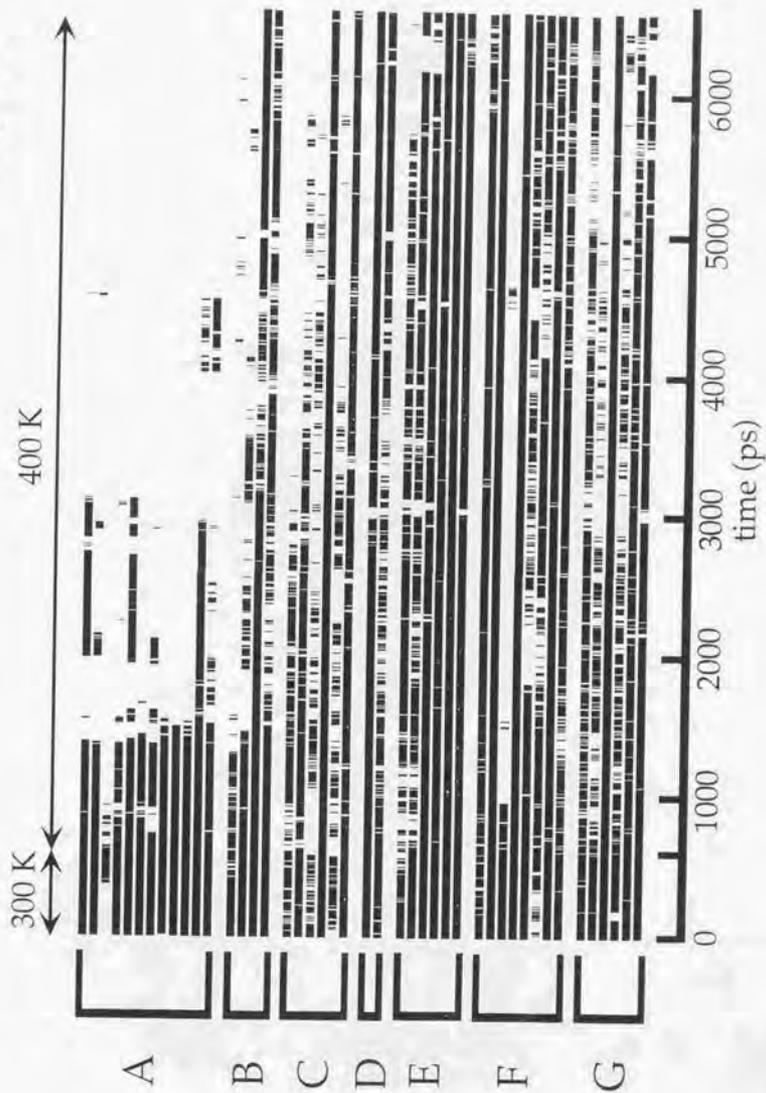
Fig. 9 b
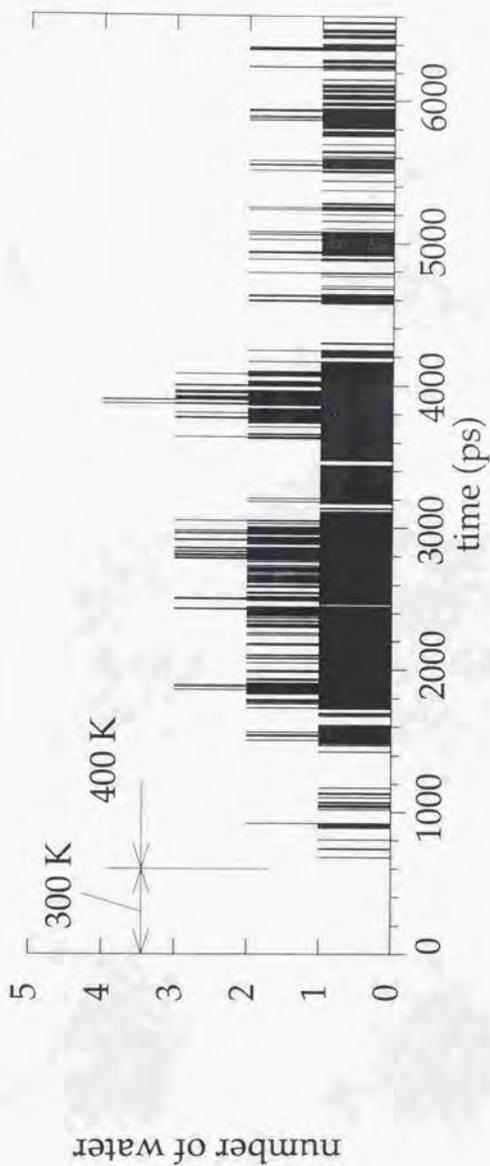
Fig. 10
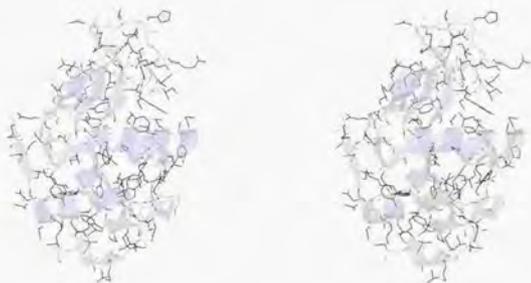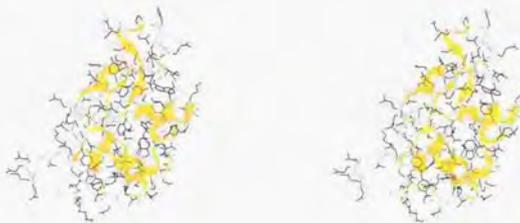
Fig. 10 (continued)

