Protein Motions Represented in Moving Normal Mode Coordinates

Kei Moritsugu and Akinori Kidera*

Graduate School of Integrated Science, Yokohama City University, Tsurumi-ku, Yokohama 230-0045, Japan

Received: October 18, 2003; In Final Form: January 11, 2004

The molecular dynamics trajectory of a protein molecule is frequently characterized by the strongly anharmonic dynamics on the rugged potential surface. We herein propose a model, moving normal mode coordinates, to describe the anharmonic protein motions recorded in the trajectory, particularly those occurring in the low-frequency or large-amplitude normal mode space. A set of normal mode coordinates is defined at each time instant of the trajectory by principal component analysis (PCA) with a small time window. The time evolution of the normal mode coordinates defines the moving normal mode coordinates, the axes of which are translated and rotated time-dependently along the trajectory. Because the harmonic part of the dynamics is accounted for by the PCA in a quasi-harmonic manner, the motions of the coordinate system are expected to represent the anharmonic part of the protein dynamics. We applied this model to the analyses of molecular dynamics trajectories of a small protein, myoglobin. Translation of the origin of the coordinates was decomposed into diffusion and motions confined in a concave potential well. Significant parts of both types of motions occur in the large-amplitude normal mode space. Rotation of the large-amplitude normal mode space was characterized by fast relaxation completed within the time window of PCA but was confined through the entire trajectory in a small space spanned by a limited number of large-amplitude normal modes. The influences of temperature and the solvation condition are also discussed.

1. Introduction

The biological function of a protein molecule can be interpreted to be a physicochemical process of sequential switching motions occurring as a part of the highly anharmonic dynamics on the rugged potential surface.[1,2] Thus, understanding such anharmonic dynamics is essential for elucidating the expression and regulation of molecular functions.

In studies of the anharmonic dynamics of proteins, molecular dynamics (MD) simulation has played a key role in the investigation of complicated anharmonic dynamics relevant to various biological functions.[3] These MD studies mainly focused on the equilibrium properties, such as the potential of mean force, and discussed the anharmonicity mostly with respect to the anharmonic part of the equilibrium distributions, for example, essential dynamics[4,5] and JAM.[6,7] On the other hand, there are few studies on the time course of the anharmonic dynamics. As an attempt to study the time-domain behavior of the anharmonic dynamics of protein, we used the perturbation scheme.[8,9] MD simulations were started by adding a small excess kinetic energy to a specified normal mode as a perturbation, and the process of the energy transfer to other modes was observed. It was found that the response to the perturbation occurred as resonance or off-resonance mode-coupling depending on temperature.[8,9] However, the perturbation scheme does not work for the motions occurring in the low-frequency (or large-amplitude) normal modes, because of their strong anharmonicity. Because it has been argued that most biologically significant motions occur in the large-amplitude normal modes,[10−12] it is important to develop a time-domain model describing the anharmonic dynamics occurring in the low-frequency normal modes.

For this purpose, we first assume that, within a sufficiently short period of time on the trajectory, the system can be approximately considered to behave as a set of harmonic oscillators. Under this assumption, it is possible to define normal mode coordinates characterizing the short time dynamics at any instant of time on the trajectory. It is, however, expected that because of the influence of anharmonicity, any two sets of coordinate systems, each derived from different portions of the trajectory, are not identical to each other, but rather have different origins and orientations. On the basis of this idea, we introduce the following time-dependent Lagrangian $L(t)$

$$L(t) = \sum_{i=1}^{N} \frac{1}{2}[\dot{\mathbf{x}} \cdot \mathbf{e}_i(t)]^2 - \frac{1}{2}\omega_i(t)^2[(\mathbf{x} - \mathbf{x}_0(t)) \cdot \mathbf{e}_i(t)]^2 \quad (1)$$

where N is the number of degrees of freedom, $\mathbf{x}$ is an N-dimensional vector representing the coordinates, $\dot{\mathbf{x}}$ is the time derivative of $\mathbf{x}$, $\omega_i(t)$ and $\mathbf{e}_i(t)$ are the frequency and eigenvector, respectively, for mode i defined at time t, and $\mathbf{x}_0(t)$ is the origin of vibration at time t. The mode number i is assigned in ascending order of frequency. If these basis functions, $\omega_i(t)$, $\mathbf{e}_i(t)$, and $\mathbf{x}_0(t)$, are time independent, then eq 1 is simply a Lagrangian for a set of harmonic oscillators. In the system of a protein, however, strong anharmonicity makes the basis functions time dependent, and thus we refer to the model as *moving normal mode coordinates*. Figure 1 explains the concept of eq 1. The projection of an MD trajectory onto a fixed normal mode coordinate shows a complicated double-peaked distribution (Figure 1a), whereas the projection onto a moving normal mode coordinate shows a single-peaked Gaussian distribution (Figure 1b). The anharmonic part of the dynamics was absorbed in the motions of the coordinate system, as shown in the time-dependent behavior of $\omega_i(t)$ (Figure 1c). This, in principle, is analogous to the model proposed by Broomhead et al.,[13] who

* Corresponding author. Phone: +81-45-508-7231. Fax: +81-45-508-7367. E-mail: kidera@tsurumi.yokohama-cu.ac.jp.

10.1021/jp037155o CCC: $27.50 © 2004 American Chemical Society
Published on Web 03/03/2004

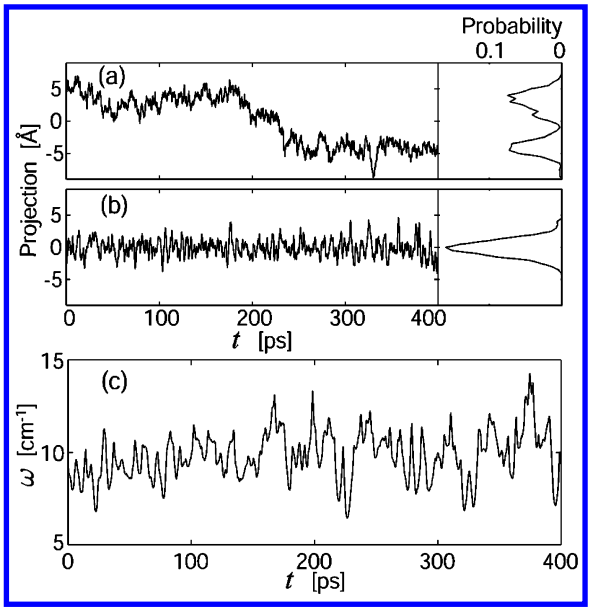

Figure 1. Projections of an MD trajectory at 300 K in water: (a) onto a fixed normal mode coordinate corresponding to the largest eigenvalue calculated at $t = 0$ and (b) onto the moving normal mode coordinate with $W = 10$ ps (see Methods for the definition). The probability distribution is also shown in the right panels. (c) The time course of frequency ω of the moving normal mode coordinate with $W = 10$ ps.

considered the trajectory as a manifold on phase space and tried to represent the trajectory by a series of locally defined manifolds.

As an application of the moving normal mode coordinates, we analyzed MD trajectories of a small protein, myoglobin, in several equilibrium conditions in order to evaluate the environmental influences on anharmonic dynamics in the protein. The details of the definition of the moving normal mode coordinates and the method to analyze the motions of the moving coordinates are given in section 2. The results and discussion are summarized in section 3.

