Eur Biophys J (2000) 29: 472–480
DOI 10.1007/s002490000103

ARTICLE

Florence Tama · Osamu Miyashita · Akio Kitao
Nobuhiro Go

Molecular dynamics simulation shows large volume fluctuations of proteins

Received: 2 May 2000 / Revised version: 9 August 2000 / Accepted: 9 August 2000 / Published online: 24 October 2000

Abstract In this paper we present a new approach to study the volume fluctuations of proteins. From a 1 ns molecular dynamics simulation, the volume fluctuation of human lysozyme has been calculated. We used two different ways for the calculation. In the first one, the volume fluctuation is extracted directly from the trajectory. For the second one, a newly developed formalism based on principal component analysis is used. The r.m.s. volume fluctuations obtained from the two analyses agree well with each other. The isothermal intrinsic compressibility was found to be larger than the one reported by experiment. The difference is discussed and suggested to exist in the assumed uncertainty of the compressibility of hydrated water to deduce the isothermal intrinsic compressibility from the experimental value. Spectral analysis shows that low-frequency dynamics dominate the total volume fluctuation. The same aspect is found in the study using principal component analysis. This low-frequency region is related to large and slow motions of proteins. Therefore a long time dynamics simulation is necessary to describe the volume fluctuations of proteins.

Key words Compressibility · Principal component analysis · Human lysozyme

Introduction

For the purpose of revealing the physicochemical properties of proteins, a lot of experimental and theoretical studies have been done by taking the temperature as a control parameter. The temperature is generally related to the energetic properties of a system. Another important parameter is the pressure. By changing the pressure as a control parameter, we can obtain new information generally related to the volumetric property of the system. The dynamic properties of the three-dimensional structures of proteins are of general interest. The three-dimensional structures of proteins can be related to the volume. From study of the effects of pressure on proteins we can obtain information, via the volume, about the dynamics of proteins.

The partial molar volume of a protein, V°, represents the sum of the intrinsic geometric volume of a protein molecule, V_M, and the hydration effects, ΔV_h. The intrinsic volume, V_M, is equal to the sum of the van der Waals' volumes of all constituent atoms of the protein, V_W, plus the volume of the cavities inside the protein, V_V, resulting from its imperfect internal packing. The quantity V_M reflects the intrinsic spatial architecture of the protein. Thus the three-dimensional structure of a protein is directly related to V_M, which is our prime interest. The hydration contribution, ΔV_h, to the partial molar volume reflects the solvent contraction due to its interaction with the surface atomic groups of a protein.

The partial molar adiabatic compressibility of a protein, $\bar{\beta}_S$, is defined as the pressure derivative of the partial molar volume, V°. The partial molar adiabatic compressibility $\bar{\beta}_S$ can be represented by a sum of the intrinsic, β_S, and hydration, $\Delta\beta_h$, contributions. The intrinsic compressibility of a protein, β_S, is a measure of the intramolecular interactions, whereas the hydration-induced change in solvent compressibility, $\Delta\beta_h$, reflects the interactions of the solvent molecules with the protein. β_S reflects the dynamic properties of a protein, because the mean-square volume fluctuation of a protein molecule is related to β_S.

There are different experimental approaches to study the compressibility of proteins. Most of the experimental data on the compressibility of proteins comes from sound velocity experiments (Gekko and Noguchi 1979;

F. Tama · O. Miyashita · A. Kitao · N. Go (✉)
Department of Chemistry, Graduate School of Science,
Kyoto University, Kyoto 606-8502, Japan
e-mail: go@qchem.kuchem.kyoto-u.ac.jp
Fax: +81-75-7533669

F. Tama
Laboratoire de Physique Quantique,
UMR 5626 of CNRS, IRSAMC, Université Paul Sabatier,
118 Route de Narbonne, 31062 Toulouse Cédex, France

Gavish et al. 1983; Gekko and Hasegawa 1986; Kharakoz and Sarvazyan 1993). Fluorescence spectroscopy (Marden et al. 1986), hole-burning (Zollfrank et al. 1991), NMR (Li et al. 1998) or crystallography (Kundrot and Richards 1987) have been also used to determine the compressibility of proteins.