d

e
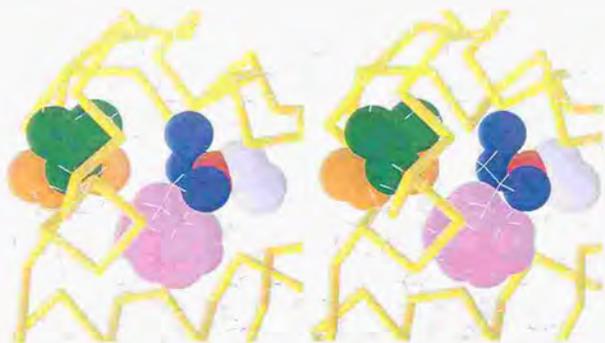
Fig. 11 a
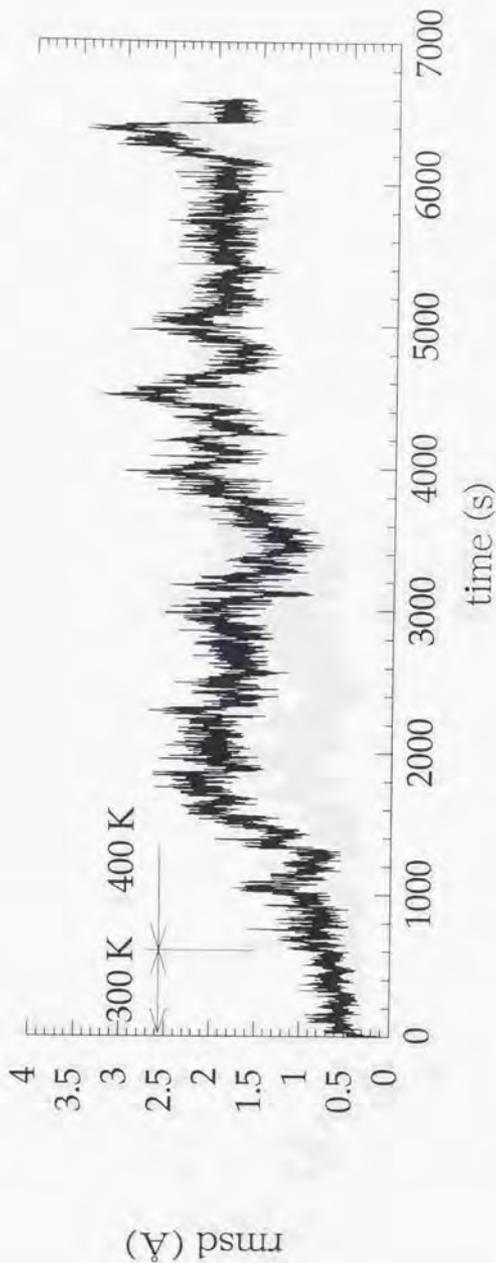
Fig. 11 b