2. Methods

2.1. Simulation System and Computations. In this study, we carried out four MD simulations of deoxymyoglobin (PDB code: 1mwd,[14] 2475 atoms), at two different temperatures, 30 and 300 K, and each in two different solvation conditions, in water and in vacuum. Each of the MD simulations was performed for 5 ns with a time step of 0.5 fs by the program PRESTO[15,16] with the AMBER parm94 all-atom parameters.[17] The simulations in water (3875 TIP3P waters[18]) were performed under the constant-temperature condition of the Nosé−Hoover algorithm[19,20] with the spherical solvent boundary potential[21] and the cell multipole method.[22] The vacuum simulations were performed under the microcanonical condition with the kinetic energy corresponding to the specified temperature and no cutoff operation.

The moving normal mode coordinates were calculated at each instant in time of the trajectories by principal component analysis (PCA) with a small time window of size W. The PCA is the eigenvalue problem for the variance−covariance matrix $\mathbf{C}$ of the distribution function, whose elements are defined by

$$c_{\alpha\beta} = \langle (x_\alpha - \langle x_\alpha \rangle)(x_\beta - \langle x_\beta \rangle) \rangle \qquad (2)$$

where x_α is the αth coordinate and $\langle \cdots \rangle$ is the average during W. Here, only the 153 Cα atoms in myoglobin were included in the calculation. Thus, the internal degrees of freedom, N, in eq 1 is 453 ($= 153 \times 3 - 6$). In order to guarantee rank ($\mathbf{C}$) = N, the size of the time window W was set larger than ~5 ps with the sampling interval of the coordinates being 5 fs. Calculations were repeated by shifting the time window every Δt ($= 0.1$ ps) through the entire MD trajectory. Assuming that the mass of all atoms is unity, the eigenvalue of $\mathbf{C}$, λ_i, is related to the frequency ω_i by kT/ω_i^2 with k and T being the Boltzmann factor and temperature, respectively, and the eigenvector is equivalent to $\mathbf{e}_i(t)$ in eq 1.

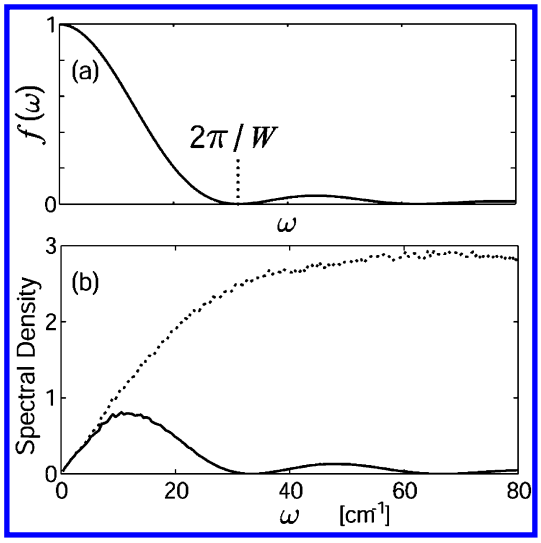

Figure 2. Implication of the moving average procedure in the analysis of an MD trajectory: (a) damping function of eq 4; (b) spectral densities of the origin of the moving normal mode coordinates for the MD trajectory at 300 K in water, without the averaging procedure (dotted curve) and with the averaging procedure with $W = 1$ ps (solid curve).

There is another way to calculate the normal mode coordinates: the instantaneous normal mode analysis (INMA),[23] the eigenvalue problem for the Hessian defined at an instant of the trajectory. There are two reasons why we chose PCA rather than INMA. First, the variance−covariance matrix of eq 2 does not have a mode with a negative eigenvalue, unlike the Hessian matrix at a finite temperature which has a number of modes with negative eigenvalues. The absence of the negative eigenvalue is important in order to evaluate the low-frequency normal modes correctly. Second, controlling the size of the time window W, we can filter out the high-frequency vibration from the trajectory. In the case of a harmonic vibration, $x(t) = A \sin \omega t$, the moving averaging procedure of the time window W gives

$$\langle x(\tau) \rangle = \frac{1}{W}\int_0^W \mathrm{d}t\, A \sin \omega(t+\tau) = \frac{\sin(\omega W/2)}{\omega W/2} A \times \sin(\omega\tau + \omega W/2) \qquad (3)$$

This indicates that the second moment of the vibration is damped by the function

$$f(\omega) = \frac{\sin^2(\omega W/2)}{(\omega W/2)^2} \qquad (4)$$

As shown in Figure 2, the damping function eliminates the vibrational motions with frequency higher than $2\pi/W$ and extracts the low-frequency motions. The moving average procedure is a simple filtering method without Fourier transformation. It may also be possible to interpret W as the average time to go over a potential barrier of height $E^\ddagger$, that is,

$$\frac{1}{W} \sim \exp(-E^{\ddagger}/kT) \quad \text{or} \quad E^{\ddagger} \sim kT \log W \tag{5}$$

2.2. Description of the Anharmonic Dynamics of the Moving Normal Mode Coordinates. Translation of the origin of the coordinate system is simply described by $\mathbf{x}_0(t)$ of eq 1, whose elements are $\langle x_\alpha \rangle$ of eq 2. Therefore, as far as the definition of translation is concerned, principal component analysis is not necessary, but the moving average is enough. On the other hand, the definition of rotation requires two sets of principal components at time t and $t + \Delta t$, and furthermore, the correspondence between the respective modes has to be determined. We regard mode i at t as corresponding to mode j at $t + \Delta t$, when their eigenvectors satisfy the following criterion

$$|\mathbf{e}_i(t) \cdot \mathbf{e}_j(t + \Delta t)| \geq 1/\sqrt{2} \tag{6}$$

The value of the criterion $1/\sqrt{2}$ comes from the normalization condition

$$\sum_j [\mathbf{e}_i(t) \cdot \mathbf{e}_j(t + \Delta t)]^2 = 1 \tag{7}$$

That is, eq 6 means that the contribution from mode j to the total variance exceeds 0.5. After defining the correspondence, rotation after the time interval τ is described quantitatively by the rotation angle of mode i

$$\cos \theta_i(\tau) = \mathbf{e}_i(t) \cdot \mathbf{e}_i(t + \tau) \tag{8}$$

However, there is a difficulty in the definition of correspondence, when two or more modes at t have similar values of eigenvalues of **C** or when those modes are in the resonance condition. Even a small variation in c_{ij} of eq 2 occurring within a small time interval Δt would cause a significant change in the coordinates of those modes. According to perturbation theory,[24] the vector $\mathbf{e}_i(t + \Delta t)$ is written for a small value of ϵ ($\sim \Delta t$)

$$\mathbf{e}_i(t + \Delta t) = \mathbf{e}_i(t) + \epsilon \sum_{j \neq i} \mathbf{e}_j(t) \frac{\Delta c_{ij}}{\lambda_i - \lambda_j} + O(\epsilon^2) \tag{9}$$

where $\epsilon \Delta c_{ij}$ is the variation in c_{ij}. When $\lambda_i \sim \lambda_j$, the second term on the right-hand side becomes large enough to change $\mathbf{e}_i$ significantly within Δt. In such a resonance case, correspondence of those modes cannot be defined, and rotation loses physical meaning. Because the density of states is extremely high in myoglobin, particularly in the high-frequency region, such resonance inevitably occurs. Therefore, rotation can be defined only for the several lowest-frequency modes having sufficiently large gaps between contiguous eigenvalues.

