PROTEINS: Structure, Function, and Genetics 10:106–116 (1991)

Projection of Monte Carlo and Molecular Dynamics Trajectories Onto the Normal Mode Axes: Human Lysozyme

Tokio Horiuchi[1] and Nobuhiro Gō[2]

[1]Protein Engineering Research Institute, Furuedai, Suita 565, and [2]Department of Chemistry, Faculty of Science, Kyoto University, Kyoto 606, Japan

ABSTRACT **A method is presented to describe the internal motions of proteins obtained from molecular dynamics or Monte Carlo simulations as motions of normal mode variables. This method calculates normal mode variables by projecting trajectories of these simulations onto the axes of normal modes and expresses the trajectories as a linear combination of normal mode variables. This method is applied to the result of the molecular dynamics and the Monte Carlo simulations of human lysozyme. The motion of the lowest frequency mode extracted from the simulations represents the hinge bending motion very faithfully. Analysis of the obtained motions of the normal mode variables provides an explanation of the anharmonic aspects of protein dynamics as due first to the anharmonicity of the actual potential energy surface near a minimum and second to trans-minimum conformational changes.**

Key words: collective motion, hinge bending motion, normal mode analysis, anharmonic motion, low-frequency motion

INTRODUCTION

Internal motions of proteins play important roles in their functions. Computer simulation methods are now a powerful theoretical tool to study the internal motions of protein. Three major methods employed for the simulations are (1) normal mode analysis,[1–5] (2) molecular dynamics calculation,[6–9] and (3) Monte Carlo method in the normal mode variable space.[10]

The normal mode analysis has been applied to various proteins.[1–5,11–13] This method describes the internal motions of proteins analytically under the assumption that the energy surface near a potential energy minimum can be approximated by a multidimensional parabola within the range of thermal fluctuations, i.e., by a harmonic function. In this analysis any internal motion in a protein is given by a superposition of normal modes. Each of the normal modes represents a collective motion in the protein. Although the above assumption of the harmonicity is now known not to hold exactly, there are such experimentally observed quantities that agree well either quantitatively or qualitatively with those calculated from the normal mode analysis. For example, the calculated root-mean-square (rms) positional fluctuations of atoms in proteins have been found to correlate well with experimentally observed X-ray temperature factors.[1,3]

Molecular dynamics calculations have been applied to a number of proteins.[6–9] These calculations are free from the assumption of harmonicity. The Monte Carlo simulation is an alternative method to simulate the protein dynamics in a way free from the assumption of harmonicity. One of the important conclusions from these calculations is that the energy surface is such a strongly anharmonic one that there are many local minima within the range of thermal fluctuations.[14–20] Existence of corresponding multiple conformational substates has been shown also experimentally.[21] Thus, proteins appear to have apparent paradoxical dual harmonic and anharmonic aspects; even though the energy surface is strongly anharmonic to have many local minima, proteins behave in some aspects as if the energy surface were harmonic. It is known that the assumption of harmonicity of the energy surface is mostly valid in motions associated with normal modes with medium and high frequencies.[1] This fact may be a hint for the resolution of the paradox. If the shape of the energy surface is such that it is highly anharmonic in such a way that there are many local minima only in the subspace spanned by the low-frequency normal modes, then the harmonic motion as predicted by the normal mode analysis could approximately simulate the motion which in reality is highly anharmonic, at least, for example, as to the amplitude of fluctuations of atomic positions. In this picture protein conformational dynamics should be

Received September 11, 1990; revision accepted November 30, 1990.

Address reprint requests to Professor Nobuhiro Gō, Department of Chemistry, Faculty of Science, Kyoto University, Kitashirakawa, Sakyo-ku, Kyoto 606, Japan.

described in terms of a highly anharmonic coupled motion of low-frequency normal mode variables plus harmonic motions of normal modes with medium and high frequencies.

With such a point of view in mind, we have developed in this study a method to describe internal motions of proteins obtained by Monte Carlo or molecular dynamics simulation as motions of the normal mode variables. This is achieved by calculating a projection of the Monte Carlo or molecular dynamics trajectory onto the normal mode axes. We applied this method to study the dynamics of human lysozyme. Attention is focused to see how various harmonic and anharmonic aspects manifest themselves in the projected motions of normal mode variables.

Special attention is paid to the projected motions of normal mode variables with very low frequencies. Motions of a small number of such modes are known to have dominant contributions to the mean atomic positional fluctuations.[22] They are also believed responsible for functionally related "soft motions." A typical example of such a soft motion is the so-called hinge bending motion of the two lobes forming the active site cleft of lysozyme. This motion in lysozyme has been discussed repeatedly with increasing degrees of refinement in the theoretical treatment.[11,23,24] In the most recent treatment[11] the normal mode analysis has been done for lysozyme and the normal mode with the lowest frequency has been found to be well describable as the hinge bending motion. In this paper we will treat the same problem from the point of view of projection.

METHOD

Projection of the Monte Carlo and the Molecular Dynamics Trajectories Onto the Normal Mode Axes

Normal modes in the dihedral angle space

An instantaneous conformation sampled from the Monte Carlo simulation is characterized by a set of values of deviation of dihedral angles, $\Delta q_j(t)$, $j=1,2,3, \ldots, N_d$ (N_d being the number of dihedral angles in the molecule), from those at a reference conformation, usually the minimum energy conformation. The deviation is expressed by the following linear combination.

$$\Delta q_j(t) = \sum_i u_{ji}\sigma_i(t) \tag{1}$$

where u_{ji} is the j^{th} component of the ith normal mode eigenvector in the dihedral angle space, and $\sigma_i(t)$ is the value of i^{th} normal mode variable at the t^{th} step. The eigenvectors are orthonormalized as follows.

$$\sum_{kl} h_{kl}\, u_{ki}u_{lj} = \delta_{ij} \tag{2}$$

where h_{kl} is an element of the coefficient matrix of the kinetic energy Lagrangian determined for motions satisfying the Eckart condition.[25] Because of Eq. (2), the normal mode variables can be calculated from $\Delta q_j(t)$ by the following projection.

$$\sigma_i(t) = \sum_{kl} \Delta q_k(t) h_{kl} u_{li} \tag{3}$$

Normal modes in the cartesian coordinate space

An instantaneous conformation sampled from the molecular dynamics trajectory is characterized by a set of displacement vectors of constituent atoms, $\Delta\mathbf{r}_a(t)$, $a=1,2,3, \ldots, N_a$ (N_a being the number of constituent atoms) . They are given by the following linear combination.