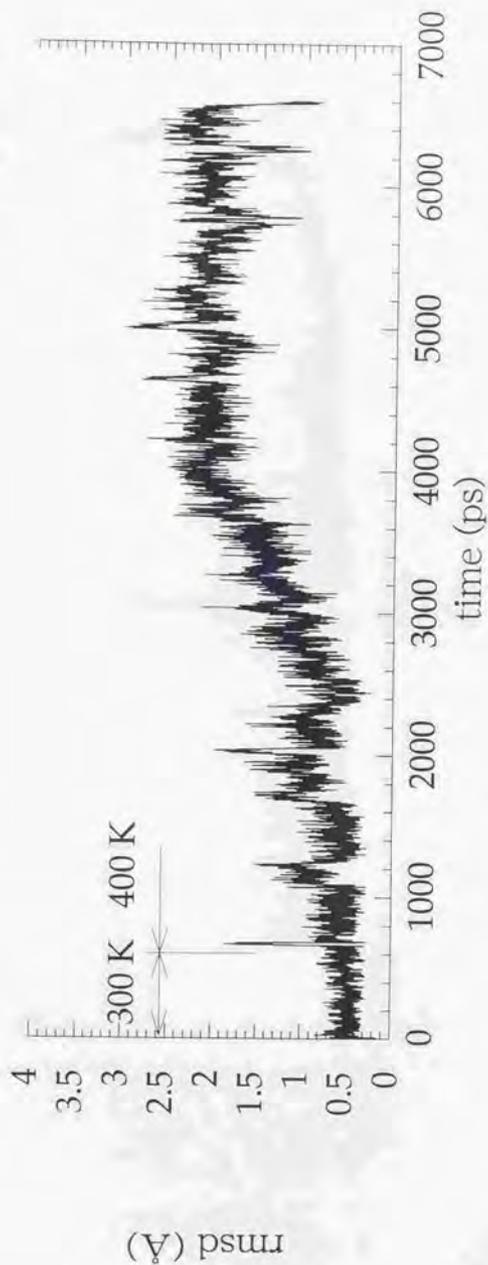
Fig. 11 c
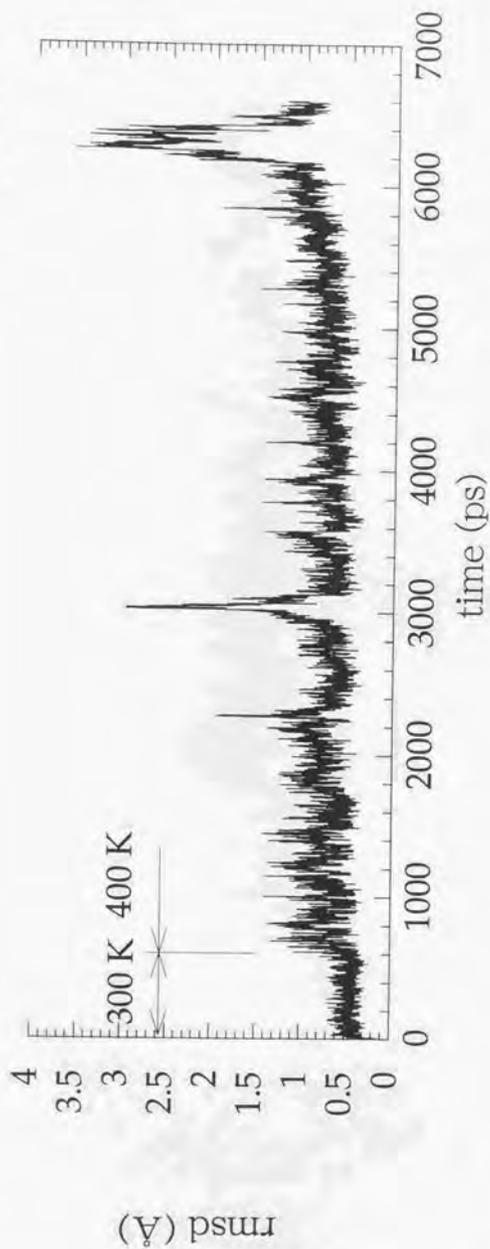
Fig. 11 d
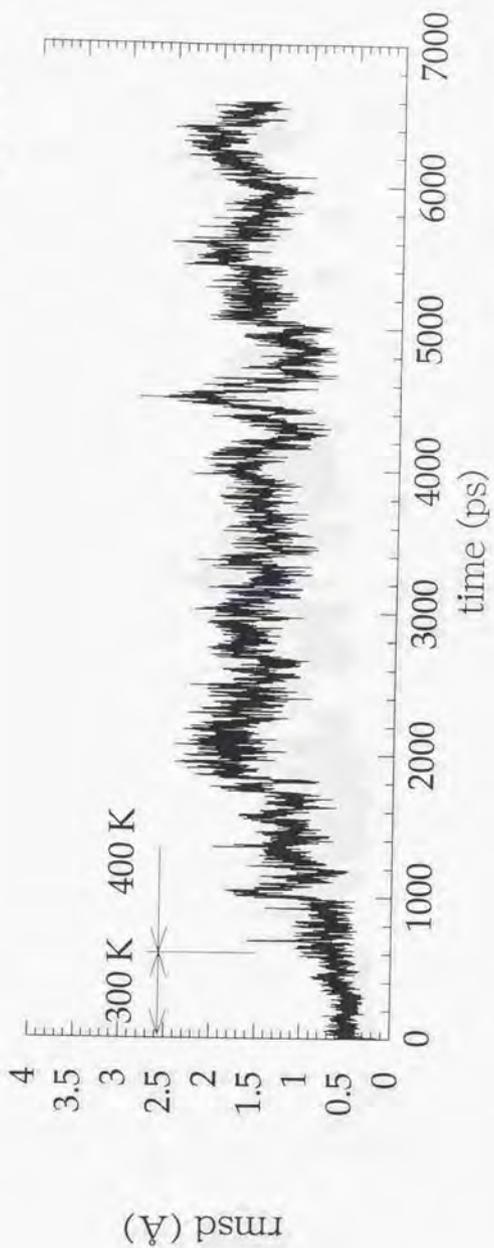