2.3. Canonical Covariance Analysis. In order to avoid the difficulty in the description of the rotation of a single mode, we analyzed the rotation of a set of normal modes. For this purpose, we used the canonical covariance analysis, which is a simple extension of the canonical correlation analysis.[25] Canonical covariance provides a quantitative measure expressing how similar or how different two sets of coordinate systems are in the scale of covariance. Here, the method is explained in terms of two sets of variance−covariance matrices, $\mathbf{C}(t)$ and $\mathbf{C}(t + \tau)$. First, a matrix **X** is defined by decomposing **C** as

$$\mathbf{C}(t) = \mathbf{X}(t)\mathbf{X}(t)^{\mathrm{t}} \tag{10}$$

where superscript t denotes the transpose. The matrix **X** represents a set of axes of coordinates with lengths corresponding to the square root of the variances and is written by the eigenvalue and eigenvector matrices of **C**, $\mathbf{\Lambda}$, and **E**, respectively, as $\mathbf{X} = \mathbf{E}\mathbf{\Lambda}^{1/2}$. The level of similarity is measured by the covariance between the following two sets of canonical variables, $\mathbf{a}^{\mathrm{t}}\mathbf{X}(t)$ and $\mathbf{b}^{\mathrm{t}}\mathbf{X}(t + \tau)$, where the coefficient vectors **a** and **b** are varied under the constraint that $\mathbf{a}^{\mathrm{t}}\mathbf{a} = 1$ and $\mathbf{b}^{\mathrm{t}}\mathbf{b} = 1$. A set of the covariances can be calculated as the eigenvalues of the cross-covariance matrix, $\mathbf{X}(t)^{\mathrm{t}}\mathbf{X}(t + \tau)$. Therefore, similarity can be evaluated by the trace of $\mathbf{X}(t)^{\mathrm{t}}\mathbf{X}(t + \tau)$. In the analysis, to describe the rotation of the coordinate system spanned by the large-amplitude normal modes, we used the partial matrix $\mathbf{C}_m$ and $\mathbf{X}_m$ ($\mathbf{C}_m = \mathbf{X}_m\mathbf{X}_m{}^{\mathrm{t}}$) containing only the m largest-amplitude modes, rather than the full matrix of **C** and **X**. That is, $\mathbf{X}_m = \mathbf{E}_m\mathbf{\Lambda}_m{}^{1/2}$, where $\mathbf{\Lambda}_m$ is the ($m \times m$) diagonal matrix containing the m largest eigenvalues and $\mathbf{E}_m$ is the corresponding ($N \times m$) eigenvectors matrix. Furthermore, the trace was normalized to give the normalized trace of canonical covariance R_{cc}

$$R_{\mathrm{cc}}(\tau) = \langle \mathrm{Tr}[\mathbf{X}_m(t)^{\mathrm{t}}\mathbf{X}_m(t + \tau)]^{1/2} / [\sum_{i=1}^{m} \lambda_i(t) \sum_{i=1}^{m} \lambda_i(t + \tau)]^{1/2} \rangle \tag{11}$$

where the average is taken in terms of time t along the whole trajectory. The difference from the conventional canonical correlation coefficient is that R_{cc} explicitly counts the effects of anisotropy in the distribution.

3. Results and Discussion

3.1. Translation. The translational motion of the origin of the moving normal mode coordinates was monitored by the autocorrelation function of the mean square displacements between t and $t + \tau$ averaged over the trajectory

$$\langle \Delta \mathbf{x}_0{}^2(\tau) \rangle = \langle [\mathbf{x}_0(t + \tau) - \mathbf{x}_0(t)]^2 \rangle \tag{12}$$

where the average is taken over time t along the whole trajectory. In parts a and b of Figure 3, the autocorrelation functions for various values of the window size W, calculated from the trajectories at 30 and 300 K, and both in water and in a vacuum, are plotted against τ.

As the data at 30 K show, the autocorrelation shows a sudden increase at a small value of τ and a slower linear increase at a longer time scale. The data at 300 K simply appear the same as the data at 30 K but stretched along the τ axis. Because the former can be ascribed to vibration and the latter to diffusion, we analyze these data by the model consisting of the contributions from vibration and diffusion. It is mentioned that the term "vibration" does not necessarily represent a real oscillation in this time scale but simply the motion in a concave potential well.

In this model, the autocorrelation function is described for mode i as

$$\langle \Delta x_i{}^2(\tau) \rangle = R_i{}^2(1 - \cos \omega_i \tau) + 2D_i\tau \tag{13}$$

where the first term represents vibration with vibrational amplitude R_i and frequency ω_i and the second term is for diffusion with diffusion coefficient D_i. Summation over all modes with the mode density $\rho(\omega_i)$ gives

$$\langle \Delta \mathbf{x}_0{}^2(\tau) \rangle = \sum_i \rho(\omega_i) \langle \Delta x_i{}^2(\tau) \rangle \sim R^2(1 - e^{-\tau/\tau_0}) + 2D\tau \tag{14}$$

