Enzyme specificity under dynamic control II: Principal component analysis of α-lytic protease using global and local solvent boundary conditions

NOBUYUKI OTA AND DAVID A. AGARD

Howard Hughes Medical Institute and Department of Biochemistry and Biophysics, University of California, San Francisco, San Francisco, California 94143–0448, USA

(RECEIVED March 1, 2001; FINAL REVISION April 10, 2001; ACCEPTED April 16, 2001)

Abstract

The contributions of conformational dynamics to substrate specificity have been examined by the application of principal component analysis to molecular dynamics trajectories of α-lytic protease. The wild-type α-lytic protease is highly specific for substrates with small hydrophobic side chains at the specificity pocket, while the Met190→Ala binding pocket mutant has a much broader specificity, actively hydrolyzing substrates ranging from Ala to Phe. Based on a combination of multiconformation analysis of cryo-X-ray crystallographic data, solution nuclear magnetic resonance (NMR), and normal mode calculations, we had hypothesized that the large alteration in specificity of the mutant enzyme is mainly attributable to changes in the dynamic movement of the two walls of the specificity pocket. To test this hypothesis, we performed a principal component analysis using 1-nanosecond molecular dynamics simulations using either a global or local solvent boundary condition. The results of this analysis strongly support our hypothesis and verify the results previously obtained by in vacuo normal mode analysis. We found that the walls of the wild-type substrate binding pocket move in tandem with one another, causing the pocket size to remain fixed so that only small substrates are recognized. In contrast, the M190A mutant shows uncoupled movement of the binding pocket walls, allowing the pocket to sample both smaller and larger sizes, which appears to be the cause of the observed broad specificity. The results suggest that the protein dynamics of α-lytic protease may play a significant role in defining the patterns of substrate specificity. As shown here, concerted local movements within proteins can be efficiently analyzed through a combination of principal component analysis and molecular dynamics trajectories using a local solvent boundary condition to reduce computational time and matrix size.

Keywords: α-lytic protease; principal component analysis; molecular dynamics; substrate specificity; solvent boundary condition

Our fundamental concepts of enzyme mechanisms are firmly based on the idea of complementarity between an enzyme and the reaction transition state (Pauling 1948). Through a combination of shape and electronic complementarity, enzymes select the appropriate substrates for a given reaction. Although there has been considerable experimental support for the importance of complementarity in determining specificity, for example from the combination of X-ray crystallographic and biochemical studies of enzyme-substrate complexes (Steitz et al. 1969; Fastrez and Fersht 1973; Brayer et al. 1979; Bone et al. 1987; Ding et al. 1994), such conclusions are generally based on a static description of protein conformation. More recently, the direct involvement of protein mobility in enzyme mechanics has been

Reprint requests to: David A. Agard, Howard Hughes Medical Institute and Department of Biochemistry and Biophysics, University of California, San Francisco, CA 94143-0448, USA; e-mail: agard@msg.ucsf.edu; fax: (415) 476–1902.

Article and publication are at http://www.proteinscience.org/cgi/doi/10.1101/ps.800101.

emphasized in fields such as nuclear magnetic resonance (NMR) imaging (Wüthrich 1986, 1995; Lipari and Szabo 1982), time-resolved X-ray crystallography (Farber 1997; Moffat 1997) and simulation methods (McCammon et al. 1977; van Gunsteren and Karplus 1982; Levitt 1983; Karplus and Petsko 1990; Kollman 1993; Karplus and Ichiye 1996). In general, protein flexibility appears to be useful in aiding the access of substrates to and the egress of products from the active site (Johnson et al. 1979). Flexibility has also been proposed to be coupled to the chemical steps of an enzyme reaction by directing the substrates to the transition state conformation (Johnson et al. 1979; Bone et al. 1989a). However, unlike such recognition of the importance of flexibility in the catalytic reaction, it has remained unclear whether or not protein dynamics are directly involved in specificity.

Previous experimental and theoretical studies of α-lytic protease (αLP) shed more light on the importance of dynamic motion in enzyme specificity. αLP, an extracellular serine protease from *Lysobacter enzymogenes*, has long served as an excellent model system for studies of enzymatic mechanisms (Hunkapiller et al. 1976) and for studies of the structural basis of substrate specificity (Bone et al. 1987, 1989a,b, 1991). Whereas the wild-type αLP shows a strong preference for the small hydrophobic residue, Ala, at the P1 site, mutation of the binding pocket residue Met190 to Ala (M190A) dramatically broadens specificity while maintaining or increasing catalytic activity (Fig. 1). The large changes in kcat/Km due to the mutation mainly result from alterations in Km and not changes in kcat (Bone et al. 1989a). It was clear from structural data that the mutant binding pocket was able to accommodate the broad range of substrates through an induced-fit mechanism. Although it

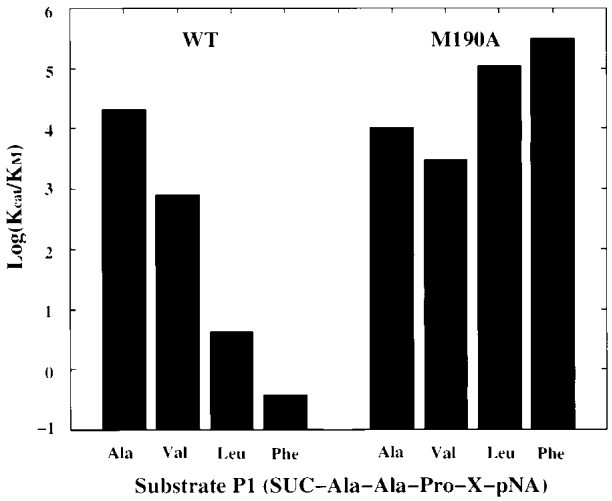

Fig. 1. The effect of mutations on specificity. Log(kcat/Km) is shown as a bar for each of five substrates, sAAP-X-p-Na where X = Ala, Val, Met, Leu, and Phe, for the wild-type enzyme and mutant M190A. (Bone et al. 1989a,b, 1991).

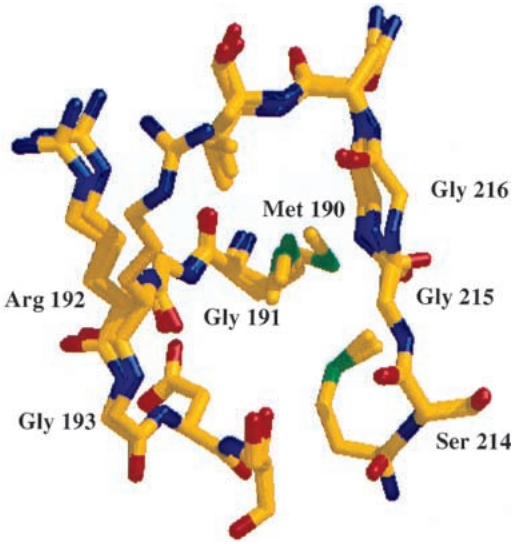

Fig. 2. Correlated motions in the S1 binding pocket of α-lytic protease obtained by multiconformer cryocrystallography. The subset of the multiconformers shows that two walls (Met190-Gly193 and Ser214-Gly216) move in the same direction with respect to the mean positions. The color of atom indicates the atom type: yellow, C; dark blue, N; red, O; green, S (Rader and Agard 1997).

appeared that the mutant was better able to adapt conformation than the wild-type αLP, the lack of change in crystallographic B factors did not suggest an increase in overall binding pocket flexibility (Bone et al. 1989a). From several lines of evidence, we proposed the following hypothesis for the significantly altered specificity in the mutant based on the patterns of concerted motion of the binding pocket walls. The walls around the wild-type S1 pocket (Met 190-Gly191-Arg192-Gly193 and Ser214-Gly215-Gly216) move in tandem (symmetric motion) such that the narrow size of the binding pocket is conserved, leading to the preference for residues with a small side chain, such as Ala. In contrast, the mutant M190A has a much broader specificity because the motion of the walls has become uncoupled (antisymmetric movement), allowing the S1 pocket to sample smaller and larger sizes. This hypothesis is supported by NMR studies showing slow exchange for binding pocket residues (Davis and Agard 1998), multiple conformation analysis of cryocrystallographic data indicating that the binding pocket walls can be trapped in multiple, closely-related conformations (Rader and Agard 1997) (Fig. 2), and normal mode analysis (NMA) (Miller and Agard 1999). In the latter study, the internal molecular vibrations of αLP in the wild-type and mutant forms were calculated in the presence and absence of ligand. Motions of the walls of the binding pocket were explored by taking the dot product of the eigenvectors with a vector set defining relative motion of the walls of the pocket, and summing the power over the modes. These calculations indicated the dominant presence

of symmetric movements in the wild-type S1 pocket and the much greater tendency towards antisymmetric movements in the mutant pocket. Although the results of NMA on αLP support the hypothesis of dynamic control of the substrate specificity, there are some potential concerns about the analysis. First, in the NMA model, the movements of all atoms are described as simple harmonic motions. It is known that, in some cases, this assumption can introduce significant error (Teeter and Case 1990; Ichiye and Karplus 1991; Hayward et al. 1994, 1995). Secondly, the effect of solvent on motional dynamics is not explicitly included in the in vacuo NMA calculations.