$$\Delta\mathbf{r}_a(t) = \sum_i \mathbf{v}_{ai}\sigma_i(t) \tag{4}$$

where $\mathbf{v}_{ai}$ is the a^{th} component of i^{th} normal mode eigenvector in the cartesian coordinate space, and $\sigma i(t)$ is the value of the i^{th} normal mode variable at the t^{th} step. The summation extends over normal modes with nonvanishing frequencies. This means that the atomic displacement vectors are such that they do not involve any overall translational and rotational motions. The absence of the overall rotational motion is defined more precisely by that the Eckart condition is always satisfied.[22] Eigenvectors are orthonormalized as follows.

$$\sum_a m_a \mathbf{v}_{ai} \cdot \mathbf{v}_{aj} = \delta_{ij} \tag{5}$$

where m_a is mass of the a^{th} constituent atom. Because of this equation, the normal mode variable $\sigma_i(t)$ can be calculated by the following projection.

$$\sigma_i(t) = \sum_a m_a \Delta\mathbf{r}_a(t) \cdot \mathbf{v}_{ai} \tag{6}$$

The physical dimension of the normal mode variables both in Eqs. (1) and (4) is $(\text{mass})^{1/2}$length and we express them in $\text{gram}^{1/2}$cm.

Relation between the two treatments

Infinitesimal changes in the dihedral angles, Δq_i, cause infinitesimal changes in the atomic positions, $\Delta\mathbf{r}_a$.

$$\Delta\mathbf{r}_a = \sum_i \mathbf{k}_{ai}\Delta q_i \tag{7}$$

An analytic expression of the coefficient $\mathbf{k}_{ai}$ has been derived from the condition that the resulting atomic displacements do not involve any overall translational or rotational motions (the latter in the sense of the Eckart condition).[25] The element, h_{ij}, of the coefficient matrix of the kinetic energy Lagrangian is given in terms of the coefficients in Eq. (7) as follows:

$$h_{ij} = \sum_a m_a \mathbf{k}_{ai} \cdot \mathbf{k}_{aj} \tag{8}$$

From Eqs. (7) and (8) we can derive the following equation of projection:

$$\sigma_i(t) = \sum_{aj} m_a \Delta \mathbf{r}_a(t) \cdot \mathbf{k}_{aj} u_{ji} \tag{9}$$

As compared with Eq. (6), which is the formula for projection onto the normal mode axes in the cartesian coordinate space, this is a formula to be used to calculate projection of molecular dynamics trajectories onto the normal mode axes in the dihedral angle space. Trajectories in the cartesian coordinate space involves, in general, such motions entailing changes in bond lengths and bond angles as well as in dihedral angles. However, bond lengths and bond angles do not change in the normal modes in the dihedral angle space. In this sense the eigenvectors u_{ij} are not complete in the cartesian coordinate space, i.e., a linear combination of u_{ij} cannot represent changes in bond lengths and bond angles. Therefore, in the process of calculating the projection of Eq. (9) changes in bond lengths and bond angles are ignored. Because such changes are generally associated with high frequency motions, we will not be interested in them in this paper.

Contribution to the mass-weighted mean-square displacement of atoms from each normal mode

Now from Eqs. (4) and (5), or from Eqs. (1), (2), (7), and (8) we can derive the following equation:

$$\sum_i \sigma_i^2 = \sum_a m_a \Delta \mathbf{r}_a^2 \tag{10}$$

This equation, derived previously in reference 22, indicates that each normal mode contributes additively to the mass-weighted mean-square displacement of atoms. In the case when the energy surface is purely harmonic, the following relation holds:

$$\sigma_{i\,\mathrm{har}}^2 = \frac{k_B T}{(2\pi\nu_i)^2} \tag{11}$$

Here k_B is the Boltzmann constant and T is the absolute temperature. The suffix har emphasizes that it is the harmonic value. Therefore, one of the measures of the harmonicity of the dynamics of the protein is the ratio of $\langle\sigma_i^2\rangle_{\mathrm{obs}}$ to the value of Eq. (11). This ratio will be referred to as R_i in the following.

Normal Mode Analysis and Monte Carlo Simulation

Normal mode analysis and Monte Carlo simulation of human lysozyme in the dihedral angle space have been carried out *in vacuo* as follows. The initial heavy atom coordinates of human lysozyme are taken from the Protein Data Bank (code:1LZ1[26]). The ECEPP function[27] is used as the empirical conformational energy function. First of all, the cartesian coordinates are regularized so that bond lengths and bond angles have exact standard values and a set of 771 dihedral angles which generates the crystallographic conformation is obtained. This is done by using the distance analysis program (DADAS) developed by Braun and Gō.[28] Next, the conformational energy of the regularized conformation is minimized by using the program based on an algorithm for rapid calculation of gradient and hessian (the second derivative matrix) of the conformational energy.[29–31] The rms displacement between the heavy atom coordinates of energy minimized and the X-ray conformation is 1.50 Å. A set of 771 normal modes is obtained at this minimum energy conformation by following the procedure described by Nishikawa and Gō.[5] For a detailed description of each of the calculated normal modes the reader is referred to the recent paper by Gibrat and Gō.[11] These normal modes are numbered according to the ascending order of their frequencies. Starting from the minimum energy conformation, three series of Monte Carlo simulation are carried out for 100,000, 100,000, and 400,000 steps at three different temperatures, 50K, 100K, and 300K, respectively, following the procedure of Noguti and Gō.[10]. After discarding the initial 20,000 steps which include the equilibration phase, 1,600 conformations are sampled at every 50 steps from the remaining 80,000 steps in the 50K and 100K simulations and 7,600 conformations are sampled in the same manner from the remaining 380,000 steps in the 300K simulation.

Molecular Dynamics Simulation

The molecular dynamics simulation has been carried out in the cartesian coordinate space with the CHARMm program according to the following procedure. The methodology and the energy function are described by Brooks et al.[32] First, polar hydrogens are added to the X-ray structure. After potential energy minimization of the resulting system consisting of 1029 heavy atoms and 270 polar hydrogens, the system is gradually heated to reach the desired temperatures of 100K and 300K. Then, the system is equilibrated for 20 psec. After the equilibration, a 60-psec molecular dynamics simulation is carried out with a time step of 1 fsec using the SHAKE algorithm,[33] out of which 1,200 conformations are sampled at intervals of 50 fsec.