A theoretical approach to understanding the volume properties of proteins has been done. Pressure effects on the equilibrium conformation and its volume fluctuation have been studied by normal mode analysis (Yamato et al. 1993; Kobayashi et al. 1997; Miyashita and Go 1999). In this analysis, the volume fluctuation is given as a sum of the contributions from individual normal modes. Contributions from low-frequency normal modes were found to dominate those from higher-frequency normal modes. However, there are some limitations in normal mode analysis. Normal mode analysis assumes that the protein conformational energy is harmonic within the range of the thermal fluctuations and simulation is usually done in vacuum. In other words, the normal mode reflects only local minima of the potential energy surface and protein dynamics of a short timescale. However, it has been shown that the real energy surface of a protein surrounded by water solvent has considerable anharmonicity and many sub-states. Therefore the treatment based on the protein normal mode analysis must be extended to treat the anharmonic character of protein dynamics in water. The molecular dynamics (MD) simulation provides an alternative to the study by the normal mode since large conformational fluctuations can be covered by long timescale dynamics.

In this sense, MD simulation can solve these limitations by treating water molecules explicitly. In a recent paper (Kitao et al. 1998) a 1 ns simulation of lysozyme has been done. In the present study we analyze the volume fluctuation of the protein along this long trajectory. Moreover, in order to obtain a more precise picture of the volume fluctuation of the protein, we develop a new formalism to describe volume dynamics from the principal mode. With the principal mode we can understand the dynamic properties of proteins in terms of the three-dimensional structure.

In the second section of this paper we present a method used to study volume properties of proteins from dynamics simulation. In the third section, results from the direct study of the simulation and the study from the principal component analysis are described. Finally, we discuss a comparison with the experimental results.

Methods

Definition of volume

Various geometric definitions of the volume of a protein have been proposed. The most widely used definitions of protein volume are the *van der Waals*, the *molecular*, the *excluded* and the *Voronoi* volumes (Paci and Velikson 1997).

In the present study, the volume of lysozyme is defined as the excluded volume in solvent water. In this definition, a lysozyme molecule is regarded as consisting of overlapping spheres, each of which represents a constituent atom. The radius of a sphere is taken as the sum of the van der Waals radii of the solvent molecule and of the atom. The effective van der Waals radius of a solvent water molecule is taken to be 1.4 Å. The volume thus defined contains the contribution from a half-layer of water molecules on the surface of lysozyme. This geometrical volume is a microscopic observable property, in the sense that its value is a function of the positions of all atoms of the protein. The excluded volume of the molecule is calculated using the NSC subroutine (Eisenhaber et al. 1995).

Principal component analysis

Principal component analysis is a powerful method to analyze the region of conformational space explored by a trajectory (Kitao et al. 1991; Hayward et al. 1994). A set of principal components is determined as a solution of eigenvalue problems of the second-moment matrix of the mass-weighted internal atomic displacements x. Therefore we will study the following second-moment matrix, $\mathbf{C}$, whose matrix element, $c_{i,j}$, is:

$$c_{i,j} = \left\langle \left(x_i - \langle x_i \rangle\right)\left(x_j - \langle x_j \rangle\right)\right\rangle \tag{1}$$

where $\langle \ldots \rangle$ means an average over instantaneous structures sampled during the simulation. Diagonalization of the symmetric matrix $\mathbf{C}$ results in a diagonal eigenvalue matrix Λ and an orthogonal matrix $\mathbf{A}$ whose columns are the eigenvectors or principal vectors. The elements Λ_m of the diagonal matrix Λ are the variances, or mean-square fluctuations of the principal component variables. The trajectory, $q_j(t)$, of the jth principal component variable is given by:

$$q_j(t) = \sum_i \left(x_i(t) - \langle x_i \rangle\right) a_{ij} \tag{2}$$

where a_{ij} is the element for atom i of the jth eigenvector of the matrix $\mathbf{A}$. The trajectory of each principal component variable is obtained from the 1 ns dynamics. From this trajectory we can calculate the correlation between principal component variables during the dynamics.