To test the results previously obtained by NMA, here we performed principal component analysis (PCA) (Levy et al. 1984; Teeter and Case 1990; Kitao and Go 1999) of αLP molecular dynamics (MD) trajectories using an explicit solvent model (Fig. 3A). Although PCA provides a useful tool to study protein dynamics and functions, its application has been limited to small molecular systems due to the associated computational cost of molecular dynamics simulations and diagonalization of a large covariance matrix. We therefore tried to develop a novel PCA method which requires only a local MD trajectory of a large biological system. The local simulation is carried out only within a focused volume using a solvent boundary condition. To our knowledge, PCA using a localized MD trajectory with a solvent boundary condition has not been reported, and it is thus necessary to verify the accuracy and validity of the method. Therefore, we compared the modes derived from a local MD trajectory with a local solvent boundary condition (LBC) and a global MD trajectory with a global solvent boundary condition (GBC) of α-lytic protease. By using the P_j test (Kitao and Go 1991), we found that the low-frequency modes from the two distinct trajectories showed good accordance, suggesting that the PCA using a local trajectory is accurate enough for our current study. Errors are mainly observed around the high-frequency modes, as expected by theoretical analysis (see Appendix). Since most protein functions such as ligand binding and induced fit are associated with low-frequency motions, the errors in high-frequency modes do not present a serious drawback for most protein studies. After the validation of the PCA using a local trajectory, molecular dynamics simulations with an LBC centered at the binding pocket were carried out for 1 nanosecond (nsec) for both the wild-type and mutant enzymes. From the localized trajectories, we could efficiently derive the low-frequency modes of the two walls of the binding pocket for both proteins. To compare the current PCA results with previous in vacuo NMA results (Miller and Agard 1999), we summarized cumulative dot products of the two walls for first 100 lowest modes in exactly the same fashion as that used by Miller and Agard. According to the dot product analysis, in the

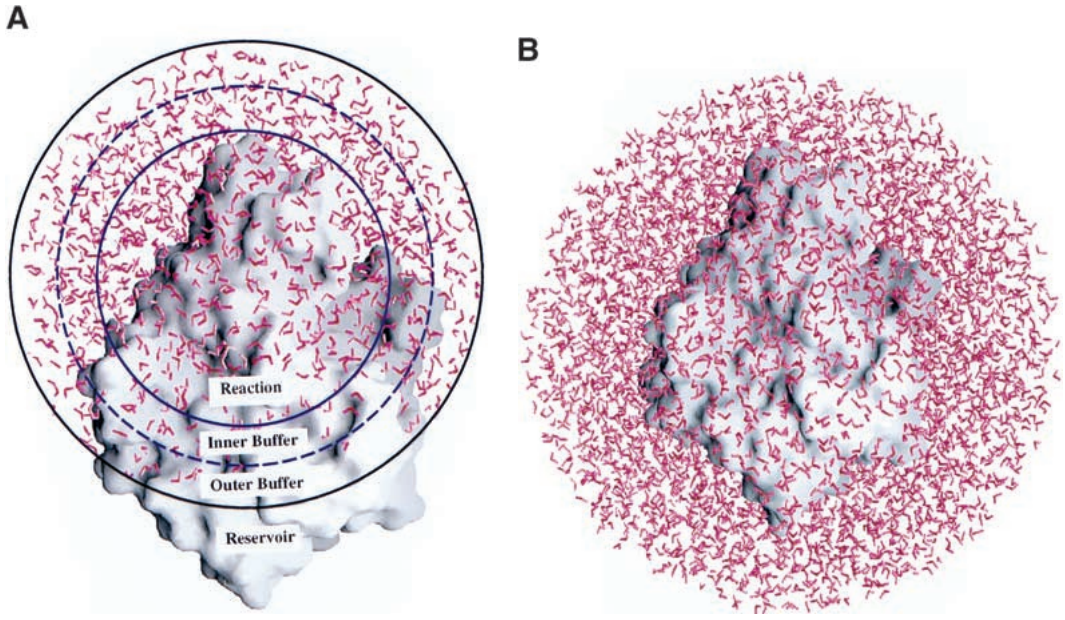

Fig. 3. (*A*) S1 binding pocket of α-lytic protease within a local solvent boundary condition (LBC). Outside of the sphere is a reservoir zone where all of the atoms are fixed during the simulation. The first and second outer shells are an outer buffer zone and an inner buffer zone, respectively. In the outer buffer zone, all protein atoms are restrained to the X-ray positions. In the inner buffer zone, only backbone protein atoms are restrained to the X-ray positions. In those buffer zones, Lanvegin dynamics are used to reduce the artificial effect of the boundary condition. Inside the inner sphere is a reaction zone where all atoms and solvents move freely without any restraint function and standard molecular dynamics simulations are carried out. (*B*) α-lytic protease within a global solvent boundary condition (GBC). The center of mass of αLP is located at the center of a water droplet with a radius of 35 Å. No restraint functions are imposed on any atoms of αLP, and the center of mass motion was turned off during the 1 nsec simulation.

wild-type case, symmetric movement of the specificity walls is dominant over antisymmetric movements, while in the case of the M190A mutant, more antisymmetric movement was observed. The PCA results show, in general, the same trend as that seen in the in vacuo NMA and reconfirm our hypothesis on the importance of collective motions in the mechanism of substrate specificity.

Results

Comparing modes from LBC and GBC analyses

The combination of PCA with a local MD simulation using a local solvent boundary condition is very useful for cases in which interest is focused on a specific region of protein, such as the active site. Because this is a novel application, it is necessary to confirm the accuracy and validity of this approach by comparing the LBC method (Fig. 3A) with the global solvent boundary condition in which αLP is located at the center of a large solvent droplet without any restraints on the protein (Fig. 3B). In principle, if the local MD using LBC can well duplicate the properties of the global MD trajectory with GBC, the resultant PCA modes from a local trajectory should be very similar to those from a global trajectory and should prove useful for dynamics studies of macromolecules.

In the LBC method, the system was divided into four regions, because the size of the spherical boundary is fairly large. As shown in Figure 3A, the first region, called the "reaction zone," includes all protein atoms within a radius of 15 Å of the site of interest. In the reaction zone, all atoms are free to move without any positional restraints, and standard MD simulations are carried out. The second and third regions are called "buffer zones." All backbone protein atoms in the first buffer zone (within the inner shell in Fig. 3A) and all atoms in the second buffer zone (within the outer shell in Fig. 3A) are restrained to their initial X-ray structure positions (Fujinaga et al. 1985) using a harmonic potential. In the buffer regions, Langevin dynamics (Brünger 1992) are performed in order to introduce stochastic motions to water molecules. The fourth "reservoir zone" contains all atoms outside the spherical solvent boundary, which are kept fixed to the X-ray structure (Fujinaga et al. 1985). In the GBC method, the entire αLP was solvated in a large solvent droplet of radius 35Å without imposing any restraints on the protein (Fig. 3B). Because a limited number of atoms are allowed to move in the LBC method, the simulation cost is significantly reduced to one-fifth of the computational time required for the GBC method (156 h for the LBC method vs. 884 h for the GBC method using an SGI R10,000 workstation). The matrix size of PCA is also reduced to 837^2 from 2376^2 when backbone atoms are selected for PCA calculations.