RESULTS

Normal Mode Variables

A set of values of the 771 normal mode variables is calculated by projecting each conformation sampled from the Monte Carlo simulations by Eq. (3) and from the molecular dynamics simulations by Eq. (9) onto the normal mode axes. As a reference conformation to measure deviations of values of either the dihedral angles or the cartesian coordinates, an average conformation of each simulation is used in-

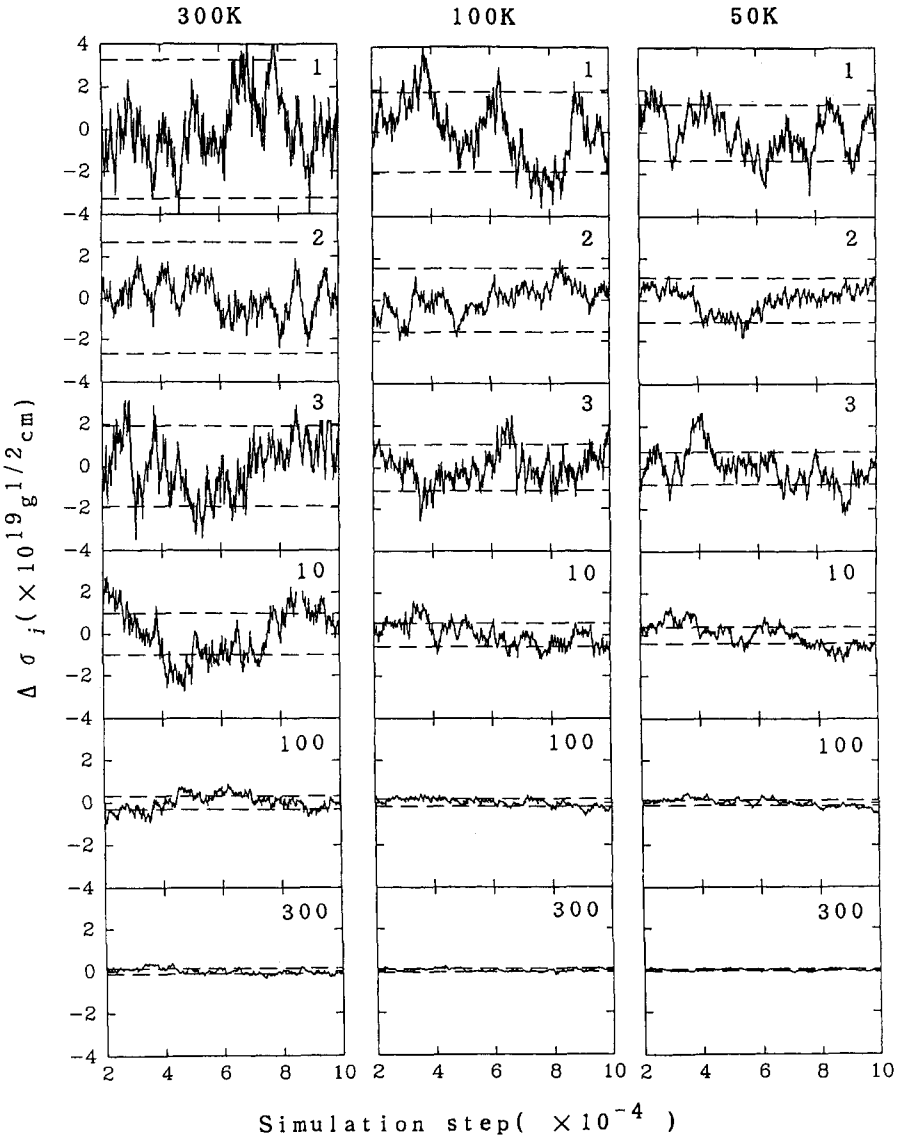

Fig. 1. Values of normal mode variables of mode 1, 2, 3, 10, 100, and 300, calculated from the Monte Carlo simulations at 50K, 100K, and 300K, plotted against simulation step number. A pair of horizontal lines have been drawn at the average rms displacement calculated for the purely harmonic potential. Mode numbers are shown in the upper right corner of each panel.

stead of the minimum energy conformation in the present calculation. Figure 1 shows the records of the conformational dynamics of the Monte Carlo simulations at the three different temperatures as changes of values of several normal mode variables with low frequencies (mode 1, 2, 3, and 10) and with medium frequencies (mode 100 and 300). In the following we will discuss the conformational dynamics of the protein from the point of view of the variation of the normal mode variables.

Hinge Bending Motion

Lysozyme has two domains that are connected by a few strands of polypeptide and shows a characteristic motion called hinge bending, i.e., an opening and closing of the active-site cleft. A number of theoretical studies have been reported on this motion.[11,23,24] In the most recent study[11] it has been shown that the normal mode with the lowest frequency is well characterized by the hinge-bending motion. Therefore, in this paper we are interested in the variation of the normal mode variable with the lowest frequency and in how it is correlated with the hinge bending motion. Figure 2a shows the variation of distance between the C^{α} atoms of residues 48 and 114 during the 380,000 steps of the Monte Carlo simulation at 300K including the first 80,000 steps already shown in Figure 1. As these atoms are lo-

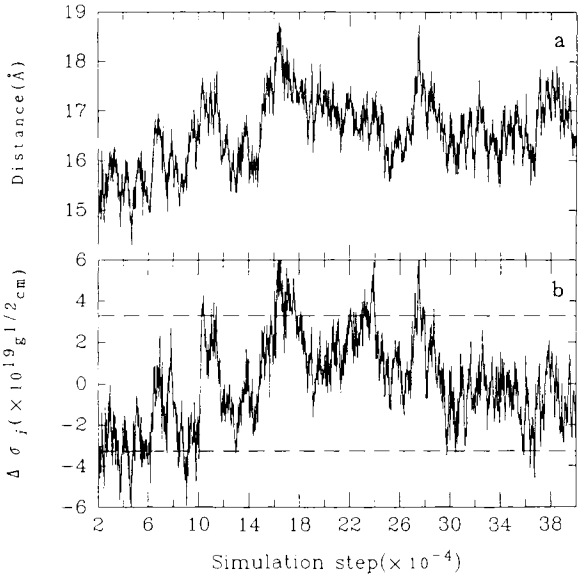

Fig. 2. **(a)** Distance between the C^{α} atoms of residues 48 and 114 and **(b)** normal mode variable with lowest frequency, both plotted against the step number of the Monte Carlo simulation at 300K.