Fig. 11 e
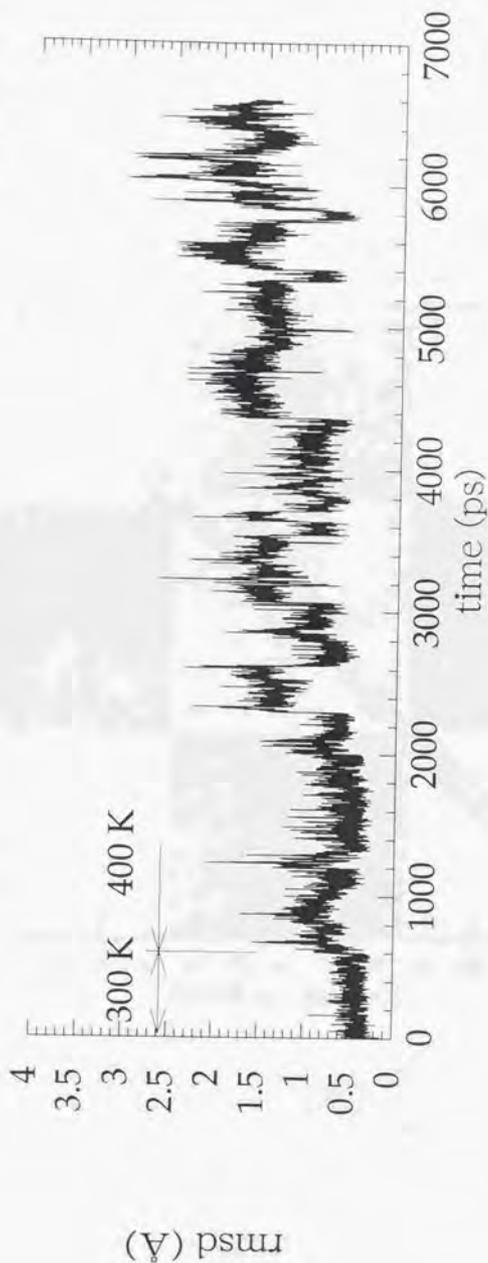
Fig. 12
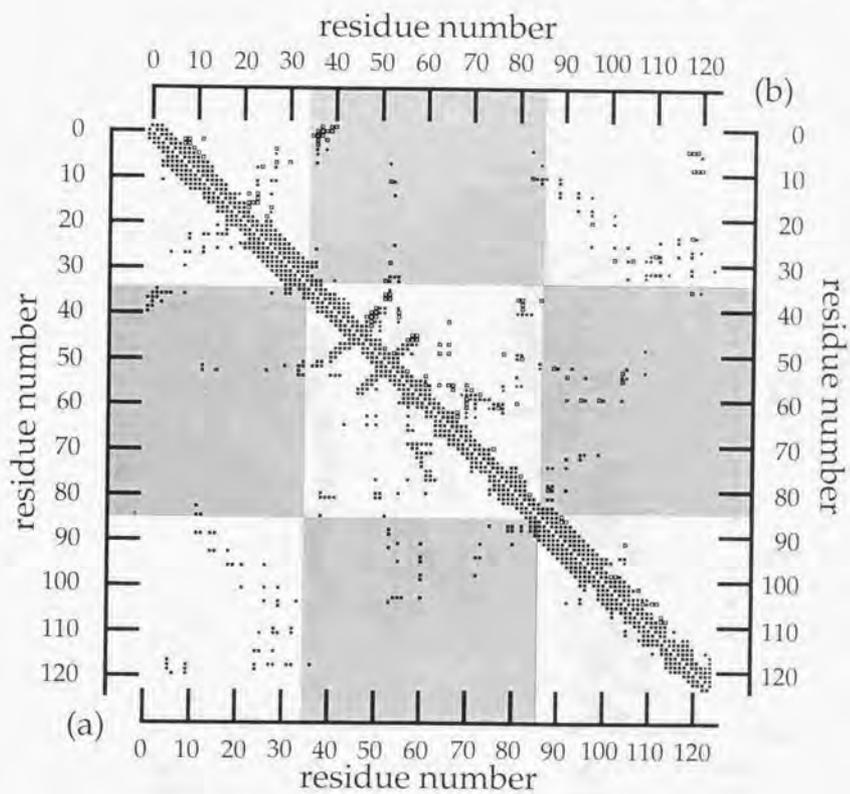
Fig. 13