Comparison of modes derived from these methods is sensitive to appropriate superimposition of the relevant structures. Here we describe that these structures are very close and comparable, which means that in the P_j test, the inner products between the normal mode vectors derived from the two different simulations will not be much affected qualitatively by the differences between the two averaged structures. As illustrated in Figure 4, the best fit between the LBC averaged structure (shown in blue) and the GBC averaged structure (red) over the 1 nsec simulations was calculated using the backbone atoms in the reaction zone. The root mean squared deviations (RMSDs) obtained by fitting the original X-ray structure (Fig. 4, white; Fujinaga et al. 1985) over all main chain heavy atoms are 0.39 Å for the LBC structure and 0.83 Å for the GBC structure. Because with the GBC method no restraint functions are imposed on any atoms of αLP, the GBC structure deviates slightly from the original X-ray structure, especially around solvent accessible loops. With the LBC method, in contrast, the atoms outside the spherical boundary are rigid and the atoms in the inner shells are restrained, thus keeping the entire αLP very close to the original X-ray structure even after 1 nsec simulation. Moreover, as seen in Figure 4, A and B, the structural deviations around the binding pocket and the two walls are quite small. Therefore we can proceed with a more detailed analysis, the P_j test, in order to check the validity of the LBC PCA by comparing it to PCA using the GBC method.

Because of the differences in the number of atoms and the use of independent simulations in the two calculations, the calculated modes are not expected to be directly comparable. Instead, we calculate the "projection values" of one mode set onto another mode set, as given by Kitao, Hayward, and Go (Kitao and Go 1991; Kitao et al. 1994; see Materials and Methods for the detailed P_j test procedure). In general, the P_j test is often used for comparison of NMA results using different force fields. Here we used the P_j test in order to check whether the two different methods can provide qualitatively similar results (see Materials and Methods and Appendix). In Figure 5, the Projection values, P_j, are plotted against the mode frequency from the GBC analysis. This P_j plot shows the degree of similarity to which each GBC mode can be expressed in terms of the LBC modes. When P_j is 1, it means that each GBC mode is completely expressed by a linear combination of all LBC modes. When P_j is small (< 0.3), each GBC mode cannot be expressed accurately enough by the LBC modes. In general, the P_j values are very high around the low-frequency region. For the ten lowest frequency modes, the averaged P_j value was 0.83 after 1 nsec. However, one cannot expect extremely high P_j values (~ 1.0) due to the aforementioned small deviations between the LBC and GBC structures (Fig. 4A). In the high-frequency region (> 200 cm^{-1}), P_j values decrease to around 0.4. The higher the frequency, the more error is expected due to the effect of the restrained atoms in

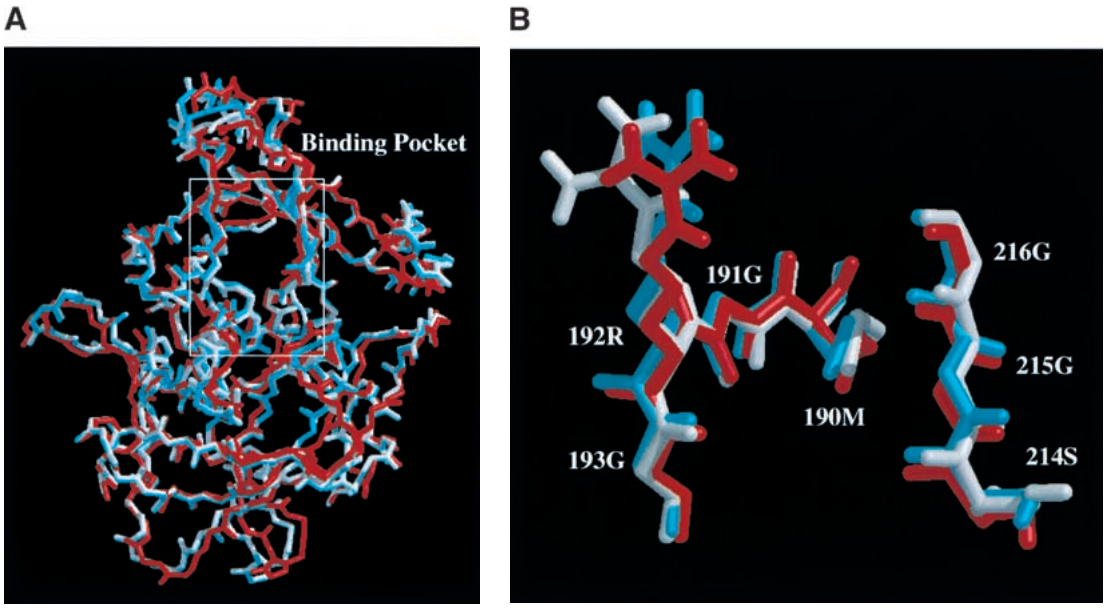

Fig. 4. (*A*) Comparison of the backbone atom structures of αLP. Superposition of the LBC averaged structure (blue) and the GBC structure (red) onto the crystal structure (white) using the backbone atoms of the reaction region. The region around the binding pocket is shown in the white rectangular box. The LBC structure is very close to the crystal structure. The deviation between the GBC structure and the crystal structure is mainly seen around the outer loop region. (*B*) Comparison of the binding pocket region of αLP. All atom representations of the three structures, that is, the crystal structure (white), the averaged LBC structure (blue), and the averaged GBC structure (red) show that the averaged structures at the binding pocket are almost overlapped to the crystal structure except for the long side chain of Arg 192.

the LBC methods, according to theoretical analyses using a static reduction method (Ookuma and Nagamatsu 1984) (see Appendix). Also, a convergence problem of the covariance-variance matrix using a 1 nsec simulation trajectory appears to be another source of the error in the high-frequency modes (Balsera et al. 1996). In another study, it was shown that a few collective coordinates are sufficient to describe functionally important movements of proteins. For example, a large hinge-bending motion of T4 lysozyme shown by crystallographic studies can be well described using only the lowest frequency modes derived from MD simulations (de Groot et al. 1998). Since P_j values of the LBC method are very high qualitatively for the low-frequency modes, the LBC method should be very useful to calculate the low-frequency modes, providing information on correlated local motions within large macromolecules.

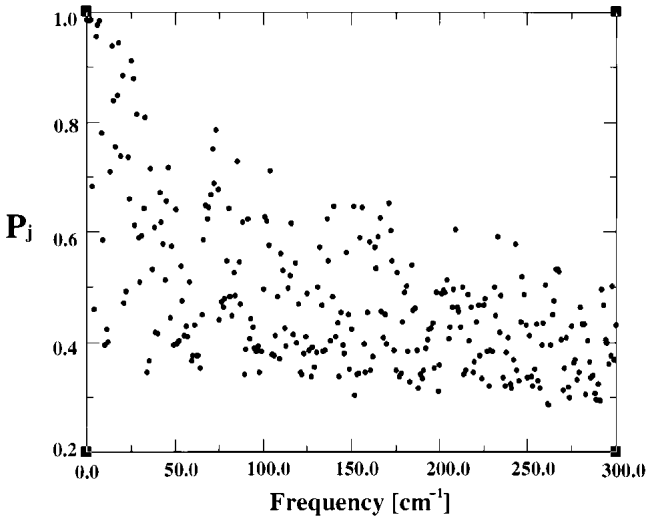

Fig. 5. Projection P_j of the ith GBC mode onto the LBC subspace. The averaged P_j value of the first ten lowest frequency modes is larger than 0.8 after 1 nsec simulation.

Symmetric and antisymmetric movements of the walls at the S1 pocket

To retest the previous NMA results and the role of dynamic motions in specificity, we performed the same dot product analysis (Miller and Agard 1999) to quantify symmetric and antisymmetric motions of the S1 pocket walls, but using the modes calculated by the PCA method. Figure 6, A and C, shows the cumulative dot products over the 100 lowest frequency modes for the wild-type and the M190A mutant, respectively. In the wild-type case, the symmetric movement is clearly dominant relative to the antisymmetric movement in the region of the S1 pocket. In the mutant, the antisymmetric movement shows much larger amplitude compared to the wild-type value and is more dominant relative to the symmetric movement.