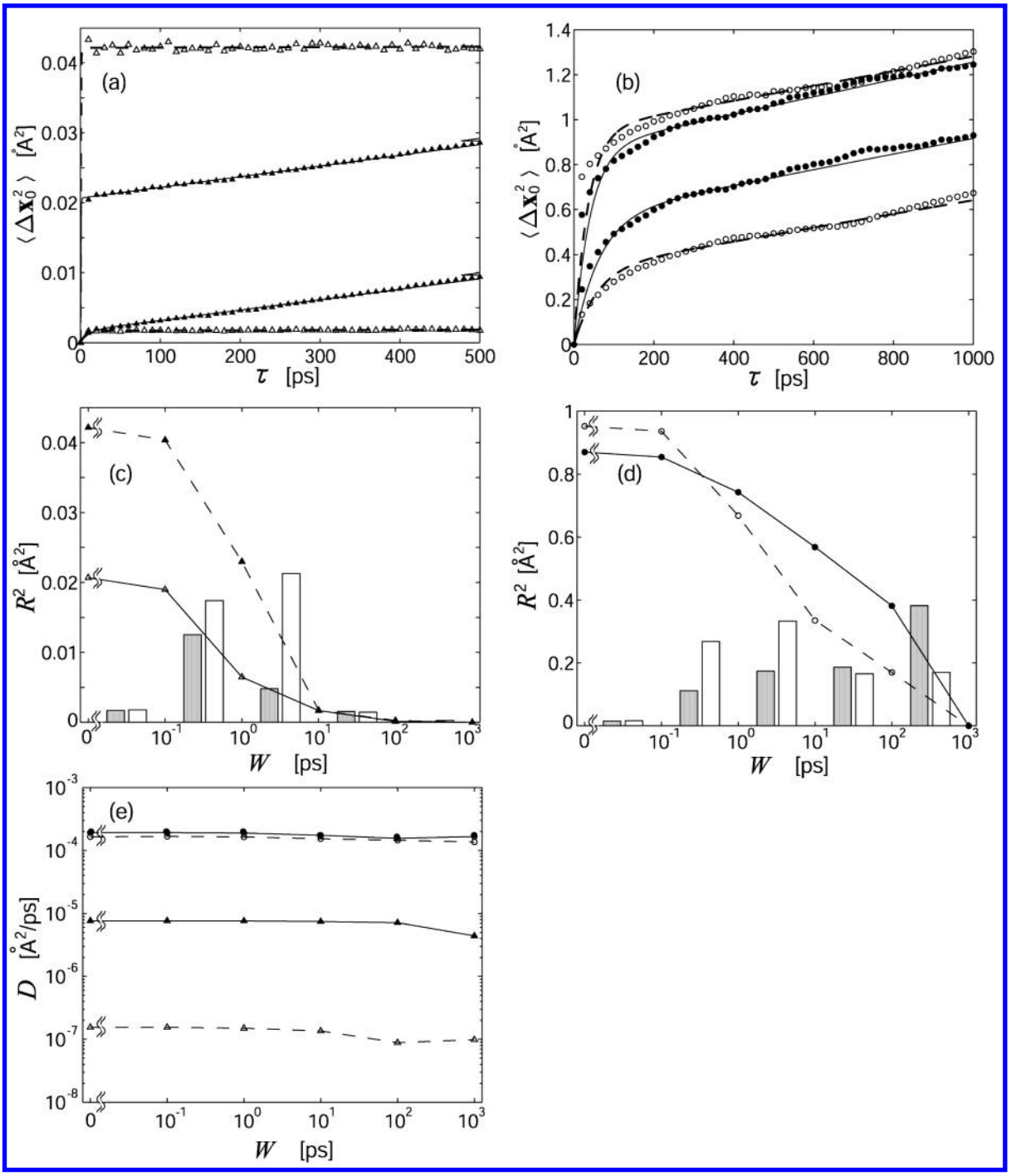

Figure 3. Mean square deviations of the origin of the moving normal mode coordinates, $\langle \Delta \mathbf{x}_0^2(\tau) \rangle$ (eq 12), calculated from (a) the trajectories at 30 K in water (filled triangles) and in vacuum (open triangles) and (b) the trajectories at 300 K in water (filled circles) and in vacuum (open circles). Each figure shows two sets of the data calculated with two different sizes of time windows, $W = 0$ (upper) and $W = 10$ ps (lower). The best-fit curves of the model function of eq 14 are shown: the solid curve for the data in water and the broken curve for the data in vacuum. The square of the vibrational amplitude R^2 and the diffusion coefficient D are plotted against W: (c) R^2 at 30 K, (d) R^2 at 300 K, and (e) D. The same symbols as those in (a) and (b) are used. In (c) and (d), the values of $\sigma^2(W_n)$ [$= R^2(W_{n-1}) - R^2(W_n)$] are also plotted in the form of a histogram. The scales of the y axes in (a−d) are those divided by the number of Cα atoms (= 153).

where

$$R^2 \equiv \sum_i \rho(\omega_i) R_i^2, \quad D \equiv \sum_i \rho(\omega_i) D_i,$$

$$\int \mathrm{d}\omega\ \rho(\omega) R(\omega)^2 \cos \omega\tau \sim e^{-\tau/\tau_0}$$

Here, the last equation is simply an approximation by an exponential function. The model of eq 14 could fit well to the values in Figure 3a and b with the adjustable parameters, R^2, D, and τ_0. Garcia and Hummer[26] calculated the autocorrelation function, $\langle \Delta x_i^2 \rangle$ (eq 12), of cytochrome *c* along some low-frequency normal modes in a fixed coordinate system and obtained a similar conclusion that the trajectory could be described in terms of diffusion. Their "subdiffusion" may correspond to the vibrational contributions defined in the present study.

The values of the squared amplitude of vibration R^2 and the diffusion constant D of eq 13 thus determined are plotted against

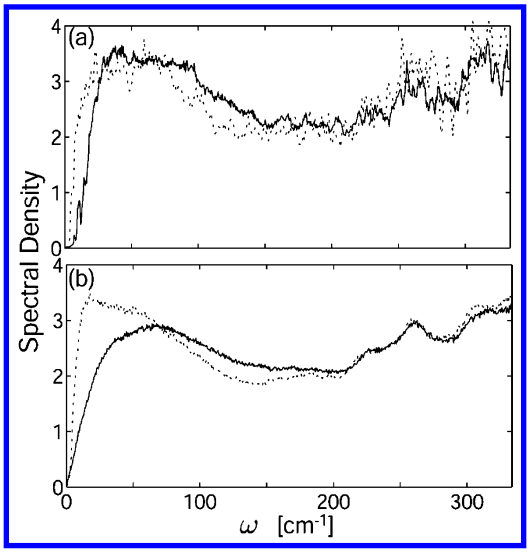

Figure 4. Spectral density of the origin of the moving normal mode coordinates calculated from the trajectories (a) at 30 K and (b) at 300 K. Each figure presents the spectra for the data in water (solid curve) and in vacuum (dotted curve).

the window size W in Figure 3c−e. Here, R^2 was found to monotonically decrease, approaching zero with increasing W, whereas D was almost constant irrespective of W. It is reasonable that D is independent of W, that is, the diffusion term is not affected by the moving average procedure. By definition, the diffusion term is determined from the asymptotic behavior of $\langle \Delta \mathbf{x}_0^2(\tau) \rangle$ at large τ or in the low-frequency limit, whereas the moving average procedure is a filter of the high-frequency contributions (eq 4).

In order to explain the behavior of R^2, we propose the following model for the dynamics of the origin of the coordinate system. The distribution of the vibrational motions, $p(x)$, has a hierarchical structure described by multiple levels of the convolution of Gaussian functions as

$$p(x) \sim \int p_1(x_1) \prod_{i=2}^{M} p_i(x_i - x_{i-1})\, \mathrm{d}x_{i-1}$$

$$\sim \exp\left[- \frac{x^2}{2\sum_{i=1}^{M} \sigma^2(W_i)} \right] \qquad (15)$$

where x is $|\mathbf{x}_0|$ after the diffusional contribution is subtracted, $x \equiv x_M$ and $p_i(x)$ is a Gaussian function with variance $\sigma^2(W_i)$ representing motions with the characteristic time range of W_i ($0 = W_1 < W_2 < \cdots < W_M$; here, we arbitrarily chose $M = 6$ with $W_i = 0$, 0.1, 1, 10, 100, and 1,000 ps), and thus $R^2(W_1) \sim \sum_{i=1}^{M} \sigma^2(W_i)$.