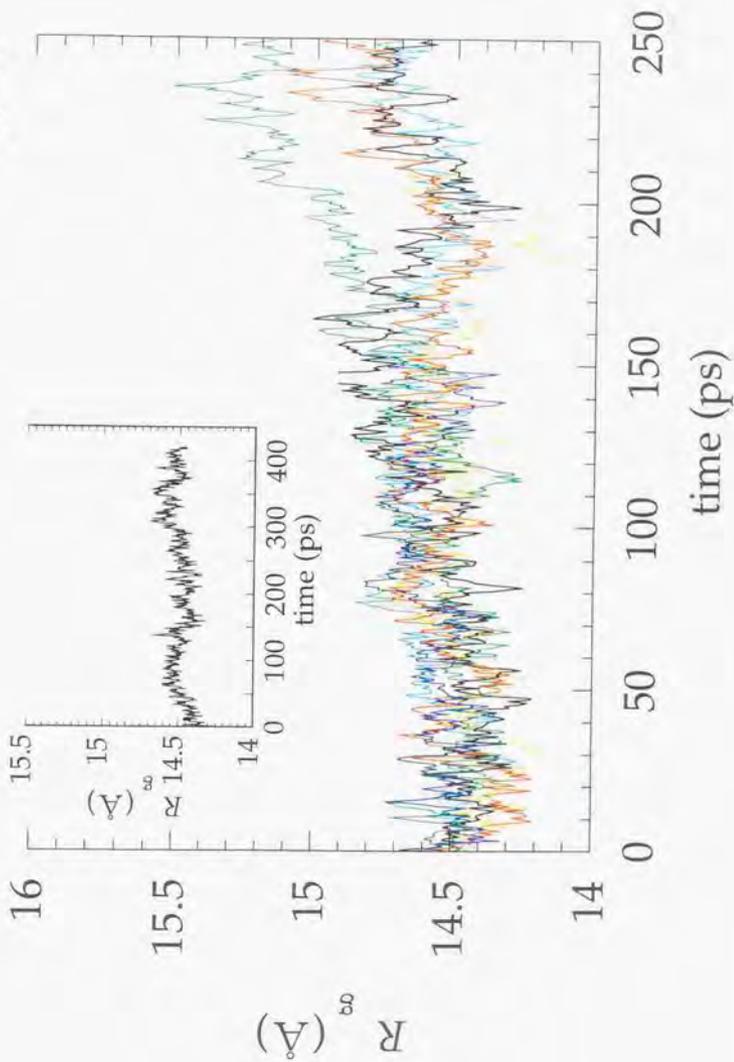
Fig. 14
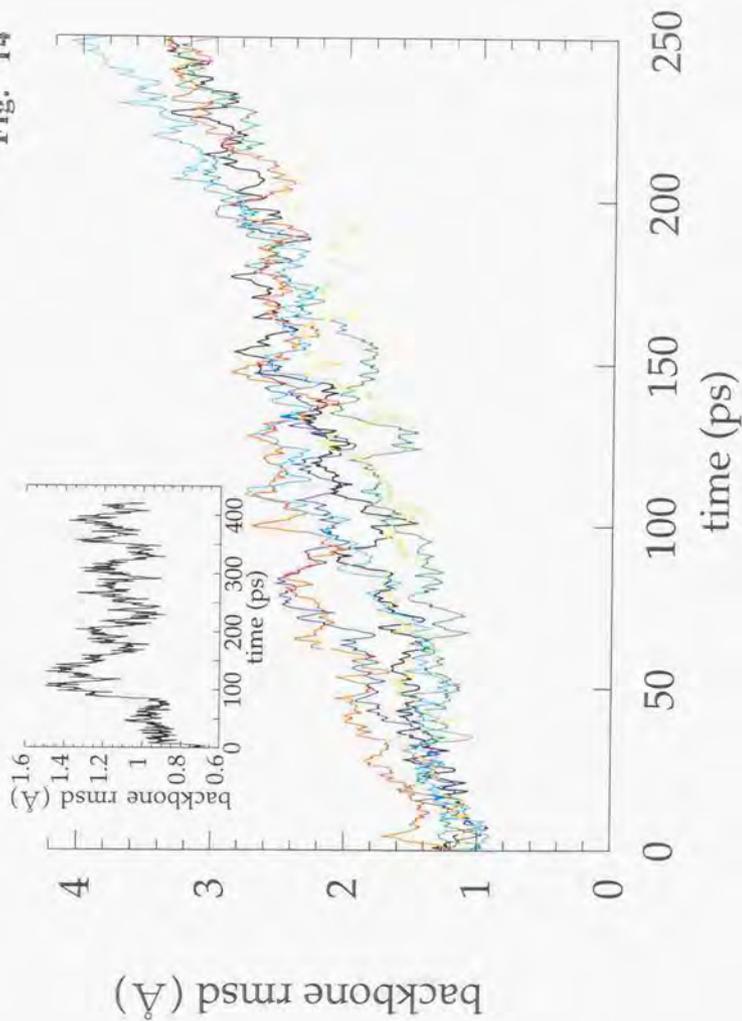
Fig. 15 a
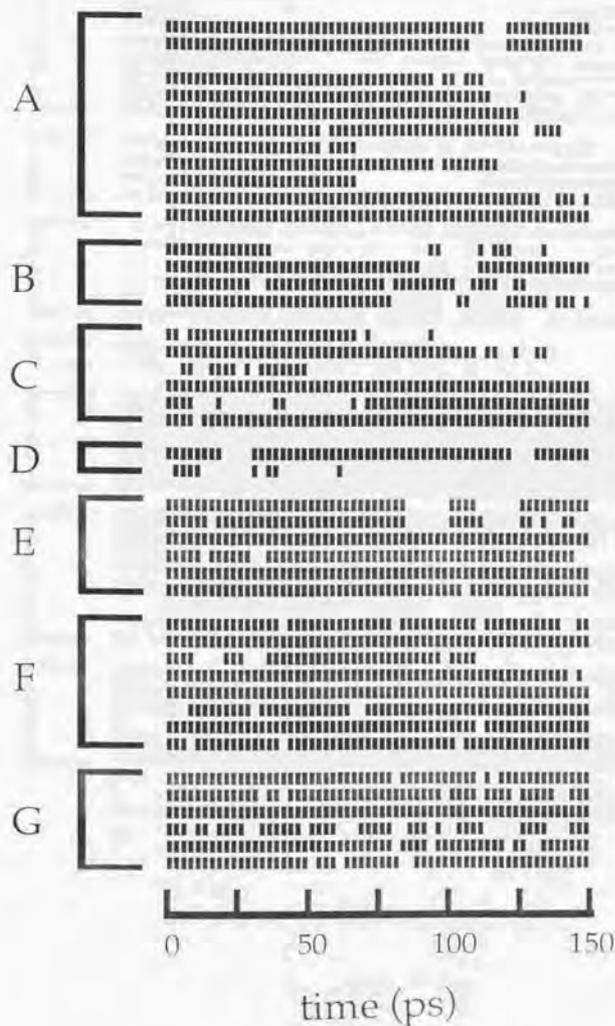