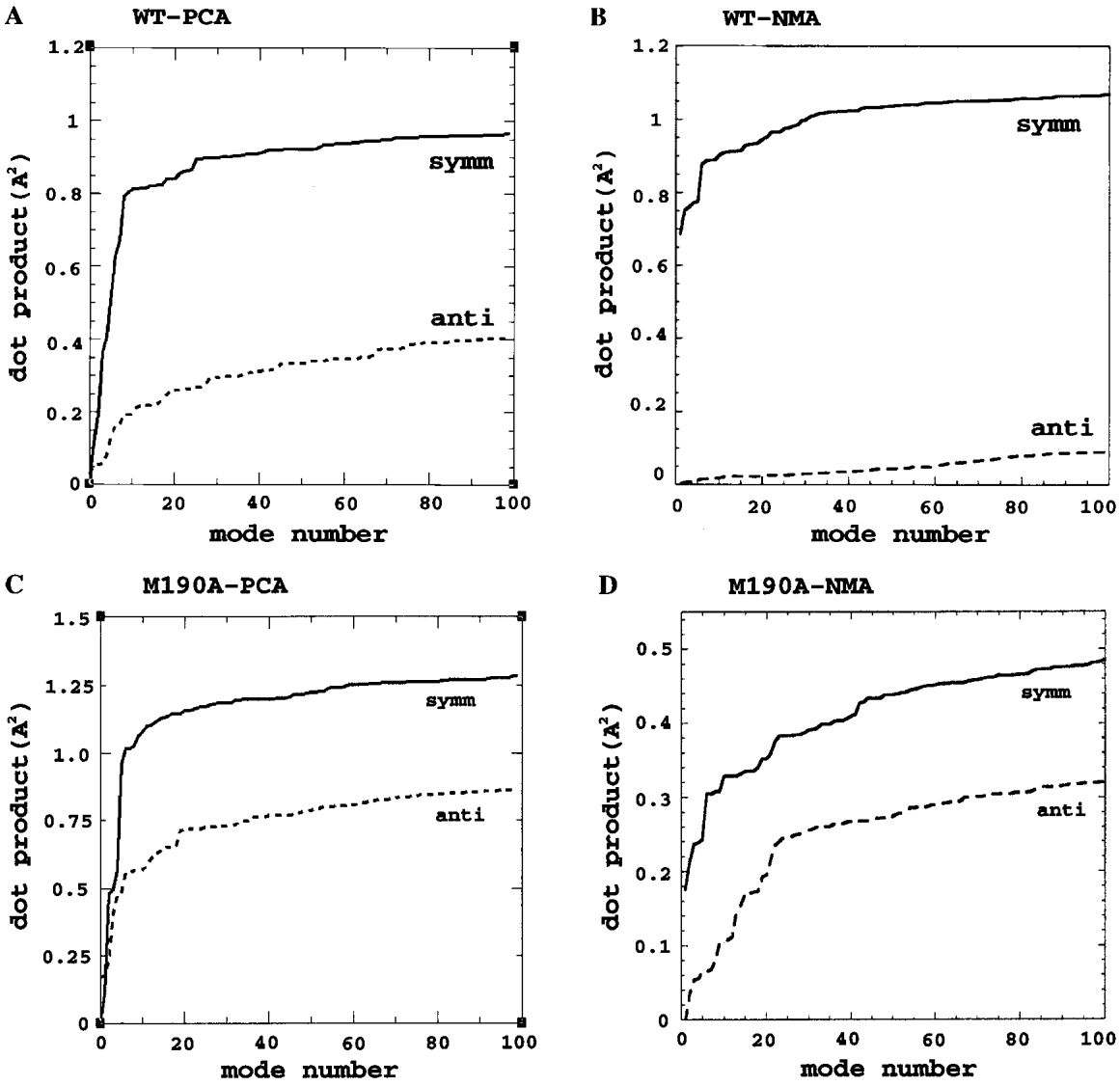

Fig. 6. Dot product calculations of two walls at the S1 binding pocket. The cumulative dot products are plotted for the 100 lowest frequency modes for comparison with the NMA results obtained by Miller and Agard (1999). (*A*) For the wild-type PCA (WT-PCA), αLP shows that symmetric movement is dominant in relation to antisymmetric movement as shown by NMA (*B*). (*C*) For the M190A mutant PCA, more antisymmetric movement is observed relative to the symmetric movement. The amplitudes are significantly increased compared with those of NMA (*D*) due to the solvent effect.

The PCA dot product results (Fig. 6A,C) provided the same trend as that seen in in vacuo NMA (Fig. 6B,D). However, an increase in amplitude of the dot product analysis was observed. It is difficult to directly compare the PCA results with the NMA results since the first 100 modes of PCA do not correspond to those of NMA. However, the increase in amplitude of the dot product of PCA appears to be due to the explicit treatment of the solvent. In the MD simulation in the explicit solution phase, water molecules constantly enter and egress from the binding pocket, which leads to the larger deviation and more dynamic motions around the binding pocket. This may indicate that the solvent water molecules play important roles in dynamic movements of protein and enzyme specificity.

Visual portrayal of the lowest (non-zero) frequency mode, representative of major motions, shows that it is consistent with the dot product analysis and our hypothesis. The top view and the side view of the lowest frequency mode of wild-type αLP are shown in Figure 7, A and B, respectively. This clearly indicates that the two walls move symmetrically so that the size of the pocket is well preserved. The main movements of the walls are the back-and-forth tandem movement of the wall (Met 190-Gly193) against the S1 pocket and the swing motion of Gly215-Gly216 keeping

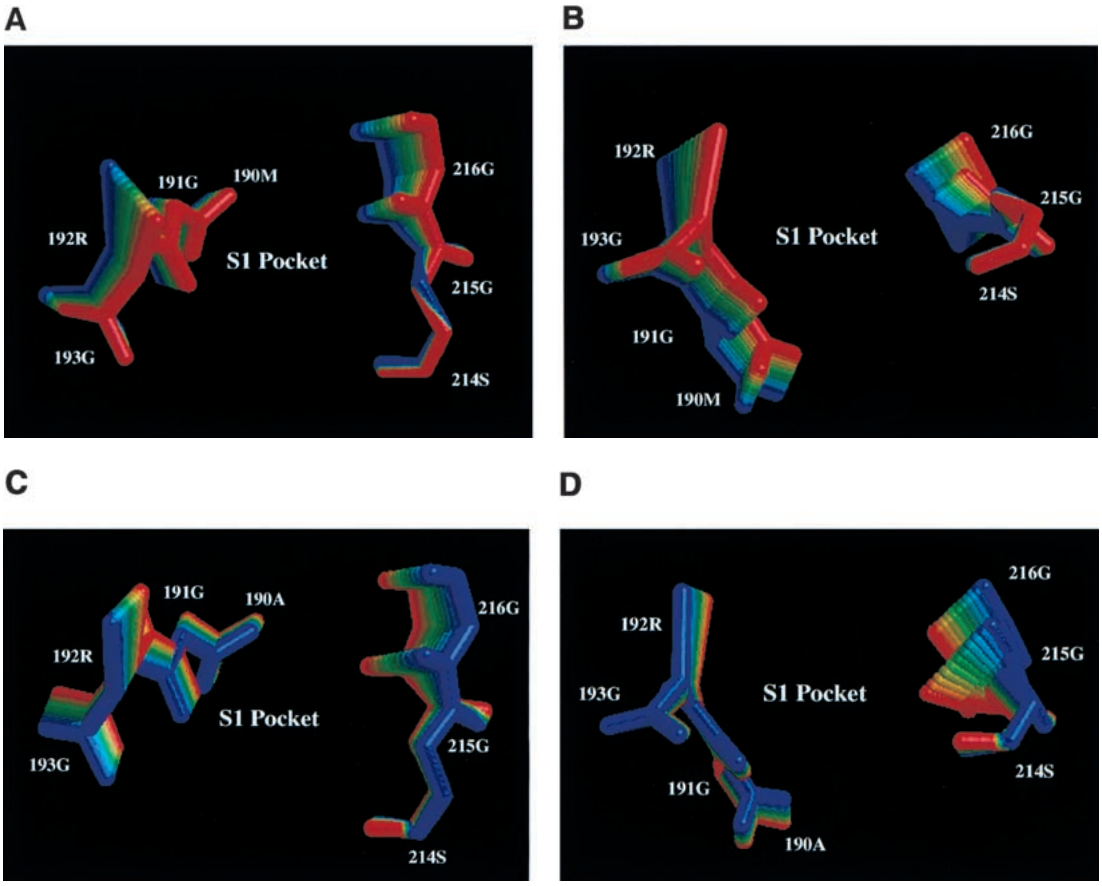

Fig. 7. Lowest frequency mode around the S1 binding pocket. The movements of the lowest mode are represented by a series of 11 structures with gradient colors from red to blue. The top and side views of the lowest binding mode of the wild-type αLP are shown in (*A*) and (*B*), respectively. The wild-type case shows the symmetric movement, so that the size of the pocket remains fixed. The top and side views of the lowest frequency mode of the M190A mutant are shown in (*C*) and (*D*), respectively. The movement of the mutant clearly shows the antisymmetric movement, indicating the expansion and contraction of the binding pocket.