Because the moving average with the time window W_n ($1 < n < M$) filters out the high-frequency contributions having time scales shorter than W_n, $R^2(W_1)$ is reduced to $R^2(W_n)$ as $R^2(W_n) \sim \sum_{i=n+1}^{M} \sigma^2(W_i)$. Therefore, $\sigma^2(W)$ represents the size of the conformational space having characteristic time scale W. In Figure 3c and d, the values of $\sigma^2(W_n)[= R^2(W_{n-1}) - R^2(W_n)]$ are plotted in the form of a histogram. In this sense, the distribution of σ^2 may have some correlation with the spectral density in the spectral analysis[27,28] shown in Figure 4. However, it is noticed in Figure 4 that the spectral analysis cannot detect motions of which the frequency is less than several cm^{-1} or of which the characteristic time range is over ~10 ps. On the other hand, the present analysis of $\sigma^2(W)$ can follow the motions of the time range up to 1 ns.

Comparing the two distributions of $\sigma^2(W)$ given in Figure 3c and d, one in a vacuum and the other in water, we can see the solvation influence on protein anharmonic dynamics. According to the damped oscillator model for the solvation effects appearing in the spectral density,[29,30] we tried to interpret the distribution at 300 K (Figure 3d) by classifying it into four W regions. (i) $W \sim 0.1$ ps: Because the friction influence from the solvent scarcely affects the high-frequency oscillations in this time domain, almost no difference is observed. (ii) $1 < W < 10$ ps: Oscillations in these low frequencies are damped by the friction of water, and thus the σ^2 values in water are smaller than those in vacuum. Such behavior corresponds to the decrease in the spectral density at the low-frequency region in Figure 4. (iii) 100 ps $< W <$ 1 ns: Over this time scale, the damped oscillator model predicts overdamping [frequency < damping coefficient ~ 1−7 ps^{-1}].[29−31] However, the σ^2 value in water is comparable to or larger than that in a vacuum, and thus the damped oscillator model is not suitable to explain the behavior in this time range. (iv) $W >$ 1 ns: The thermal energy at 300 K seems to be insufficient to go over the potential barrier corresponding to $W >$ 1 ns. Instead, the system moves from one basin to another to contribute only to diffusion. In summary, solvation appeared to cause the shift of the population of motions from the low-W (1−10 ps) region to the high-W (~ 1 ns) region. In other words, this made the potential surface far more rugged.

At 30 K (Figure 3c), the overall feature of the solvation influence is the same as that at 300 K. However, because of the small kinetic energy at 30 K, the motions are predominantly confined in a single basin of the potential surface, and thus motions of a time scale > 100 ps were scarcely observed.

The response of the diffusion coefficient D to solvation is not necessarily straightforward (Figure 3e). At 30 K, upon solvation, the value of D was increased by almost 2 orders of magnitude. It is considered that the activation energy required for Brownian motion was reduced by the lubrication effects of water.[32] In contrast, we find a very small difference in D at 300 K [1.65×10^{-4} (in vacuum)/1.93×10^{-4} (in water) $\mathrm{Å}^2$/ps, ($W = 0$ ps)]. This difference might arise because the gain from the lubrication effects is masked by the large kinetic energy at 300 K.

The diffusive motions extracted here are highly stochastic motions on a flat potential surface, which is totally different from the vibrational motions occurring within a concave potential surface. According to the apparent flat energy landscape, the diffusive motions should be much more susceptible to external perturbations compared to the vibrational motions. Considering that the functional motion of a protein occurs as a response to an external perturbation such as ligand binding, we suppose that the functional motion, or the structural change upon ligand binding, should preferentially occur along the diffusive motions rather than along the vibrational motions.

Finally, we examine the correlation between the diffusive and vibrational motions. Figure 5 plots the square of the velocity for the diffusive motion projected onto each vibrational mode, v_i^2, against the eigenvalue of the vibrational mode, λ_i. The velocity of the diffusive motion was calculated from the trajectory averaged with $W = 1$ ns, which contains almost exclusively diffusion contributions, as shown in Figure 3c and d. The vibrational mode was defined by the principal component analysis of the matrix $\mathbf{C}$ (eq 2) with the smallest size of time window, or $W = 5$ ps. Figure 5 shows strong positive correlation between v_i^2 and λ_i, both at 30 and 300 K, which indicates that

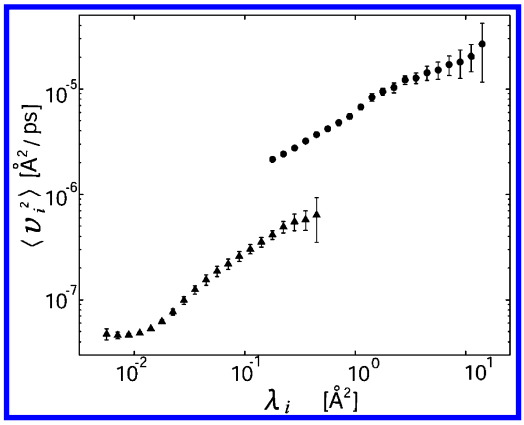

Figure 5. Mean square velocity of the origin of the moving normal mode coordinates for $W = 1$ ns projected onto the moving normal mode coordinate i defined with $W = 5$ ps, $\langle v_i^2 \rangle$, are plotted against the eigenvalue of the ith normal mode λ_i. For the sake of clarity, the data were averaged over the bin size of 0.1 along the abscissa, and the values of deviations within each bin (error bars) are also shown. Two sets of the data are presented, those at 300 K in water (filled circles) and those at 30 K in water (filled triangles).

diffusive motions occur preferentially along the direction of the vibrational modes with large amplitudes. This correlation may be relevant to the recent observations that the functional structural changes occur along the normal modes with larger amplitudes.[33−35]

3.2. Rotation. Rotation of the moving normal mode coordinates was first monitored by the average of the cosine of the rotation angle, $\langle \cos \theta_i(\tau) \rangle$ (eq 8), where the average was taken for time t along the entire trajectory. Parts a and b of Figure 6 show the decay curves of $\langle \cos \theta_i(\tau) \rangle$ at 30 and 300 K, respectively, each of which is produced from data for different solvation conditions, and for two sizes of time window, $W =$ 100 ps and $W = 1$ ns. The data shown in Figure 6 are the average values calculated for the two to three largest-amplitude modes, which have sufficiently large eigenvalues to satisfy the criterion $\lambda_i > \lambda_{max}/10$, where λ_{max} is the maximum eigenvalue found in the trajectory. Actually, all of the other smaller-amplitude modes with $\lambda_i < \lambda_{max}/10$ suffer from the problem of resonance mentioned in section 2.