Fig. 15 b
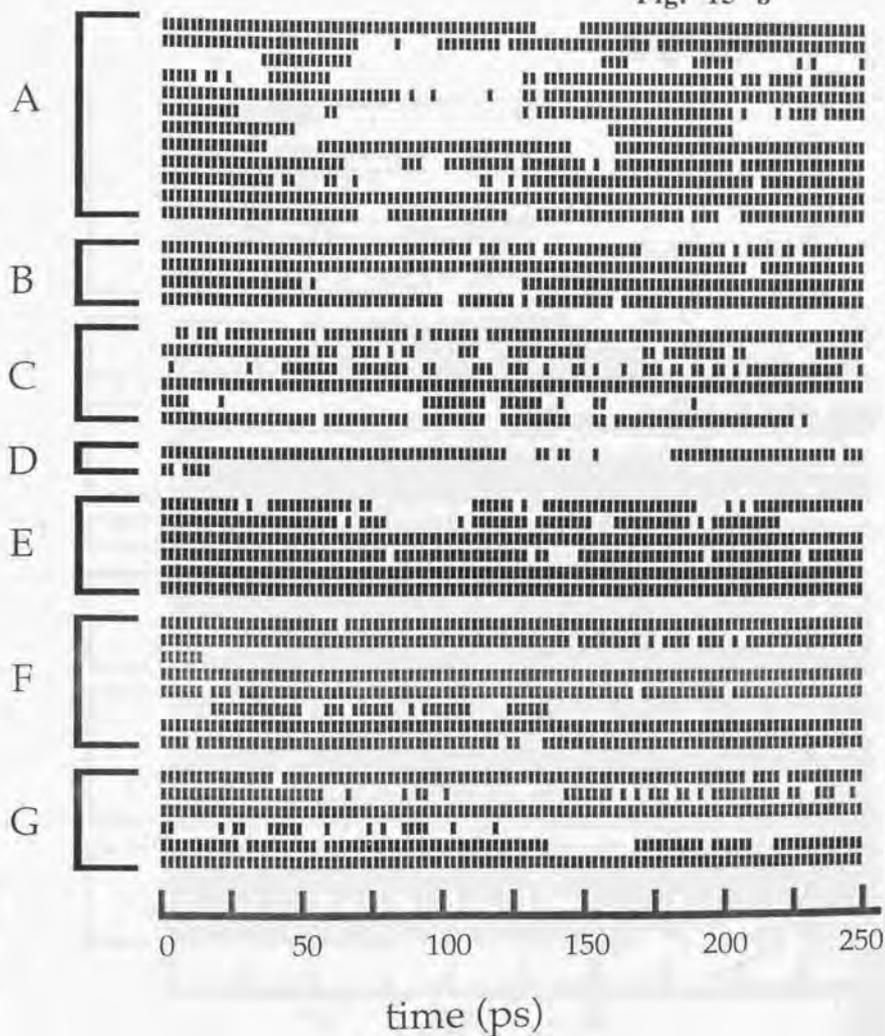
Fig. 15 c
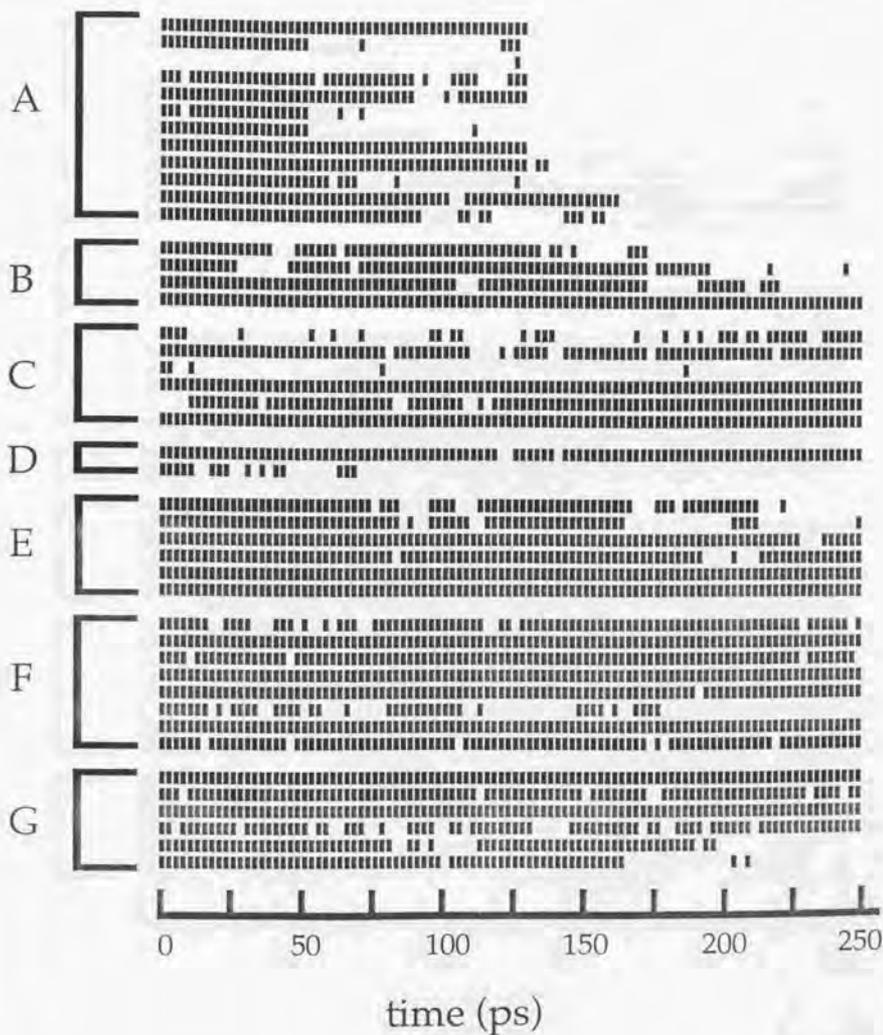
Fig. 15 d