Ser214 as the nodal point. This lowest mode shows significant similarity to the multiple conformer structures at cryogenic temperature (Fig. 2) (Rader and Agard 1997), which further indicates the accuracy and validity of our principal component analysis. Careful analysis of the angular orientation of maximal S1 pocket motions is also in good agreement with the previous NMA study (data not shown). In contrast, the two walls in the lowest frequency mode of M190A mutant show more antisymmetric movement relative to the wild type. The wall (Ala190-Gly193) vibrates up and down along the backbone strand between Ala190 and Gly 192 (Fig. 7C,D). The other wall (Gly214-Gly216) shows a gate-like movement from the open form and the closed form as an entrance for the S1 pocket, changing the size of the S1 pocket (Fig. 7B), which accounts for the broad specificity of the mutant M190A.

Wall distance at the S1 binding pocket

Direct observations of the distance between the two S1 pocket walls throughout the dynamics simulations also support our hypothesis. During the entire course of the 1 nsec MD simulations, the distances between the center of masses of the two walls for both the wild type and the mutant were calculated and are shown in Figure 8, A and B, respectively. For the wild type, the average wall distance was 7.90 Å and the standard deviation was 0.17 Å. The size of the S1 pocket is preserved with very small fluctuations between ~ 7.5 Å and ~ 8.5 Å. In the case of the M190A mutant, the wall distance shows much larger fluctuations, between ~ 6.5 Å and ~ 9.0 Å with larger deviations (0.46 Å). However, the average distance of the mutant (7.8 Å) remained almost identical to that of the wild type (7.9 Å). This suggests that the specificity mechanism of αLP is influenced by "dynamic motion" around the S1 binding pocket in addition to the static structural characteristics.

Discussion

Principal component analysis of αLP indicates that the dynamic motion of the two walls at the S1 binding pocket is

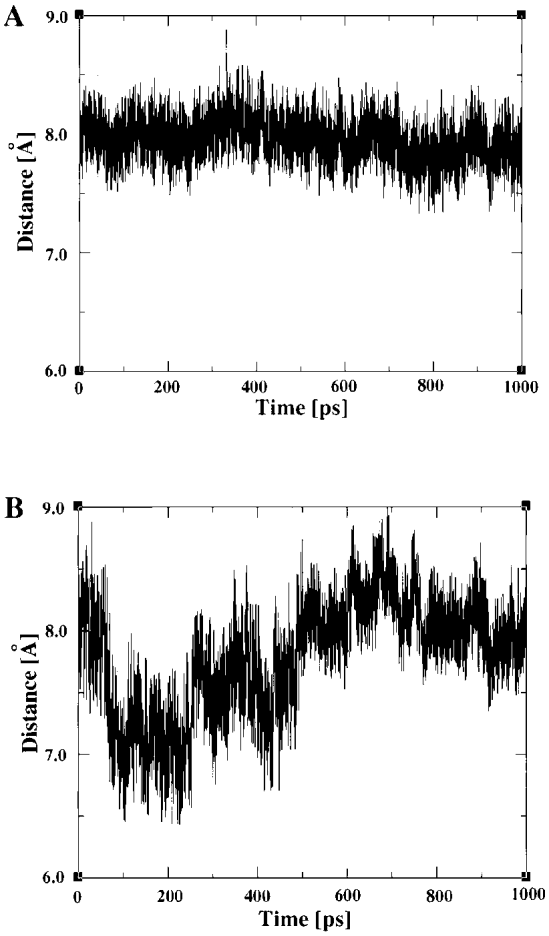

Fig. 8. Time course of the distance between the two walls at S1 pocket for the wild type (WT) (*A*) and the M190A mutant (*B*). The distributions are plotted every 0.05 ps for the 1 nsec simulations.

significantly altered when Met190 is mutated to Ala. In the wild type, the two walls move more symmetrically, preserving the size of the binding pocket. In contrast, in the mutant, more antisymmetric movement is observed so that the size of the binding pocket shows more expansion and contraction. However, the average sizes of the binding pocket over the 1 nsec simulations are very similar in both the wild-type and the mutant cases, as X-ray crystallographic studies previously suggested (Bone et al. 1989a,b, 1991). The results obtained by PCA methods appear to be consistent with the observed specificity profiles of the wild type and the mutant (Bone et al. 1989a) and also strongly support our hypothesis regarding the specificity mechanism (Rader and Agard 1997; Miller and Agard 1999). According to the PCA and the MD simulations, the two walls at the S1 pocket of the wild type move in unison, such that the interior dimensions of the pocket remain fixed, and therefore the wild-type specificity for small residues such as Ala is maintained. In contrast, the two walls of the mutant M190A enzyme show a greater degree of antisymmetric motion and a larger fluctuation in the size of the S1 pocket, allowing the various-sized P1 residues to access the active site and undergo catalysis. Thus, in addition to the static structure, dynamic motions around the binding pocket appear to play a significant role in determining the specificity profile for the enzyme.

We have applied a new method to study enzyme specificity, principal component analysis using molecular dynamics with a local solvent boundary condition. The combined method enables us to accurately elucidate the detailed motions of a critical region within a protein, with dramatically reduced computational costs. As anticipated from theory, the primary errors are in the high-frequency range owing to restraint effects. In general, this is not a problem, as the low-frequency collective motions of groups of atoms are most likely to be of functional significance. Such coupled motions are well described by the low-frequency modes. Thus, this combination of principal component analysis with a local solvent boundary condition provides a very promising method for studying functionally important local dynamic motions within large macromolecules.

Materials and methods

Solvent boundary condition and molecular dynamics (MD) simulation

All molecular dynamics (MD) simulations were carried out using X-PLOR (Brünger 1992) and run on a Silicon Graphics R10,000 workstation. The coordinates of the wild type and the M190A mutant of αLP were taken from 2ALP (Fujinaga et al. 1985) and 1GBJ (Mace and Agard 1995) in the Protein Data Bank (PDB). The OPLS force field (Jorgensen and Tirado-Rives 1998) with polar hydrogens was used for the proteins, and the TIP3P model (Jorgensen et al. 1983) was used for the solvent. The dielectric constant was set to 1.0, and the cut-off radius of nonbonded interactions was 12.0 Å. Two distinct simulation studies using spherical boundary methods (Brooks and Karplus 1983; Brünger et al. 1983, 1985) with two different radii, local boundary condition (LBC) and global boundary condition (GBC), were performed for both wild-type αLP and M190A.

The local solvent boundary condition (LBC) method in our system consists of four zones (Fig. 3A). First, inside the inner 15 Å radius "reaction" zone, standard molecular dynamics simulations were carried out. No restraints are imposed on any atoms in this region. Second, in the shell of a radius of 15–20 Å (inner buffer zone), the backbone atoms were weakly restrained to the X-ray structures by a harmonic energy term (10 kcal/Å^2). Third, in the outermost shell of 20–25 Å radius (outer buffer zone), all protein atoms were restrained to the initial X-ray structures (50 kcal/Å^2). Langevin dynamics were used for atoms in the buffer zone. The atoms outside this shell (reservoir zone) were fixed during the course of the simulation. All 1185 water molecules in the LBC method are free to move for all simulations.