In order to explain the behaviors of the decay curves in Figure 6, we adopted the following damped oscillator model[36]

$$\langle \cos \theta(\tau) \rangle = \exp(-\gamma_R \tau)\left[\cos(\tilde{\omega}_R \tau) + \frac{\gamma_R}{\tilde{\omega}_R} \sin(\tilde{\omega}_R \tau)\right] \quad (16)$$

where $\tilde{\omega}_R = (\omega_R^2 - \gamma_R^2)^{1/2}$ with the angular velocity ω_R and the damping coefficient γ_R. The model of eq 16 fit well the values in Figure 6a and b with the adjustable parameters $\omega_R$and γ_R. Parts c and d of Figure 6 show the values of $\omega_R$and γ_R thus obtained. All sets of the parameters ω_R and γ_R indicate near-critical damping, that is, the ratio $\gamma_R/\omega_R \simeq 0.8$, irrespective

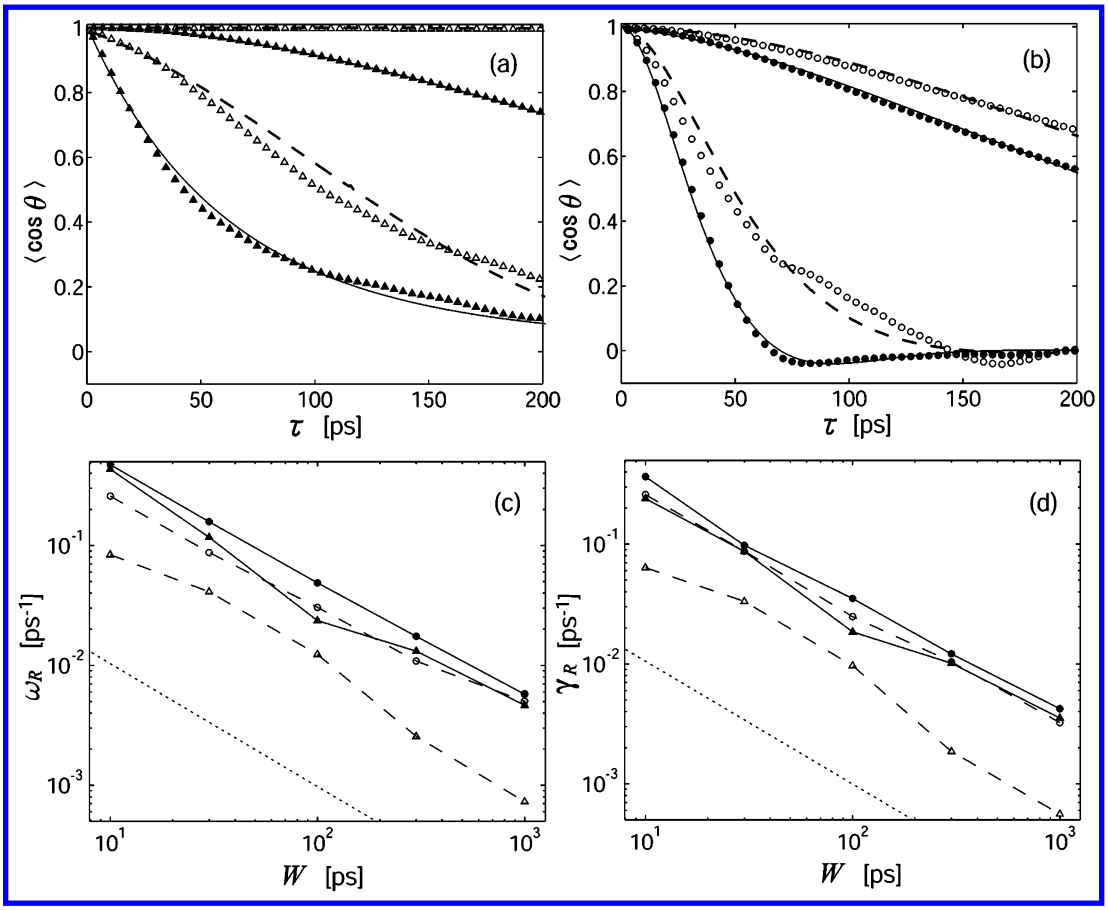

Figure 6. Cosine of the rotation angle of the moving normal mode coordinate, $\langle \cos \theta(\tau) \rangle$, is plotted as a function of time τ. These values were calculated as the average for the modes whose eigenvalues λ_i are large enough to satisfy the relation $\lambda_i > \lambda_{max}/10$, where λ_{max} is the maximum eigenvalue found in the trajectory. (a) Values for the trajectories at 30 K in water (filled triangles) and in vacuum (open triangles); (b) values for the trajectories at 300 K in water (filled circles) and in vacuum (open circles). Each figure contains the data with different time windows, $W = 100$ ps and $W = 1$ ns. The best-fit curves of the model function of eq 16 are also shown: the solid curve for the data in water and the broken curve for the data in vacuum. In (c) and (d), the angular velocity ω_R and the damping coefficient γ_R derived by fitting to the values in (a) and (b) are shown for the data at 30 K (c) and at 300 K (d). The same symbols as those in (a) and (b) were used, and ω_R (or γ_R) $= 1/W$ with an arbitrary intercept were drawn by dotted lines.