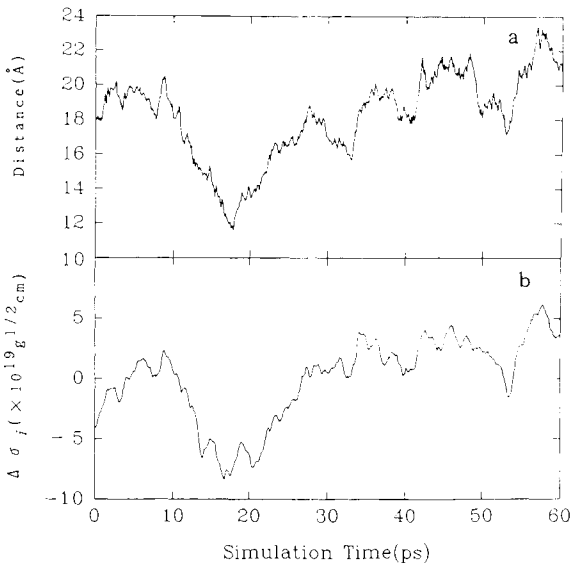

Fig. 3. **(a)** Distance between the C^{α} atoms of residues 48 and 114 and **(b)** normal mode variable with lowest frequency plotted against the time of the molecular dynamics simulation at 300K.

cated on the edges of the active-site cleft, the distance between them is used to monitor the hinge bending motion. Figure 2b shows the variation of the normal mode variable of the lowest frequency mode during the same 380,000 steps of the Monte Carlo simulation. The zero point of the ordinate in each panel of Figure 1 and Figure 2 is taken to coincide with the average of the values of the normal mode variable during the steps shown in each panel. Figure 3a shows the variation of the distance between the two C^{α} atoms during 60 psec of the molecular dynamics simulation at 300K. Figure 3b shows the variation of the normal mode variable of the lowest frequency mode during the same trajectory. The curves in Figures 2a and b and those in Figures 3a and b are very similar to each other. The correlation coefficients of the two lines in Figure 2 and those in Figure 3 and 0.82 and 0.91, respectively. The same analysis is done also for the records of the Monte Carlo simulation at 50K and 100K and for the trajectory of the molecular dynamics simulation at 100K to find again very good correlations of 0.91, 0.89, and 0.80, respectively. These very high correlations indicate that the normal mode with the lowest frequency represents the hinge bending motion quite faithfully. At the same time the method of projection developed in this paper is shown to be a powerful way to extract an important collective motion in a protein which is otherwise rather difficult to detect in the simulated records.

Note that the energy function and the number of degrees of freedom are quite different in the two types of simulation. Yet, essentially the same results are obtained. This fact indicates that the very soft or low frequency motions do not depend very sensitively on each specific choice of the conformational energy function.

Contribution of Each Normal Mode to the Whole Internal Motion

The contribution of each normal mode to the whole internal motion in the Monte Carlo simulations expressed as the mass-weighted mean-square deviation is given by $\langle \sigma_i^2 \rangle$ as shown in Eq. (10). Under the harmonic potential approximation each contribution is related to the frequency as shown in Eq. (11). For the duration of 80,000 steps of each of the three Monte Carlo simulations, the contributions from individual modes are calculated and averaged. The cumulative average contributions from the 37 lowest frequency normal modes (4.7% of 771 modes), which have frequencies lower than 20 cm^{-1}, are 58.9% (50K), 56.4% (100K), 49.0% (300K), and 69.5% (harmonic approximation). The large contribution from a small number of normal modes has been pointed out in the case of the harmonic approximation.[5,22] This is observed also in the projected motions of normal mode variables, even though the latter motions should reflect the anharmonic nature of the protein conformational dynamics. The observed deviations of the percentage contributions from the harmonic approximation reflects the anharmonicity of the Monte Carlo simulation. The harmonicity and anharmonicity of the protein dynamics will be discussed in the following.

TABLE I. Degree of Anharmonicities of the Projected Normal Mode Variables, R_i, and of the Actual Energy Cross Section Along Each of the Normal Modes, R_i'

Mode no.	Frequency (cm^{-1})	Temperature (K)	R_i*	R_i'*
		50	0.793	0.830
1	3.298	100	0.797	0.780
		300	0.480 (0.682)†	0.663
		50	0.573	0.697
2	4.031	100	0.458	0.593
		300	0.322 (0.588)	0.481
		50	1.108	0.872
3	5.624	100	0.730	0.814
		300	0.702 (0.797)	0.700
		50	1.397	0.946
10	10.744	100	0.980	0.919
		300	1.265 (1.143)	0.844
		50	1.513	0.995
100	34.938	100	1.052	0.975
		300	1.240 (1.382)	0.932
		50	1.127	0.976
300	82.567	100	1.006	0.954
		300	1.148 (1.220)	0.907

*For detailed definition, see text.
†Values in parentheses are those calculated for longer 380,000 step Monte Carlo simulation at 300K. Others are calculated for 80,000 step simulation at respective temperatures.

Harmonic Aspect of the Conformational Dynamics

Some statistics of the results shown in Figure 1 are given in Table I. Figure 1 indicates that amplitudes of fluctuation of the normal mode variables agree rather well with those expected theoretically for a purely harmonic potential. In order to quantify this observation the ratio R_i of the observed to the theoretical rms values is calculated and is shown in Table I. The same ratio is calculated for all modes and is shown in Figure 4. This ratio stays roughly between 0.5 and 1.5 for modes with a wide range of frequencies. This fact indicates that the harmonic picture is basically valid to describe the protein dynamics. However, Figure 4 reveals some systematic trends of deviation as well as some abrupt deviations from the harmonicity. (1) The values of R_i for modes with very low frequencies are found smaller than unity. (2) Those with medium and high frequencies are usually equal to or larger than unity. (3) Abrupt deviations such as at mode 102 in 50K simulation are conspicuous. We shall later study the mechanism of such deviations.

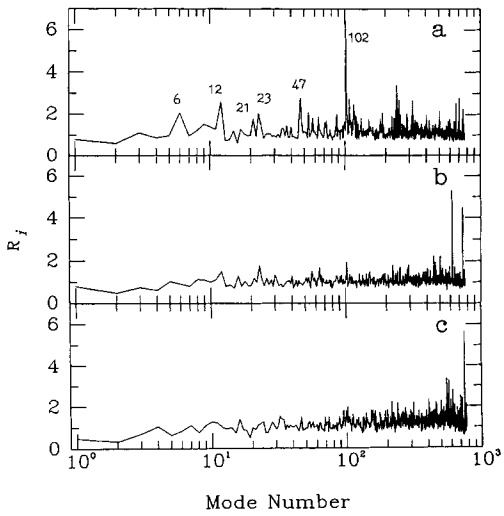

Fig. 4. Variation of the values of R_i calculated from the record of the Monte Carlo simulations at **(a)** 50K, **(b)** 100K, **(c)** 300K, each plotted against the mode number. Mode numbers which have extraordinary large values of R_i at 50K are shown in panel **(a)**.