The global spherical boundary condition (GBC) simulation was performed only for the wild-type αLP for the purpose of comparing the LBC results with those of the GBC method. In the global method, the entire αLP was immersed in a large spherical solvent droplet (radius 35 Å centered at the center of mass of αLP (Fig.

3B)). Because the entire protein migrates slightly during the simulation with the center of mass motion enabled, making it very difficult to directly compare modes obtained by the LBC and the GBC methods, the center of mass motion of α-lytic protease was turned off. In the GBC method, the total number of water molecules is 4202. No restraint functions were imposed on αLP.

The Verlet algorithm (Verlet 1967) was used to integrate the equation of motions. Temperature coupling (Berendsen et al. 1984) was used to keep the temperature at 300K with a coupling constant of 10 ps^{-1}. With the SHAKE algorithm (Ryckaert et al. 1977), the geometry of each water molecule was kept rigid and bond lengths within the proteins were kept constant. The time step was 2 fs. In the equilibration, the entire system was first thermalized gradually from 10K to 300K for 100 ps with the protein being restrained to the X-ray structure (Fujinaga et al. 1985). After the initial solvent equilibration, protein restraints were removed and thermalized again from 10K to 300K for 100 ps. The equilibrated structures were used as the initial model for the rest of the simulations. The total simulation time for both the local and the global boundary methods was performed up to 1 nsec. The 20,000 coordinates were stored at every 50 fs for principal component analysis (PCA). The CPU times were 156 h for the LBC method and 884 h for the GBC method.

Principal component analysis (PCA)

To compare the current MD results with the previous NMA results (Miller and Agard 1999), we need to convert the time course results of the MD simulation to frequency course results. This becomes possible by using principal component analysis (PCA) (Levy et al. 1984; Teeter and Case 1990; Kitao and Go 1999). Although there are two distinct ways to calculate modes and frequencies from MD trajectories, we used 'quasi-harmonic analysis' (Karplus and Kushick 1981; Levy et al. 1984; Teeter and Case 1990). In this method, the effective modes can be computed from the second moment matrix obtained from the MD trajectory through the following relationship between the force matrix and the variance-covariance matrix.

$$F = k_b T \sigma^{-1} \tag{1}$$

where F is a matrix of second derivatives of the potential energy, k_b is the Boltzmann constant, T is the absolute temperature, and σ_{ij} is the variance-covariance matrix given by

$$\sigma_{ij} = \langle (x_i - \langle x_i \rangle)(x_j - \langle x_j \rangle) \rangle \tag{2}$$

As shown in the above equation, the force matrix, F, can be obtained by the inverse of the fluctuation matrix found in the MD simulations. As in the case of NMA, diagonalization of the force matrix F can yield the eigenvectors corresponding to the frequency of each mode. In the case of the LBC method, only atoms in the "reaction zone" are included in the PCA. Therefore, the number of atoms in the LBC method was 628 in both the wild type and the mutant. The matrix size was 1884 by 1884. In contrast, in the GBC method, only backbone atoms of the entire αLP are included for the PCA, to reduce the size of the matrix. The number of backbone atoms in αLP is 729, and therefore the matrix size was 2376 by 2376. To achieve a direct comparison between the LBC and GBC methods, PCA was performed using backbone atoms in the "reaction zone" of the LBC method. The number of backbone atoms in the reaction zone in the LBC method was 279, and the matrix was 837 by 837. Because the systems in our studies are restrained to a specific space given by the X-ray coordinates and the center of mass motion was turned off, the PCA frequency spectrum consists of 3N nonzero frequencies rather than 3N-6 (Hodel et al. 1995).

Projection (P_j) test for comparing results from the two methods

To compare the similarity of the two normal mode vector sets, u_j from MD with the LBC, and w_i from MD with the GBC method, we adopted the projection test (P_j-test) (Kitao et al. 1991, 1994; Hayward et al. 1994). Each averaged conformation from the two methods is brought into its best fit position using the backbone atoms in the 'reaction zone' of the LBC method. Then, all of the vector sets u_j and w_i are reoriented and mapped onto the best fit positions. To quantify the similarity between the two vector sets obtained by two distinct methods, their inner product is calculated as shown below.

$$C_{ij} = w_i \cdot u_j \tag{3}$$

However, since the total number of vectors, w_i, is not equal to that of u_j, it is necessary to select the same number of renormalized vector sets of w_i and u_j to satisfy the normality condition. Here, the vector set corresponding to the backbone atoms matched to the reaction zone of the LBC method is selected as the new normalized w_i from the vectors of the GBC method. Then, the renormalized w_i is now compared with u_j by taking the inner product as shown above (Eq. 3). If the two vectors are identical, the inner product, C_{ij}, must be unity due to the normality. Since the space defined by u_j given by the LBC method lies within the space span by w_i given by the GBC method, any w_i should be expressed as a linear combination of the u_j vector sets. Therefore, the projection value,

$$P_j = \sum_{i=1}^{3N_{local}} C_{ij}^{\,2} \tag{4}$$

where P_j is the sum over the projections of the w_i vectors onto a single u_j vector. This P_j value indicates to what extent the vector u_j given by the GBC method is represented in the w_i vector space given by the LBC method. The closer the P_j value is to 1, the more completely the vectors obtained by the LBC methods can express the vectors obtained by the GBC methods.

Dot product analysis

In order to compare the PCA results and the NMA results (Miller and Agard 1999), we performed the dot product analysis described by Miller and Agard. Briefly, to determine the extent of symmetric and antisymmetric movement of the two walls at the binding pocket of αLP, we selected a group of 25 "specificity atoms" around the two walls of the S1 pocket (the Met190-Cly193 strand and the Ser214-Gly216 strand). Each of the 25 specificity atoms was then assigned a vector with a direction defined by a third line that intersects and is perpendicular to the two parallel walls. The length of all vectors was chosen to be 1.0 Å. In the symmetric case, the direction vectors of one wall are in the same direction of those of the other wall. In the antisymmetric case, the direction vectors of one wall head in the direction opposite to that of the other wall. The dot product is then calculated between those 25 difference vectors and 25 eigenvectors of each mode obtained by PCA. For

each mode, the sum of the 25 dot products was squared. The cumulative dot products are shown in Figure 6, A and C.

Acknowledgments

This work was supported by funds from the Howard Hughes Medical Institute to N.O. and D.A.A. We thank Professor Nobuhiro Go for helpful comments on the PCA with the local solvent boundary condition. We also thank Seth Harris, Luke Rice, Mark Wilson, Cynthia Fuhrmann, and members of Agard group for helpful discussions.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

Appendix

Theory

When macromolecules are very large or the local movement of a part of a protein is of interest, standard normal mode analysis (NMA) or principal component analysis (PCA) of the whole molecule would be time-consuming and computationally very intensive. Theoretically, it is possible to reduce the computational time and the matrix of second derivatives of energy, F, by MD with a localized solvent boundary condition. The original concept of the matrix reduction was developed as the static reduction method and later applied to mechanical engineering (Ookuma and Nagamatsu 1984).

First, we apply the theory of static reduction to our use of a solvent boundary condition. The dynamic movement of a macromolecule with N atoms is approximated as harmonic motions in terms of displacement vectors {x} as shown in the following equation:

$$[M]\{\ddot{x}\} + [F]\{x\} = \{0\} \tag{5}$$

where [*M*] is the mass matrix of the macromolecule, and [*F*] is the force matrix whose elements are given by the second derivatives of the potential energy of the macromolecule. In MD simulation with solvent boundary conditions, the system is usually divided into three regions: reaction region, buffer region, and reservoir region. In the reaction region, where the region of interest is located, standard MD simulation is carried out. In the buffer region, the atoms are positionally restrained because they are assumed to remain very close to the X-ray structure and prevent the MD trajectory from deviating too far from the X-ray structure. The atoms in the reservoir region are then treated as rigid since the interaction between reaction atoms and reservoir atoms is very weak. Since the displacements of rigid atoms and their second derivatives of the displacement vectors with respect to time are zero, the descriptions are eliminated from equation (5). Accordingly, the equation of motion of the system with no external forces can be partitioned in the following form:

$$\begin{pmatrix}[M_{rr}]\ [M_{rb}]\\ [M_{br}]\ [M_{bb}]\end{pmatrix}\begin{Bmatrix}\{\ddot{x}_r\}\\ \{\ddot{x}_b\}\end{Bmatrix} + \begin{pmatrix}[F_{rr}]\ [F_{rb}]\\ [F_{br}]\ [F_{bb}]\end{pmatrix}\begin{Bmatrix}\{x_r\}\\ \{x_b\}\end{Bmatrix} = \begin{Bmatrix}\{0\}\\ \{0\}\end{Bmatrix} \tag{6}$$

where $M_{\alpha\beta}$ and $F_{\alpha\beta}$ (α, β = r, b) are submatrices of the mass [*M*] and the force [*K*] matrices of the protein. The subscripts r and b indicate the reaction region and the buffer region, respectively.