Volume fluctuations from principal component analysis

In the present paper we study the volume fluctuation in terms of the principal component variables. Volume is defined as a function of coordinates, i.e., principal component variables. The volume V is expanded in powers of q and approximated to the second order:

$$V(q) = V_0 + \sum_i \frac{\partial V}{\partial q_i} q_i + \frac{1}{2} \sum_{ij} \frac{\partial^2 V}{\partial q_i \partial q_j} q_i q_j \tag{3}$$

where q_i is the ith principal component variable. The deviation of the volume from the average is:

$$\Delta V(q) = V(q) - \langle V \rangle \tag{4}$$

where $\langle \ldots \rangle$ means an average over the trajectory. Then:

$$\Delta V(q) = \sum_i \frac{\partial V}{\partial q_i} \left(q_i - \langle q_i \rangle\right) + \frac{1}{2} \sum_{ij} \frac{\partial^2 V}{\partial q_i \partial q_j} \left(q_i q_j - \left\langle q_i q_j \right\rangle\right) \tag{5}$$

Volume fluctuation can be derived from this equation as:

$$\begin{aligned} \Delta V^2(q)\rangle = \Bigg\langle & \sum_{ij} \frac{\partial V}{\partial q_i} \frac{\partial V}{\partial q_j} \left(q_i - \langle q_i \rangle\right)\left(q_j - \left\langle q_j \right\rangle\right) \\ & + \sum_{ijk} \frac{\partial V}{\partial q_i} \frac{\partial^2 V}{\partial q_j \partial q_k} \left(q_i - \langle q_i \rangle\right)\left(q_j q_k - \left\langle q_j q_k \right\rangle\right) \\ & + \frac{1}{4} \sum_{ijkl} \frac{\partial^2 V}{\partial q_i \partial q_j} \frac{\partial^2 V}{\partial q_k \partial q_l} \left(q_i q_j - \left\langle q_i q_j \right\rangle\right)\left(q_k q_l - \langle q_k q_l \rangle\right) \Bigg\rangle \end{aligned} \tag{6}$$

and finally, with $\langle q_i \rangle = 0$ and $\langle q_i q_j \rangle = 0$ for $i \neq j$, the volume fluctuation is given by:

$$\langle \Delta V^2(q) \rangle = \sum_i \left(\frac{\partial V}{\partial q_i} \right)^2 \langle q_i^2 \rangle + \sum_{ijk} \frac{\partial V}{\partial q_i} \frac{\partial^2 V}{\partial q_j \partial q_k} \langle q_i q_j q_k \rangle + \frac{1}{4} \sum_{ijkl} \frac{\partial^2 V}{\partial q_i \partial q_j} \frac{\partial^2 V}{\partial q_k \partial q_l} \left(\langle q_i q_j q_k q_l \rangle - \langle q_i q_j \rangle \langle q_k q_l \rangle \right) \quad (7)$$

In the following, according to the order of q_i, the first term of this equation will be referred to as second order, the second term as third order, and the last term as fourth order.

Simulation

The simulation used in this study and the subsequent principal component analysis are the same as the ones described in a previous paper (Kitao et al. 1998). The crystal structure of human lysozyme determined by the method of normal mode refinement (Kidera et al. 1992) is employed as an initial coordinate of the simulation. We use a newly developed program package (Sugita and Kitao 1998) based on the framework of the minimization/molecular dynamics program PRESTO (Morikami et al. 1992). The spherical solvent boundary potential (SSPB) (Beglov and Roux 1994), the cell multipole method (CMM) (Ding et al. 1992) and the Nosé-Hoover algorithm (Nosé 1984; Hoover 1985) are also used. Two chloride ions and a total of 4684 explicit waters are included in the system. AMBER potential energy functions (Weiner et al. 1986) and TIP3P water models (Jorgensen et al. 1983) are employed. The MD simulation is carried out as follows: 50 ps MD simulation is performed in order to equilibrate the system to 300 K and 1 atm, by gradually relaxing the constraints; 50 ps MD is performed without constraints for further equilibration; 1 ns MD is carried out and is used for analysis.

Results

Volume fluctuation directly from trajectory

The change of volume during the dynamics is calculated from the 1 ns trajectory; 20,000 structures have been taken from the trajectory (each 50 fs) in order to evaluate the volume fluctuations. Figure 1 shows the root-mean-square deviation (r.m.s.d.) for heavy atoms and the value of the volume observed during the 1 ns trajectory. From the dynamics, the value obtained for the r.m.s. volume fluctuation is 352 Å^3. The average volume for the 1 ns of the dynamics is 27,633 Å^3. We obtain in this case a relative volume fluctuation of 1.27%.