Fig. 15 e

Fig. 15 f

time (ps)
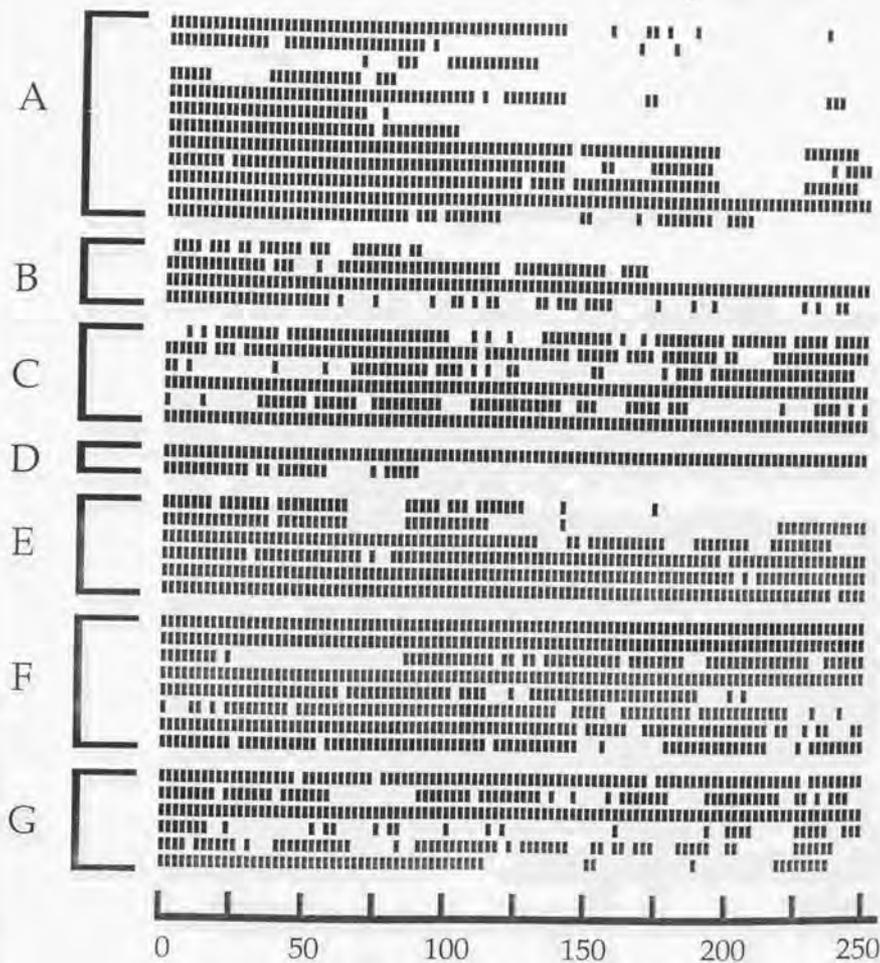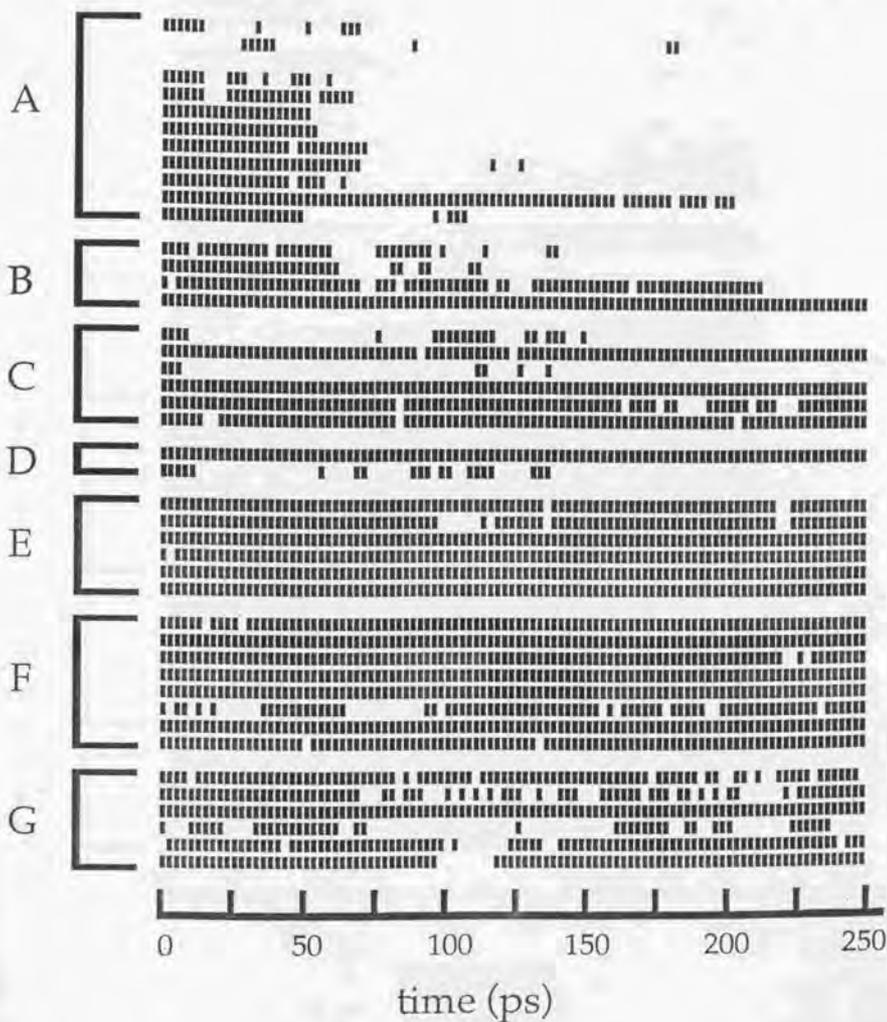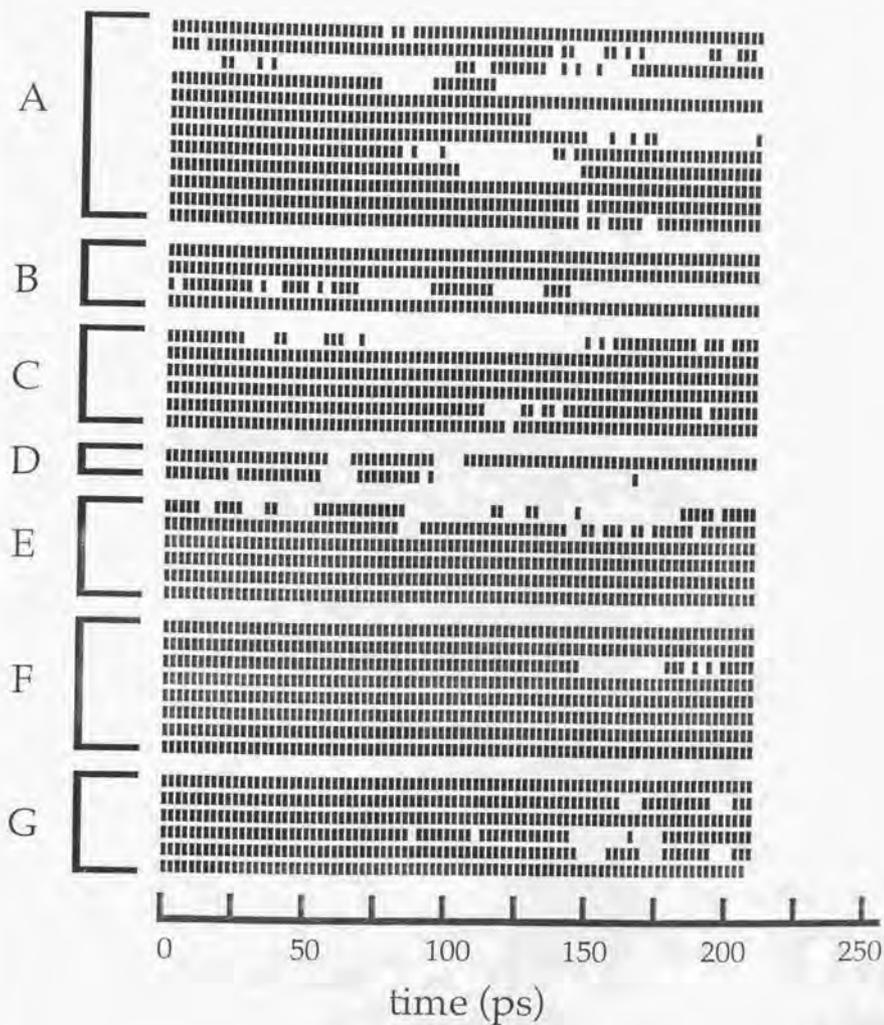