According to the static reduction theory, only the masses in the system without restraint functions are to be retained in the analysis. Here, we want to describe only the movement of atoms in the reaction region, therefore set $M_{rb} = M_{br} = M_{bb} = 0$. Then equation (6) can be simplified as follows.

$$[M_{rr}]\{\ddot{x}_r\} + [[F_{rr}] - [F_{rb}][F_{bb}]^{-1}[F_{bf}]]\{x_r\} = \{0\} \tag{7}$$

This shows that the equation of harmonic motion of the whole system with boundary conditions can be expressed in terms of the degree of freedom of the atoms in the reaction region. It is more visible when Eq. (7) is converted to Eq. 8.

$$[\mathbf{M}]\{\ddot{x}_r\} + [\mathbf{F}]\{x_r\} = \{0\} \tag{8}$$

where $[\mathbf{M}] = [M_{rr}]$, $[\mathbf{F}] = [[F_{rr}] - [F_{rb}][F_{bb}]^{-1}[F_{br}]]$.

The important property of the above equation is that normal mode analysis of the reaction region using MD simulation with local boundary conditions, whose degrees of freedom are reduced, is a reasonable approximation to that describing the whole system if [**F**] can be estimated accurately. In principle, [**F**] can be obtained by directly calculating the second derivatives of energy including the restraining functions. However, since explicit solvents are included in our study, it is quite difficult to estimate [**F**] accurately. To circumvent this problem, we used the inverse relationship between the force matrix **F** and the variance-covariance matrix, σ,

$$F_{ij} = \frac{\partial^2 E'}{\partial x_i \partial x_j} = k_b T \sigma_{ij}^{-1} \tag{9}$$

where E′ is the approximate potential energy function whose matrix elements can be derived from a molecular dynamics trajectory, k_b is the Boltzmann constant, and T is the absolute temperature (Karplus and Kushick 1981). F_{ij} is an element of the second derivative matrix of the system, and σ_{ij} is an element of the second moment matrix, which is defined as:

$$\sigma_{ij} = \langle (x_i - \langle x_i \rangle)(x_j - \langle x_j \rangle) \rangle \quad (10)$$

Where < > is the ensemble average over the sampled structures during the period of simulation, and x_i is the mass-weighted atomic coordinates.

To obtain the modes and the corresponding frequencies, there are two distinct paths. One is the diagonalization of the mass-weighted Hessian matrix H, whose diagonal elements are the squared angular frequencies as shown below:

$$H = M^{-1/2} F M^{-1/2} \quad (11)$$

$$Q^{-1} H Q = diag\{\omega_1^2, \omega_2^2, \ldots, \omega_{3N}^2\} \quad (12)$$

where Q is the transformation matrix whose columns consist of eigenvectors (normal modes). The other way is to directly diagonalize the second moment matrix, σ. By a process similar to that shown above, σ can be transformed to the diagonalized matrix, yielding a set of eigenvalues and eigenvectors. The eigenvalue represents the mean square fluctuation along the axis of collective coordinates in the conformational space given by the eigenvector:

$$T^{-1} \sigma T = diag\{<\Delta\sigma_1^2>, <\Delta\sigma_2^2>, \ldots, <\Delta\sigma_{3N}^2>\} \quad (13)$$

where T is a transform matrix to diagonalize the second moment matrix, σ.

The two methods described above are connected through the classical relationship:

$$\omega_i = \sqrt{\frac{k_B T}{\langle \Delta\sigma_i^2 \rangle}} \quad (14)$$

where ω_i and $\Delta\sigma_i$ are the ith normal mode angular frequency and corresponding variable, respectively.

Here, it is worthwhile to consider the errors of the PCA method using MD under the solvent boundary condition. First consider the second moment matrix, including both the reaction region and the buffer region. In the system, σ_{ij} can be expressed as:

$$\sigma_{ij} = \langle (x_i - \langle x_i \rangle)(x_j - \langle x_j \rangle) \rangle \; i,j \in \textit{reaction or buffer} \quad (15)$$

Then, diagonalize the matrix, yielding

$$T^{-1} \sigma T = diag\{\Delta\sigma_1^2>, <\Delta\sigma_2^2>, \ldots <\Delta\sigma_k^2>, \ldots, <\Delta\sigma_{3N}^2>\}$$
$$\textit{when } k \in \textit{reaction}, <\Delta\sigma_r^2> = \textit{mean square deviation of no restrained atoms when } k \in \textit{buffer}, <\Delta\sigma_b^2> \cong 0 = \textit{mean square deviation of restrained atoms} \quad (16)$$

Although previously we assumed $m_i = m_j = 0$ when i and j are restrained atoms in the buffer region, in reality m_i and m_j are not zero. Therefore, the element of second moments σ of atoms in the buffer region cannot be ignored, though they are very small. The eigenvalues of σ, mean squared deviations, are nearly zero for the atoms. Since there is an inverse relationship, Eq. 14, between the angular frequencies and their mean squared deviation, the frequencies of restrained atoms will be shown as high-frequency modes, compared to those of unrestrained atoms in the reaction zone.

$$\omega_r << \omega_b \quad (17)$$

where ω_r is the frequency of reaction atoms, and ω_b is the frequency of buffer atoms. The frequency of the buffer atoms is dependent on the strength of restraining function and the detailed procedure of a simulation.

In the static reduction model, we completely eliminate the degrees of freedom of the restrained atoms. Therefore, the error due to the restrained atoms in the solvent boundary condition should be seen mainly on the high-frequency modes but not on the low-frequency modes. However, it is very difficult in practice to estimate the error caused by the restraints in using the P_j test since there are other possible sources of the errors. First, the convergence of variance-covariance matrix is sensitive to the simulation length and the number of ensemble structures (Balsera et al. 1996). Second, the average structures using two different MD methods showed deviations with respect to each other (Fig. 4), which makes it difficult to perform the P_j test for the precise comparison of two methods. Therefore, we used the P_j test for a qualitative comparison rather than a quantitative comparison of the two distinct PCA methods, using a global solvent boundary condition and a local solvent boundary condition.

References

Balsera, M.A., Wriggers, W., Oono, Y., and Schulten, K. 1996. Principal component analysis and long time protein dynamics. *J. Phys. Chem.* **100:** 2567–2572.

Berendsen, H.J.C., Postma, J.P.M., van Gunsteren, W.F., Dinola, A., and Haak, J.R. 1984. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **81:** 3684–3690.

Bone, R., Frank, D., Kettner, C.A., and Agard, D.A. 1989b. Structural basis for broad specificity in α-lytic protease complexes with analogues of reaction intermediates. *Biochemistry* **28:** 7600–7609.

Bone, R., Fujishige, A., Kettner, C.A., and Agard, D.A. 1991. Structural basis for broad specificity in α-lytic protease mutants. *Biochemistry* **30:** 10388–10398.

Bone, R., Shenvi, A.B., Kettner, C.A., and Agard, D.A. 1987. Serine protease mechanism: Structure of an inhibitory complex of α-lytic protease and a tightly bound peptide boromic acid. *Biochemistry* **26:** 7609–7614.

Bone, R., Silen, J.L., and Agard, D.A. 1989a. Structural plasticity broadens the specificity of an engineered protease. *Nature* **339:** 191–195.

Brayer, G.D., Delbaere, L.T., and James, M.N.G. 1979. Molecular structure of the α-lytic protease from Myxobacter 495 at 2.8 Å resolution. *J. Mol. Biol.* **131:** 743–775.