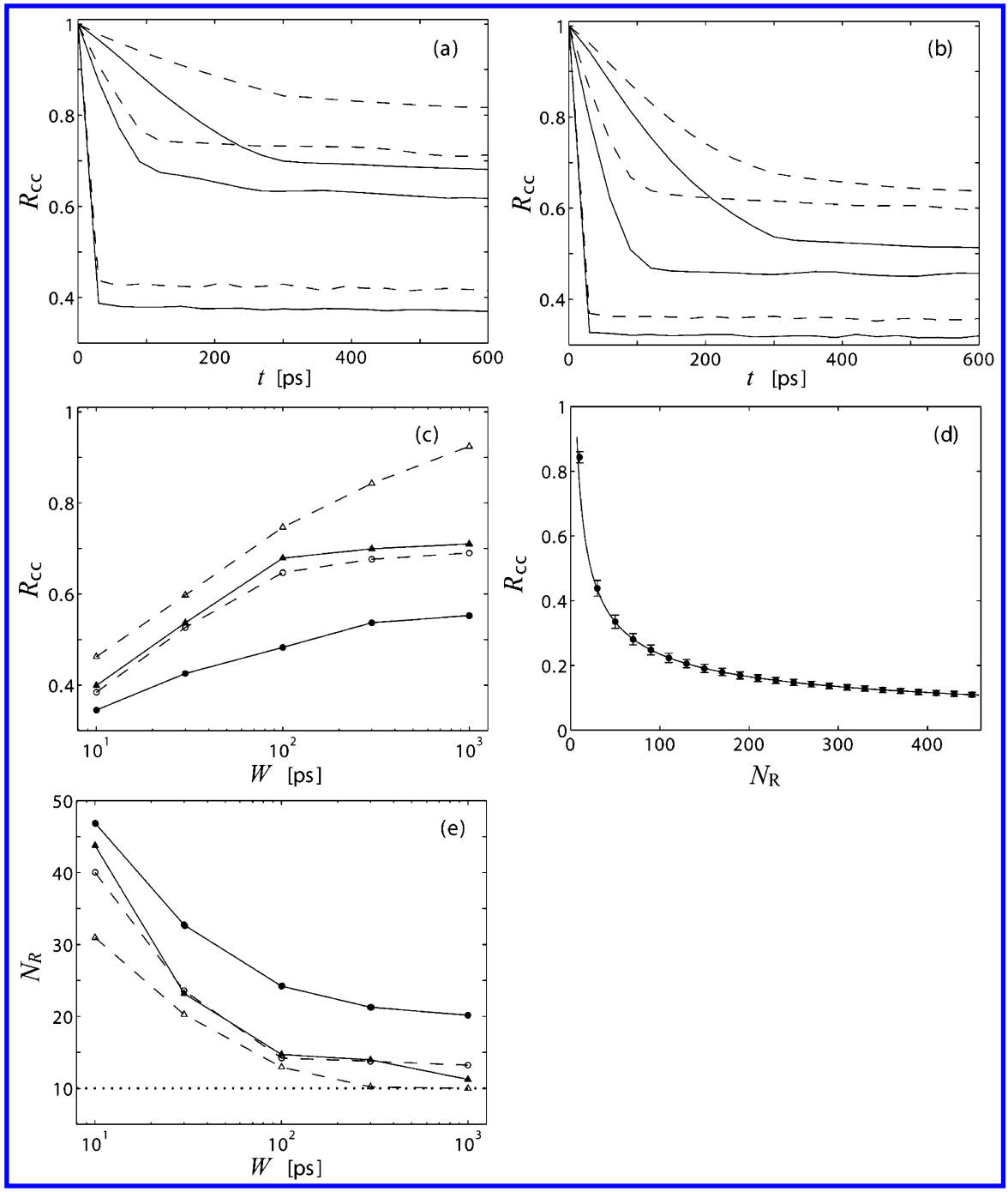

Figure 7. Normalized trace of the canonical covariance, $R_{cc}(\tau)$ (eq 11), calculated (a) from the trajectories at 30 K in water (solid curves) and in vacuum (dashed curves) and (b) from the trajectories at 300 K in water (solid curves) and in vacuum (dashed curves). Each figure contains three sets of data, which were calculated with time windows of three different sizes: $W = 300$ ps, $W = 100$ ps, and $W = 10$ ps, from top to bottom. (c) Converged values of R_{cc} or $R_{cc}(\tau > W)$ are plotted against W. Symbols are the same as those in Figure 3. (d) Correspondence between $R_{cc}(\tau > W)$ and the rotational degrees of freedom N_R derived from the model calculations (see text). The value of m in eq 11 was set to 10. The data presented here are the average of the values calculated from the trajectories at 300 K in water with $W = 100$ ps, together with the deviations (error bars). (e) Rotational degrees of freedom N_R of the configurational space are plotted against the time window W. The dotted line at $N_R = 10$ indicates the lowest limit of R_{cc}.

of temperature or solvation condition. It was further found that the W dependence of ω_R or $\gamma_R$is

$$\omega_R \text{ (or } \gamma_R) \sim 1/W \tag{17}$$

Because PCA filters out the vibrational motions with frequencies higher than $2\pi/W$, eq 17 means that the rotational velocity of the moving normal mode coordinates is determined entirely by the fastest motions recorded in the trajectory. In other words, the rotation is so quick that the same pattern of motion never appears after one cycle of vibration, or the protein continuously changes the pattern of motion by rotating the mode space. Thus, at this level of analysis, fast rotation does not seem to be compatible with the existence of the long-standing diffusional motions observed in the translational motions. We will address this point below using another type of analysis.

In the above analyses, we could investigate the rotational dynamics of only a limited number of large-amplitude modes because of the problem of resonance. In order to avoid the

difficulty in analyzing the rotation of a single mode, we tried to analyze the rotational dynamics of a subset of normal modes by using the canonical covariance analysis explained in section 2. Parts a and b of Figure 7 show the decay of the normalized trace of the canonical covariance $R_{cc}(\tau)$ (eq 11), which was calculated for the 10 largest-amplitude modes ($m = 10$) explaining as much as 60% of the total variance on average. These curves show the common features that R_{cc} decreases almost linearly up to $\tau = W$ and becomes almost constant beyond this point ($\tau > W$), irrespective of temperature or solvation condition. These features indicate that the relaxation of the rotation completes within the time window W, and no further decay occurs for $\tau > W$. Such fast rotation of the set of the mode axes is consistent with the rotation of a single mode axis found in Figure 6.

In Figure 7c, the average values of R_{cc} for $\tau > W$ are plotted as a function of W. Even though R_{cc} is almost constant for $\tau > W$, we took the average of R_{cc} for $\tau > W$. In order to explain the behavior in Figure 7c, we propose the following interpretation of the rotation of the coordinate system.

The partial matrix of variance−covariance $\mathbf{C}_m(t)$ is a random matrix in a space of a small number of degrees of freedom N_R ($\geq m$). The matrix at $t = t + W$, $\mathbf{C}_m(t + W)$, is another entirely new random matrix in the same space of N_R-dimension. Between t and $t + W$, $R_{cc}(\tau)$ linearly decreases with the decreasing number of the common coordinate sets counted both in $\mathbf{C}_m(t)$ and $\mathbf{C}_m(t + \tau)$. At $\tau = W$, $\mathbf{C}_m(t + W)$ does not contain any of the coordinate sets counted in $\mathbf{C}_m(t)$ and gives the convergent value of R_{cc}. The dimension N_R of the space in which rotation from $\mathbf{C}_m(t)$ to $\mathbf{C}_m(t + W)$ occurs determines the value of $R_{cc}(\tau > W)$.

On the basis of this idea, we obtained the relation of $R_{cc}(\tau > W)$ to $N_{R,}$ as shown in Figure 7d. The values of N_R were calculated as a function of $R_{cc}(\tau > W)$ in the following manner. The random matrix $\mathbf{C}_m(t)$ is given by

$$\mathbf{C}_m(t) = \mathbf{E}_m{}^t(t)\mathbf{\Lambda}_m\mathbf{E}_m(t) \quad \text{with} \quad \mathbf{E}_m(t) = \mathbf{E}_{N_R}\mathbf{R}(t) \quad (18)$$

Here, the m eigenvectors $\mathbf{E}_m$ are given as linear combinations of fixed N_R eigenvectors $\mathbf{E}_{N_R}$with a randomly generated coefficient $\mathbf{R}$ (an $N_R \times m$ matrix satisfying the normalization condition $\mathbf{R}^t\mathbf{R} = \mathbf{I}$ with $\mathbf{I}$ being the $m \times m$ identity matrix). The corresponding eigenvalues $\mathbf{\Lambda}_m$ were taken from the average values in the trajectory. In the same way as eq 18, the random matrix $\mathbf{C}_m(t + W)$ is given by

$$\mathbf{C}_m(t + W) = \mathbf{E}_m{}^t(t + W)\mathbf{\Lambda}_m\mathbf{E}_m(t + W) \quad \text{with}$$
$$\mathbf{E}_m(t + W) = \mathbf{E}_{N_R}\mathbf{R}(t + W) \quad (19)$$

where $\mathbf{\Lambda}_m$ and $\mathbf{E}_{N_R}$ are the same matrices as those in eq 18 and $\mathbf{R}(t)$ and $\mathbf{R}$ $(t + W)$ are different random matrices. Therefore, the numerator of $R_{cc}(W)$ in eq 11 is calculated by

$$\mathrm{Tr}[\mathbf{X}_m{}^t(t)\mathbf{X}_m{}^t(t+W)] = \mathrm{Tr}[\mathbf{\Lambda}_m{}^{t/2}\mathbf{R}(t)^t\mathbf{E}_{N_R}{}^t\mathbf{E}_{N_R}\mathbf{R}(t+W)\mathbf{\Lambda}_m{}^{1/2}]$$
$$= \mathrm{Tr}[\mathbf{\Lambda}_m\mathbf{R}(t)^t\mathbf{R}(t+W)] \quad (20)$$

After preparing 10^3 pairs of random matrices, the average and the standard deviation of R_{cc} was calculated by eq 11 for various values of N_R and was then plotted in Figure 7d.