Anharmonic Aspects of the Conformational Dynamics

Anharmonicity of the actual potential energy surface

As stated above, the values of R_i of modes 1, 2, and 3 are smaller than unity, i.e., 0.480, 0.322, and 0.702 at 300 K, respectively. In order to find the reason for this phenomenon a cross section of the actual energy surface along each of the normal mode variables is calculated. Six examples of the calculated cross sections are shown in Figure 5. Even within the range of the theoretical thermal fluctuations at 50K, significant deviations of the cross section from the harmonic surface are found. On the other hand, at higher frequencies (modes 100 and 300), no significant deviations are found from the harmonic surface. Therefore, the actual internal motions of very low-frequency modes are restricted in narrower ranges than those derived by the harmonic approximation. In order to express this fact quantitatively, the rms deviation of each of the normal mode variables is calculated from the Boltzmann distribution on the obtained cross section and its ratio to the theoretical value for the purely harmonic energy surface is listed in Table I as R_i'. We see that R_i correlates well with R_i'. This indicates that the discrepancy between the observed and theoretical rms values of normal mode variables reflects the anharmonicity of the actual potential energy surface. The phenomenon that the motion of mode 2 is very small [R_2 = 0.322 (at 300K)] is well explained by the extreme anharmonicity of the cross section of actual energy surface along mode 2 [R_2' = 0.481 (at 300K)]. The temperature dependence of the values

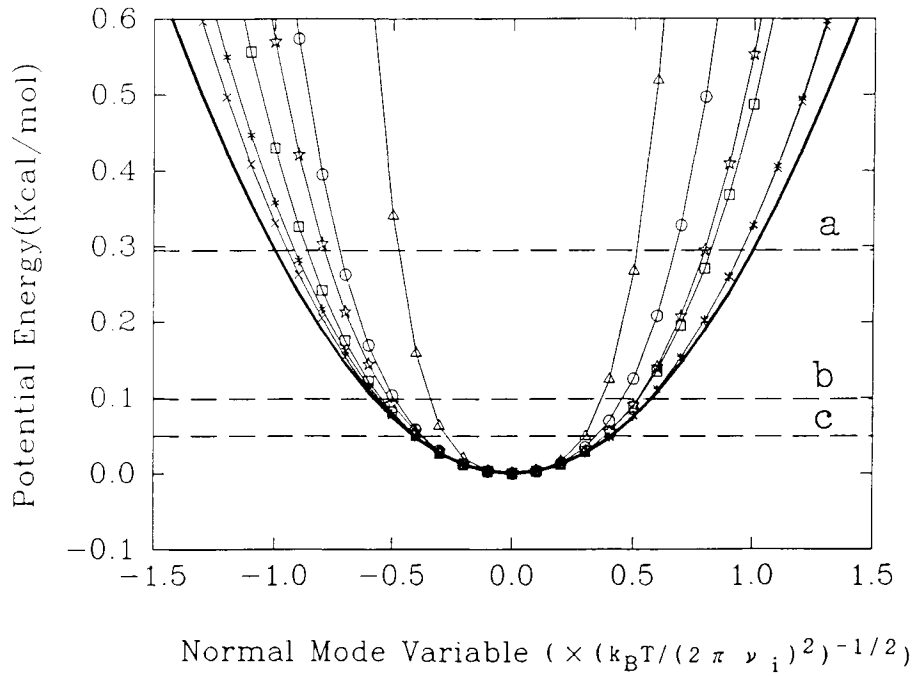

Fig. 5. Cross sections of the actual potential energy surface along each of the normal mode axes of modes 1(○), 2(△), 3(|), 10(□), 100(×) and 300(*) as compared with the purely harmonic potential energy surface (thick solid line). The abscissa is the normal mode variable scaled in each mode by the theoretical rms value of the corresponding purely harmonic variable. In other words, scaling is done in such a way that the second derivatives of the scaled cross sections at the minimum are the same for all curves. Therefore all cross sections in this figure would be identical, if they were purely harmonic. Three broken lines **(a), (b),** and **(c)** represent the average potential energies of $k_BT/2$ at 300K, 100K, and 50K, respectively.

of R_i i.e., the decrease of the values of R_i at low-frequency modes with increase of simulation temperature, is also well explained by the anharmonicity of the potential energy surface.

The values of R_i of mode 100 and 300 at all temperatures examined are larger than unity, although, as shown in Figure 5, the cross sections of the actual potential energy surface along the axes of those normal mode are a bit narrower than those of the harmonic surface. This phenomenon can also be explained by the anharmonicity of the actual potential energy surface. In order to explain this phenomenon, the actual potential energy surface which consists of two normal modes is calculated. Figure 6 shows a contour map of the actual potential energy surface in the two-dimensional space corresponding to modes 2 and 100. In this contour the minimum along each line parallel to the ordinate is located and is shown by a thick line as a function of the value of the abscissa. For a given value, x, of the variable of normal mode 2, a value, y, of the variable of normal mode 100 fluctuates around the minimum point indicated by the thick line. Because the thick line is almost identical with the abscissa for small values of x, motions of the two normal mode variables are independent of each other as long as x is small. Then, the range of fluctuation of the motion of mode 100 is determined by the cross section of energy surface shown in Figure 5. However, when $|x|$ becomes large, y begins to fluctuate around a value dependent on x. In other words, a coupling between the motions of the two variables can be observed when $|x|$ becomes large. Therefore, the range of fluctuation of y is determined by two factors: (1) the change of the minimum point of y as the value of x changes, and (2) the range of fluctuation of y for a given value of x. Factor (1) contributes to increase the value of R_{100}.

Conformational change associated with particular modes

We focus our attention on modes 6, 12, 21, 23, and 102 which in Figure 4 show large deviations from harmonicity in the 50K Monte Carlo simulation. Even though modes such as mode 47 also show some deviation, their absolute contributions to the whole internal motion are small, and therefore we will not be concerned with them. To see what is happening, the calculated variations of these normal mode variables at 50K are plotted in Figure 7. We see that they undergo a large drift between the 60,000th and 80,000th steps. This suggests that a large conformational change occurred during these steps which could be described in terms of a change mainly in these normal mode variables. In order to elucidate such a possible conformational change in detail, the

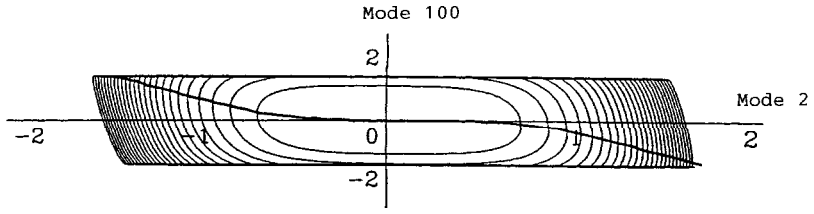

Fig. 6. Contour map of the actual potential energy surface in the two-dimensional space corresponding to modes 2 and 100. Both abscissa and ordinate are scaled normal mode variables as in the abscissa of Figure 5, but actual unit lengths of the abscissa and ordinate are taken to be proportional to the real magnitude of the normal mode variables, so that the contour map shows the actual shape of the potential energy surface. The minima along each line parallel to the ordinate have been connected by a thick line. Contours are drawn at every 1.0 kcal/mol from the minimum value.