Brooks III, C.L. and Karplus, M. 1983. Deformable stochastic boundaries in molecular dynamics. *J. Chem. Phys.* **79:** 6312–6325.

Brünger, A.T. 1992. *X-PLOR Version 3.1. A system for X-ray crystallography and NMR.* Yale University Press, New Haven and London.

Brünger, A.T, Brooks III, C.L., and Karplus, M. 1983. Stochastic boundary conditions for molecular dynamics simulations of ST2 water. *Chem. Phys. Lett.* **105:** 495–500.

Brünger, A.T., Brooks III, C.L., and Karplus, M. 1985. Active site dynamics of ribonuclease. *Proc. Natl. Acad. Sci. U S A* **82:** 8458–8462.

Davis, J.H. and Agard, D.A. 1998. Relationship between enzyme specificity and the backbone dynamics of free and inhibited α-lytic protease. *Biochemistry* **37:** 7696–7707.

de Groot, B.L., Hayward, S., van Aalten, D.M.F., Amadei, A., and Berendsen, H.J.C. 1998. Domain motions in bacteriophage T4 lysozyme: A comparison between molecular dynamics and crystallographic data. *Proteins* **31:** 116–127.

Ding, X.C., Rasmussen, B.F., Petsko, G.A., and Ringe, D. 1994. Direct structural observation of an acyl-enzyme intermediate in the hydrolysis of an ester substrate by elastase. *Biochemistry* **33:** 9285–9293.

Farber, G.K. 1997. Light! Camera! Action! *Curr. Biol.* **7:** R352–R354.

Fastrez, J. and Fersht, A.R. 1973. Demonstration of the acyl-enzyme mechanism for the hydrolysis of peptides and anilides by chymotrypsin. *Biochemistry* **12:** 2025–2034.

Fujinaga, M., Delbaere, L.T.J., Brayer, G.D., and James, M.N.G. 1985. Refined structure of α-lytic protease at 1.7 Å resolution analysis of hydrogen bonding and solvent structure. *J. Mol. Biol.* **183:** 479–502.

Hunkapiller, M.W., Forgac, M.D., and Richards, J.H. 1976. Mechanism of ation of serine proteases: Tetrahedral intermediate and concerted proton transfer. *Biochemistry* **15:** 5581–5588.

Hayward, S., Kitao, A., Hirata, F., and Go, N. 1993. Effect of solvent on collective motions in globular protein. *J. Mol. Biol.* **234:** 1207–1217.

Hayward, S., Kitao, A., and Go, N. 1994. Harmonic and anharmonic aspects in the dynamics of BPTI: A normal mode analysis and principal component analysis. *Protein Sci.* **3:** 936–943.

Hayward, S., Kitao, A., and Go, N. 1995. Harmonicity and anharmonicity in protein dynamics: A normal mode analysis and principal component analysis. *Proteins* **23:** 177–186.

Hodel, A., Rice, L.M., Simonson, T., Fox, R.O., and Brünger, A.T. 1995. Proline cis-trans isomerization in staphylococcal nuclease: Multi-substrate free energy perturbation calculations. *Protein Sci.* **4:** 636–654.

Ichiye, T. and Karplus, M. 1991. Collective motions in proteins: A covarience analysis of atomic fluctuations in molecular dynamics and normal mode simulations. *Proteins* **11:** 205–217.

Johnson, L.N., Stura, E.A., Wilson, K.S., Sansom, M.S.P., and Weber, I.T. 1979. Nucleotide binding to clycogen phosphorylase b in the crystal. *J. Mol. Biol.* **134:** 639–653.

Jorgensen, W.L., Chandrasekhar, J., and Madura, J.D. 1983. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79:** 926–935.

Jorgensen, W.L. and Tirado-Rives, J. 1988. The OPLS potential functions for proteins. Energy minimizations for crystals of cyclic peptides and crambin. *J. Am. Chem. Soc.* **110:** 1657–1666.

Karplus, M. and Ichiye, T. 1996. Comments on a "fluctuation and cross correlation analysis of protein motions observed in nanosecond molecular dynamics simulations". *J. Mol. Biol.* **263:** 120–122.

Karplus, M. and Kushick, J.N. 1981. Method for estimating the configurational entropy of macromolecules. *Macromolecules* **14:** 325–332.

Karplus, M. and Petsko, G.A. 1990. Molecular dynamics simulations in biology. *Nature* **347:** 631–639.

Kitao, A. and Go, N. 1991. Conformational dynamics of polypeptides and proteins in the dihedral angle space and in the Cartesian coordinate space: Normal mode analysis of deca-alanine. *Jour. Comp. Chem.* **12:** 359–368.

Kitao, A. and Go, N. 1999. Investigating protein dynamics in collective coordinate space. *Curr. Opin. Struc. Biol.* **9:** 164–169.

Kitao, A., Hayward, S., and Go, N. 1994. Comparison of normal mode analyses on a small globular protein in dihedral angle space and Cartesian coordinate space. 1994. *Biophys. Chem.* **52:** 107–114.

Kollman, P. 1993. Free energy calculations: Applications to chemical and biochemical phenomena. *Chem. Rev.* **93:** 2395–2417.

Levitt, M. 1983. Molecular dynamics of native protein II. Analysis and nature of motion. *J. Mol. Biol.* **168:** 621–657.

Levy, R.M., Karplus, M., Kushick, J., and Perahia, D. 1984. Evaluation of the configurational entropy for proteins: Application to molecular dynamics simulations of an α-helix. *Macromolecules* **17:** 1370–1374.

Lipari, G. and Szabo, A. 1982. Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity. *J. Am. Chem. Soc.* **82:** 4546–4559.

Mace, J.E. and Agard, D.A. 1995. Kinetic and structural characterization of mutation of glycine 216 in alpha-lytic protease: A new target for engineering substrate specificity. *J. Mol. Biol.* **254:** 720–736.

McCammon, J.A., Celin, B.R., and Karplus, M. 1977. Dynamics of folded proteins. *Nature* **267:** 585–590.

Miller, D.W. and Agard, D.A. 1999. Enzyme specificity under dynamic control: A normal mode analysis of α-lytic protease. *J. Mol. Biol.* **286:** 267–278.

Moffat, K. 1997. Laue diffraction. In *Methods in enzymology. Macromolecular crystallography Part B* (eds. C.W. Carter et al.), **277:** pp. 433–447. Academic Press, New York.

Ookuma, M. and Nagamatsu, A. 1984. Analysis of vibration by component mode synthesis method. (Part 4. Natural frequency and natural mode (II)). *Japan. Soc. Mech. Eng.* **27:** 529–533.

Pauling, L. 1948. Nature of forces between large molecules of biological interest. *Nature* **161:** 707–709.

Rader, S.D. and Agard, D.A. 1997. Conformational substates in enzyme mechanism: The 120K structure of α-lytic protease at 1.5 Å resolution. *Biochemistry* **6:** 1375–1386.

Ryckaert J-P., Ciccotii, G., and Berendsen, H.J.C. 1977. Numerical-integration of Cartesian equations of motion of a system with constraints — Molecular dynamics of N-alkanes. *J. Comput. Phys.* **23:** 327–341.

Steitz, T.A., Henderson, R., and Blow, D.M. 1969. Structure of crystalline alpha-chymotrypsin. 3. Crystallographic studies of substrates and inhibitors bound to the active site of alpha-chymotrypin. *J. Mol. Biol.* **46:** 337–348.

Teeter, M.M. and Case, D.A. 1990. Harmonic and quasiharmonic descriptions of crambin. *J. Phys. Chem.* **94:** 8091–8097.

van Gunsteren, W.F. and Karplus, M. 1982. Protein dynamics in solution and in a crystalline environment: A molecular dynamics study. *Biochemistry* **21:** 2259–2274.

Verlet, L. 1967. Computer experiments on classical fluids. I. Thermodynamic properties of Lennard–Jones molecules. *Phys. Rev.* **159:** 98–105.

Wüthrich, K. 1986. *NMR of proteins and nucleic acids.* Wiley, New York.

Wüthrich, K. 1995. NMR — This other method for protein and nucleic acid structure determination. *Acta Crystallogr.* **D51:** 249–270.