By use of the results shown in Figure 7d, the R_{cc} values were converted into the rotational degree of freedom N_R and then plotted against W in Figure 7e. It is clearly seen in Figure 7e that the rotation of the space spanned by the 10 mode axes occurs in a small configurational space. For example, at 300 K in water, the rotation of the 10 largest-amplitude modes for $W = 1$ ns occurs only in the 20-dimensional space out of the 453 dimensions. Despite the fast rotation of the mode axes, because of the smallness of the space of the rotation, diffusional motion having a characteristic time scale as long as 1 ns can be observed in protein anharmonic dynamics.

Finally, the dependence of N_R on various conditions is discussed. The following behaviors were found to be consistent with the findings described above. Upon solvation, the protein receives more anharmonic perturbations from the solvent molecules and shows more anharmonic dynamics. Thus, the rotation in water occurs in a larger space than that in a vacuum [N_R(in water) > N_R(in vacuum)]. This influence is enhanced at higher temperature [N_R(300 K) > N_R(30 K)]. Averaging a larger value of W makes the potential surface smoother, and N_R becomes a monotonically decreasing function of W.

4. Conclusion

The model of the moving normal mode coordinates was applied to the analyses of MD trajectories of myoglobin. The translation of the origin of the coordinates was decomposed into the contributions of vibration and diffusion. It was found that myoglobin in water at 300 K contains vibrational contributions of the time scale of less than 1 ns. As for the diffusion contributions, the relevance to the functional motions was mentioned. The analyses of the rotation of the coordinate axes revealed that the rotational relaxation of the coordinate system completed quickly within the time window but occurred within a small configurational space spanned by a limited number of large-amplitude normal modes.

We will apply the moving normal mode coordinates to the analyses of nonequilibrium MD simulations of opening/closing motions upon ligand binding to elucidate the role of the anharmonic dynamics in the functional motions.

Acknowledgment. This study was supported by a Grant to A.K. from MEXT. The computations were performed in the Computer Center of the Institute for Molecular Science, in the Center for the Promotion of Computer Science and Engineering of JAERI, and in the Science of Biological Supramolecular Systems of Yokohama City University. We thank Professors T. Komatsuzaki, M. Ikeguchi, and E. K. Lee for helpful discussions.

References and Notes

(1) Frauenfelder, H.; Slinger, S. G.; Wolynes, P. G. *Science* **1991**, *254*, 1598.

(2) Frauenfelder, H.; Parak, F.; Young, R. D. *Annu. Rev. Biophys. Biophys. Chem.* **1988**, *17*, 451.

(3) Karplus, M.; McCammon, J. A. *Nat. Struct. Biol.* **2002**, *9*, 646.

(4) Amadei, A.; Linssen, A. B. M.; Berendsen, H. J. *Proteins* **1993**, *17*, 283.

(5) de Groot, B. L.; van Aalten, D. M. F.; Amadei, A.; Berendsen, H. J. *Biophys. J.* **1996**, *71*, 1707.

(6) Kitao, A.; Hayward, S.; Go, N. *Proteins* **1998**, *33*, 496.

(7) Kitao, A.; Go, N. *Curr. Opin. Struct. Biol.* **1999**, *9*, 164.

(8) Moritsugu, K.; Miyashita, O.; Kidera, A. *Phys. Rev. Lett.* **2000**, *85*, 3970.

(9) Moritsugu, K.; Miyashita, O.; Kidera, A. *J. Phys. Chem. B* **2003**, *107*, 3309.

(10) Hinsen, K.; Thomas, A.; Field, M. J. *Proteins* **1999**, *34*, 369.

(11) Berendsen, H. J.; Hayward, S. *Curr. Opin. Struct. Biol.* **2000**, *10*, 165.

(12) Tama, F.; Wriggers, W.; Brooks, C. L., III. *J. Mol. Biol.* **2002**, *321*, 297.

(13) Broomhead, D. S.; Indik, R.; Newell, A. C.; Rand, D. A. *Nonlinearity* **1991**, *4*, 159.

(14) Krzywda, S.; Murshudov, G. N.; Brzozowski, A. M.; Jaskolski, M.; Scott, E. E.; Klizas, S. A.; Gibson, Q. H.; Olson, J. S.; Wilkinson, A. J. *Biochemistry* **1998**, *37*, 15896.
(15) Morikami, K.; Nakai, T.; Kidera, A.; Saito, M.; Nakamura, H. *Comput. Chem*. **1996**, *100*, 2567.
(16) Sugita, K.; Kitao, A. *Proteins* **1998**, *30*, 388.
(17) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M., Jr.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.: Kollman, P. A. *J. Am. Chem. Soc*. **1995**, *117*, 5179.
(18) Jorgensen, W. D.; Chandrasekhar, J.; Madura, J. D. *J. Chem. Phys.* **1983**, *79*, 926.
(19) Nośe, S. *Mol. Phys.* **1984**, *52*, 255.
(20) Hoover, W. G. *Phys. Rev. A* **1985**, *31*, 1695.
(21) Beglov, D.; Roux, B. *J. Chem. Phys*. **1994**, *100*, 9050.
(22) Ding, H. Q.; Karasawa, N.; Goddard, W. A., III. *J. Chem. Phys*. **1992**, *97*, 4309.
(23) Stratt, R. M. *Acc. Chem. Res*. **1995**, *28*, 201.
(24) Sakurai J. J.; Tuan, S. F. *Modern Quantum Mechanics*, revised ed.; Addison-Wesley: New York, 1995.
(25) Hair, J. F.; Tatham, R. L.; Anderson, R. E.; Black, W. *Multivariate Data Analysis*, 5th ed.; Prentice Hall: New Jersey, 1998.
(26) Garcia, A. E.; Hummer, G. *Proteins* **1999**, *36*, 175.
(27) Berens, P. H.; Mackay, D. H.; White, G. M.; Wilson, K. R. *J. Chem. Phys*. **1983**, *79*, 2375.
(28) Sessions, R. B.; Osguthorpe, P. D.; Osguthorpe, D. J. *J. Mol. Biol*. **1988**, *209*, 617.
(29) Kitao, A.; Hirata, F.; Gō, N. *Chem. Phys.* **1991**, *158*, 447.
(30) Hayward, S.; Kitao, A.; Hirata, F.; Gō, N. *J. Mol. Biol.* **1993**, *234*, 1207.
(31) Evans, G. T. *J. Chem. Phys*. **2002**, *117*, 11284.
(32) Shea, J.-E.; Brooks, C. L., III. *Annu. Rev. Phys. Chem.* **2001**, *52*, 499.
(33) Tame, F.; Brooks, C. L., III. *J. Mol. Biol*. **2002**, *318*, 733.
(34) Tame, F.; Wrrigers, W.; Brooks, C. L., III. *J. Mol. Biol*. **2002**, *321*, 297.
(35) Tama, F. *Protein Pept. Lett.* **2003**, *10*, 119.
(36) Chandrasekhar, S. *Rev. Mod. Phys*. **1943**, *15*, 1.