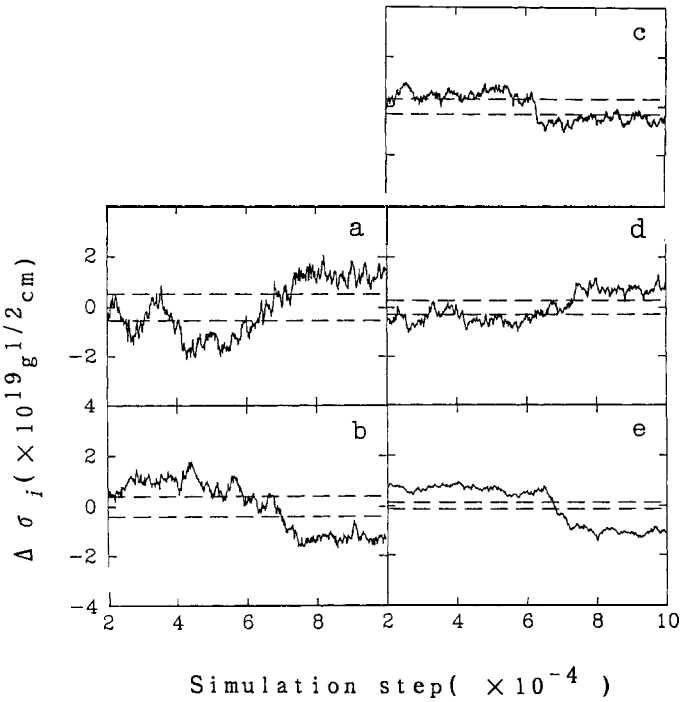

Fig. 7. Same as Figure 1 but for **(a)** mode 6, **(b)** mode 12, **(c)** mode 21, **(d)** mode 23, and **(e)** mode 102 calculated from the Monte Carlo simulation at 50K.

following two average conformations are calculated from the result of the simulation; one averaged over the conformations sampled during 55,000th and 65,000th step (first average conformation) and another over those during 75,000th and 85,000th steps (second average conformation). The rms displacements of heavy atoms between the first and the second average conformations is 0.36 Å. In Figure 8 the difference between positions of heavy atoms in these two conformations is shown by displacement vectors (only those with the displacement larger than 0.5 Å are shown). Obviously, the conformational change is caused by a deformation localized in a small number of residues. Contributions from individual modes to the conformational change shown in Figure 8 are calculated. Four (modes 6, 12, 23, and 102) out of these five modes contribute as much as 35% of this conformational change. Figure 9 shows the displacements of heavy atoms in these four normal modes. A superposition of motions of these four normal modes reproduces the large local conformational change shown in Figure 8 fairly well. This result indicates that the conformational change that occurred during these steps is caused mainly by the motions of these modes. Then, the abrupt deviations from harmonicity suggests the occurrence of a large conformational change such as the "transition from minimum to minimum" described by Noguti and Gō.[15–19] In other words, these results indicate the possibility that this method can be used to detect large conformational changes happened in simulations.

The values of R_i at 300K are seen to increase with increasing number of steps in Table I. This phenom-

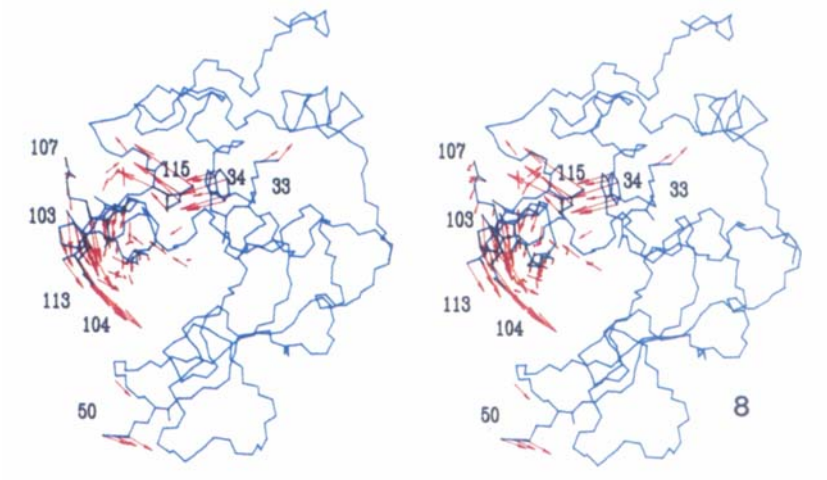

Fig. 8. Stereodrawing illustrating the difference between two conformations, one obtained by averaging over the conformations sampled during step 55,000 through 65,000 and the other during step 75,000 through 85,000 of the 50K simulation. Main chain and several side chains of the first average conformation are shown by blue lines. Vectors connecting heavy atoms in the two average conformations are shown by red arrows (magnified by a factor of 3.0). Only heavy atoms whose displacements are larger than 0.5 Å are shown.

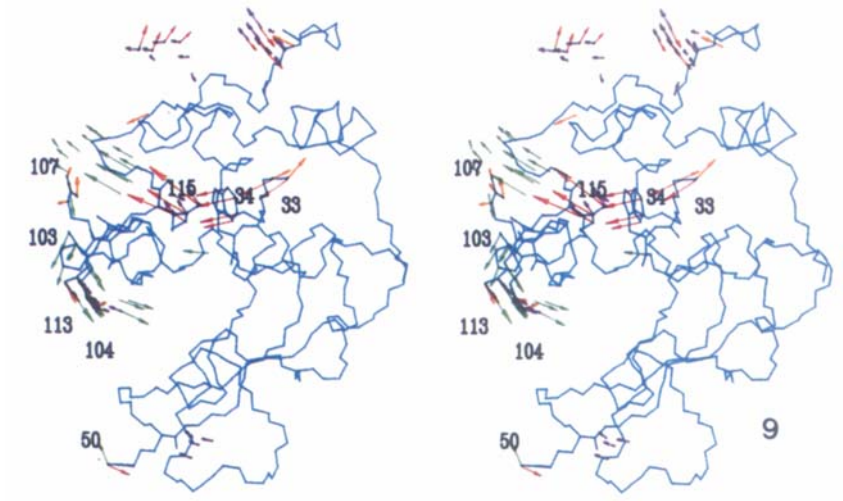

Fig. 9. Stereodrawing of atomic displacement vectors in mode 6 (green), mode 12 (red), mode 23 (violet), and mode 102 (orange). Only the displacement vectors of heavy atoms whose time averaged fluctuations are larger than 0.2, 0.2, 0.1, and 0.05 Å are shown after magnification by factors of 10, 10, 10, and 30 in each of the four modes, respectively. The magnification of the last mode is taken larger because of its large R_i value (see Fig. 4).

enon can also be explained by the conformational change described above. A large number of conformational changes accumulated in a long time simulation inevitably enlarges the value of R_i. But even in the case of 380,000 steps of the Monte Carlo simulation, the calculated ranges of the motion of very low-frequency modes are narrower than those derived from the harmonic approximation. This means that within the range of relatively short steps of simulation the anharmonicity of motions of these modes is mainly explained by the anharmonicity of the actual potential energy surface near each minimum.