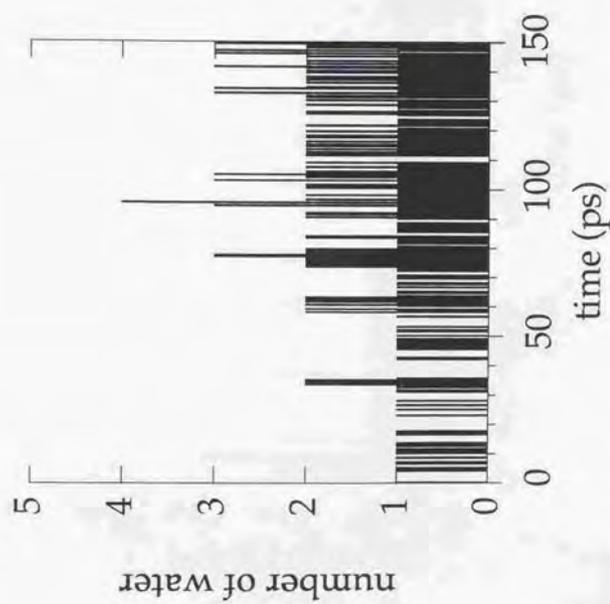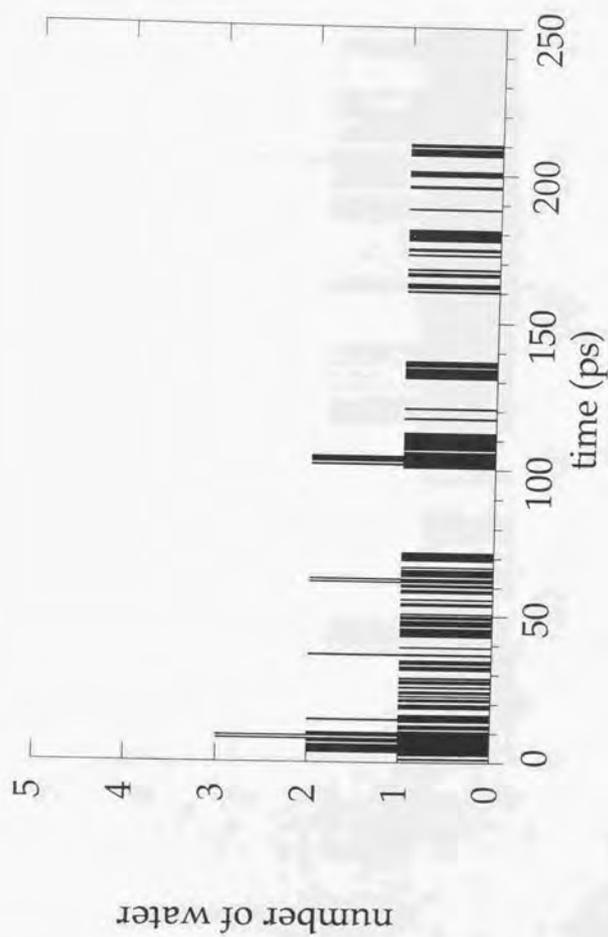
Fig. 15 g

Fig. 15 h

Fig. 15 i

Fig. 15 j

Fig. 15 k

Fig. 15

Fig. 16

α

β

α / β

Fig. 17