DISCUSSION

In this paper we have introduced a method to describe anharmonic protein dynamics obtained by computer simulation methods, such as molecular dynamics simulation and Monte Carlo simulation, as a motion of normal mode variables. This is achieved by projecting the results of simulations onto the axes of normal modes. This method is applied to the Monte Carlo and the molecular dynamics simulations of human lysozyme. The internal motions observed in these simulations can be described by the motions of normal mode variables.

The motions of the lowest frequency normal mode extracted from the Monte Carlo simulation and from the molecular dynamics simulation are found to correlate very well with the hinge bending motions observed in corresponding simulations. This result suggests a more general possibility that a particular functionally important soft motion in a protein can be described by a motion of a normal mode or a coupled motion of several normal modes with very low frequencies, and can be detected from simulation trajectories by the method of projection.

The projected Monte Carlo simulations have then been analyzed mainly from the point of view of amplitude of variation of each normal mode variable as compared with the value expected for the purely harmonic case. The result revealed that the harmonic picture is basically a valid description of the protein dynamics. However, some systematic trends of deviation as well as some abrupt deviations from the harmonic behavior are observed. These anharmonic behaviors observed in protein dynamics are classified into three types: (1) The anharmonicity of the motions of normal modes with very low frequencies. (2) Those with medium or high frequencies. (3) The abrupt anharmonic behavior observed concurrently in several modes. The first two types can be explained by anharmonicities of the actual potential energy surface near a minimum point. The last is associated with a large-scale conformational change which is localized in a part of the protein molecule. Such a motion corresponds in all likelihood to a transition between different local minimum energy conformations. These results indicate that the present method is useful (1) to extract functionally important collective motions or structural transitions as motions of a small number of low-frequency normal mode variables and (2) to study theoretically the anharmonic aspects of internal motions.

At this point we shall discuss the relation of the method of projection developed here to two recent papers.

In the first paper a new method of refinement of protein dynamic structure by X-ray crystallography was developed based on the concept of effective normal modes.[34] This method is to be used in a later stage of the process of X-ray crystallographic analysis, i.e., after a crude structure is determined by conventional methods. By starting from such a crude structure, a minimum energy structure is determined by minimizing an empirical conformational energy function, and normal mode analysis is carried out for this minimum energy structure. Then, by using the calculated normal mode eigenvectors, an instantaneous atomic displacement is expressed by Eq. (4). If the conformational energy surface can be approximated by a multidimensional parabola within the range of thermal fluctuation, then the distribution of each normal mode variable $\sigma_i(t)$ is (1) independent, and is (2) distributed individually according to a Gaussian distribution. However, as we have seen in this paper, the distribution of the values of $\sigma_i(t)$ deviates from such ideal harmonic behavior. Now, in this new method of refinement we make the assumption that distribution of $\sigma_i(t)$, $(i=1,2,\ldots)$ is not necessarily independent, but is still distributed according to a multidimensional Gaussian distribution. With this assumption, the distribution is completely characterized by its second moments, i.e., $\langle\sigma_i(t)\sigma_j(t)\rangle$, $(i,j=1,2,\ldots)$. Such a multidimensional Gaussian distribution occurs, when σ_i is a linear combination of a set of variables, τ_k, which are distributed according to independent Gaussian distributions, i.e.,

$$\sigma_i = \sum_k \rho_{ik}\tau_k \tag{12}$$

where ρ_{ik} is an element of an orthogonal matrix which diagonalizes $\langle\sigma_i(t)\sigma_j(t)\rangle$ as follows:

$$\langle\sigma_i\sigma_j\rangle = \sum_k \rho_{ik}\rho_{jk}\langle\tau_k^2\rangle \tag{13}$$

The variables τ_k are defined as the effective normal mode variables. The eigenvector of the effective normal mode is then given by

$$\mathbf{w}_{ak} = \sum_i \mathbf{v}_{ai}\rho_{ik} \tag{14}$$

The refinement of the dynamic structure of a protein is carried out by treating the second moments $\langle\sigma_i(t)\sigma_j(t)\rangle$ for a small number of normal modes with very low frequencies as adjustable parameters to fit the temperature factors against the experimentally observed diffraction intensities. It has been shown that this new method is quite powerful in improving the refinement and even more in obtaining information about anisotropic and concerted motions of atoms in a protein molecule.[34] This success is in a sense a surprise, because the method is based on the nontrivial assumption of the multidimensional Gaussian distribution of the normal mode variables, σ_i, $(i=1,2,\ldots)$.

The method developed in the present paper provides a firm basis of discussing the real distribution of the normal mode variables. A few typical modes of the anharmonic behaviors have been discussed. By further carrying out studies along this line, it may become possible to improve the newly proposed refinement method by explicitly taking into account some of the anharmonic behavior.

In the second paper Sessions et al. introduced the method of filtering motions from molecular dynamics trajectories in a specified frequency range.[35] This is done by Fourier-transforming the trajectory and back-transforming only those Fourier components that lie in a specified frequency range. This method was applied to reveal low-frequency collective motions in phospholipase A_2. Both the projection method developed in this paper and the filtering

method in frequency space are similar in the spirit of focusing attention on low-frequency motions. Both methods are able to elucidate complex anharmonic motions in proteins. One difference between the two methods is that in the projection method information specific to a given protein under consideration, i.e., specific forms of eigenvectors, is taken into account already in the method. This may make the description of some specific motions such as hinge bending motion in lysozyme simpler in the projection method.

ACKNOWLEDGMENTS

We are grateful for helpful discussions about the normal mode analysis of lysozyme to Dr. J. Higo, Dr. R. Kuroki, and Dr. J.-F. Gibrat. This work has been supported at Kyoto University by grants to N.G. from the Ministry of Education, Science and Culture, Japan and from the Science and Technology Agency, Japan.

REFERENCES

1. Gō, N., Noguti, T., Nishikawa, T. Dynamics of a small globular protein in terms of low-frequency vibrational modes. Proc. Natl. Acad. Sci. U.S.A. 80:3696–3700, 1983.
2. Brooks, B., Karplus, M. Harmonic dynamics of proteins: Normal modes and fluctuations in bovine pancreatic trypsin inhibitor. Proc. Natl. Acad. Sci. U.S.A. 80:6571–6575, 1983.
3. Levitt, M., Sander, C., Stern, P.S. The normal modes of a protein: Native bovine pancreatic trypsin inhibitor. Int. J. Quant. Chem.: Quant. Biol. Symp. 10:181–199, 1983.
4. Levitt, M., Sander, C., Stern, P.S. Protein normal-mode dynamics: Trypsin inhibitor, crambin, ribonuclease and lysozyme. J. Mol. Biol. 181:423–447, 1985.
5. Nishikawa, T., Gō, N. Normal modes of vibration in bovine pancreatic trypsin inhibitor and its mechanical property. Proteins 2:308–329, 1987.
6. Karplus, M., McCammon, J.A. The internal dynamics of globular proteins. CRC Crit. Rev. Biochem. 9:293–349, 1981.
7. Levitt, M. Protein conformation, dynamics and folding by computer simulation. Annu. Rev. Biophys. Bioeng. 11: 251–271, 1982.
8. Van Gunsteren, W.F., Berendsen, H.J.C., Hermans, J., Hol, W.G.J., Postma, J.P.M. Computer simulation of the dynamics of hydrated protein crystals and its comparison with X-ray. Proc. Natl. Acad. Sci. U.S.A. 80:4315–4319, 1983.
9. Levy, R.M., Keepers, J.W. Computer simulation of protein dynamics: Theory and experiment. Comments Mol. Cel. Biophys. 3:273–294, 1986.
10. Noguti, T., Gō, N. Efficient Monte Carlo method for simulation of fluctuating conformations of native proteins. Biopolymers 24:527–546, 1985.
11. Gibrat, J.-F., Gō, N. Normal mode analysis of human lysozyme: Study of the relative motion of the two domains and characterization of the harmonic motion. Proteins 8: 258–279, 1990.
12. Seno, Y., Gō, N. Deoxymyoglobin studied by the conformational normal mode analysis: I. Dynamics of globin and the heme-globin interaction. J. Mol. Biol., in press.
13. Seno, Y., Gō, N. Deoxymyoglobin studied by the conformational normal mode analysis: II. The conformational change upon oxygenation. J. Mol. Biol., in press.
14. Elber, R., Karplus, M. Multiple conformational states of proteins: A molecular dynamics analysis of myoglobin. Science 235:318–321, 1987.
15. Noguti, T., Gō, N. Structural basis of hierarchical multiple substates of a protein. I: Introduction. Proteins 5:97–103, 1989.
16. Noguti, T., Gō, N. Structural basis of hierarchical multiple substates of a protein. II: Monte Carlo simulation of native thermal fluctuations and energy minimization. Proteins 5:104–112, 1989.
17. Noguti, T., Gō, N. Structural basis of hierarchical multiple substates of a protein. III: Side chain and main chain local conformations. Proteins 5:113–124, 1989.
18. Noguti, T., Gō, N. Structural basis of hierarchical multiple substates of a protein: IV: Rearrangements in atom packing and local deformations. Proteins 5:125–131, 1989.
19. Noguti, T., Gō, N. Structural basis of hierarchical multiple substates of a protein. V: Nonlocal deformations. Proteins 5:132–138, 1989.
20. Gō, N., Noguti, T. Structural basis of hierarchical multiple substates of a protein. Chem. Scripta 29A:151–164, 1989.
21. Frauenfelder, H., Parak, F., Young, R.D. Conformational substates in proteins. Annu. Rev. Biophys. Biophys. Chem. 17:451–479, 1988.
22. Gō, N. A theorem on amplitudes of thermal atomic fluctuations in large molecules assuming specific conformations calculated by normal mode analysis. Biophys. Chem. 35:105–112, 1990.
23. McCammon, J.A., Gelin B., Karplus, M., Wolynes, P.G. The hinge-bending mode in lysozyme. Nature (London) 262:325–326, 1976.
24. Brooks, B., Karplus, M. Normal modes for specific motions of macromolecules: Application to the hinge-bending mode of lysozyme. Proc. Natl. Acad. Sci. U.S.A. 82:4995–4999, 1985.
25. Noguti, T., Gō, N. Dynamics of native globular proteins in terms of dihedral angles. J. Phys. Soc. Jpn. 52:3283–3288, 1983.
26. Artymiuk, P.J., Blake, C.C.F. Refinement of human lysozyme at 1.5Å resolution: Analysis of non-bonded and hydrogen-bond interactions. J. Mol. Biol. 152:737–762, 1981.
27. Némethy, G., Pottle, M.S., Scheraga, H.A. Energy parameters in polypeptides. 9. Updating of geometrical parameters, nonbonded interactions, and hydrogen bond interactions for the naturally occurring amino acids. J. Phys. Chem. 87:1883–1887, 1983.
28. Braun, W., Gō, N. Calculation of protein conformations by proton-proton distance constraints: A new efficient algorithm. J. Mol. Biol. 186:611–626, 1985.
29. Noguti, T., Gō, N., A method of rapid calculation of a second derivative matrix of conformational energy for large molecules. J. Phys. Soc. Jpn. 52:3685–3690, 1983.
30. Abe, H., Braun, W., Noguti, T., Gō, N. Rapid calculation of first and second derivatives of conformational energy with respect to dihedral angles for proteins. General recurrent equations. Comput. Chem. 8:239–247, 1984.
31. Wako, H., Gō, N. Algorithm for rapid calculation of Hessian of conformational energy function of proteins by supercomputer. J. Comp. Chem. 8:625–635, 1987.
32. Brooks, B.R., Bruccoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S., Karplus, M. CHARMm: A program for macromolecular energy, minimization and dynamics calculation. J. Comput. Chem. 4:187–217, 1983.
33. van Gunsteren, W.F., Berendsen, H.J.C. Algorithm for macromolecular dynamics and constraint dynamics. Mol. Phys. 34:1311–1327, 1977.
34. Kidera, A., Gō, N. Refinement of protein dynamic structure: Normal mode refinement. Proc. Natl. Acad. Sci. U.S.A. 87:3718–3722, 1990.
35. Sessions, R. B., Dauber-Osguthorpe, P., Osguthorpe, D.J. Filtering molecular dynamics trajectories to reveal low-frequency collective motions: Phospholipase A2. J. Mol. Biol. 209:617–633, 1988.