The r.m.s. volume fluctuation is relatively large compared to the value observed by Paci and Marchi (1996). In their simulation of egg lysozyme the relative volume fluctuation was 0.27%. This difference comes from the fact that they considered only 100 ps of dynamics compared to our 1 ns simulation. As can be seen from Fig. 1c, the r.m.s. volume fluctuation depends strongly on the simulation period. If only the first 100 ps of trajectory is used from our simulation, the relative volume fluctuation is 0.49%, which is closer to the value obtained by Paci and Marchi.

In order to know the spectral character of the volume fluctuation, we calculate the Fourier transform of the volume $V(t)$:

$$V(\omega) = \int_0^T \mathrm{d}t (V(t) - \langle V \rangle) \mathrm{e}^{i\omega t} \quad (8)$$

We obtain the power spectral density as follows:

$$P(\omega) = \frac{2}{T} |V(\omega)|^2 \quad (9)$$

Power spectral densities of the volume obtained for the 1 ns and 100 ps simulations are shown in Fig. 2a and b, respectively. In Fig. 2c the corresponding cumulative

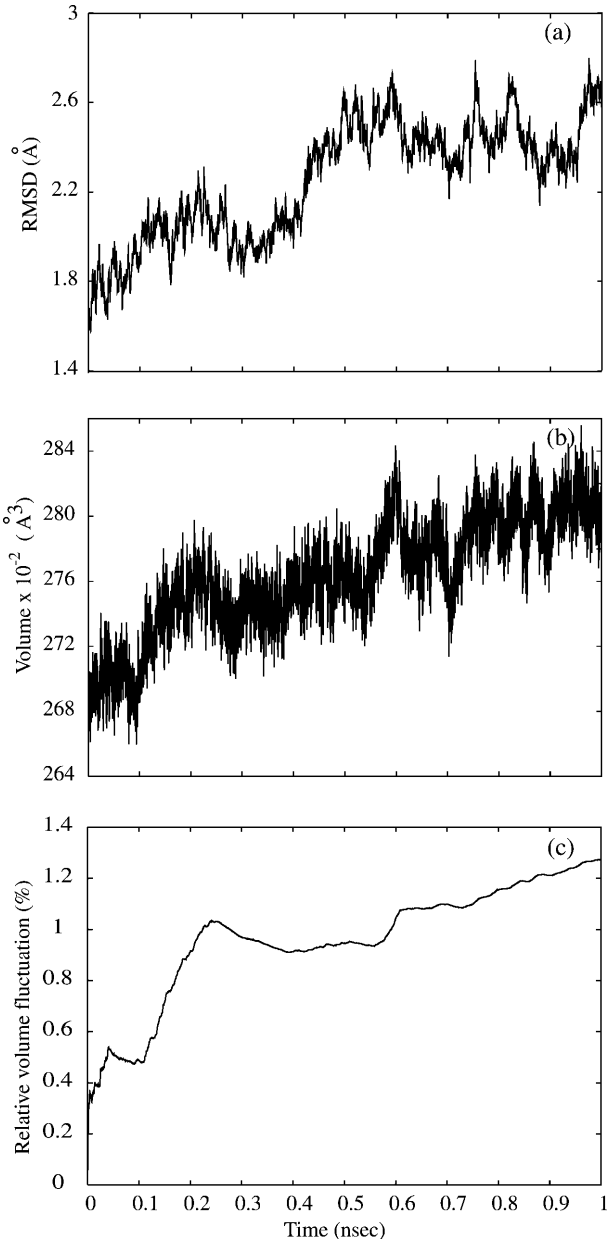

Fig. 1 **a** The root-mean-square deviation (r.m.s.d.) for heavy atoms as a function of time. **b** Volume as a function of time. **c** The dependence of the r.m.s. volume fluctuation on the period, the initial time of the period being time zero

sums (from higher frequencies) are shown (also for the 200 ps and 500 ps trajectories). The cumulative sum over the whole frequency range gives the mean-square fluctuation along the trajectory. We see in Fig. 2c that the contributions from frequencies 1 cm^{-1} and above do not depend on the time length of the trajectory. Also we see that most of the volume fluctuation is due to contributions from the very-low-frequency region. Moreover, the time length of the dynamics is important. We see a large difference in the contributions from the low-frequency region to the volume fluctuation when different time lengths are used.

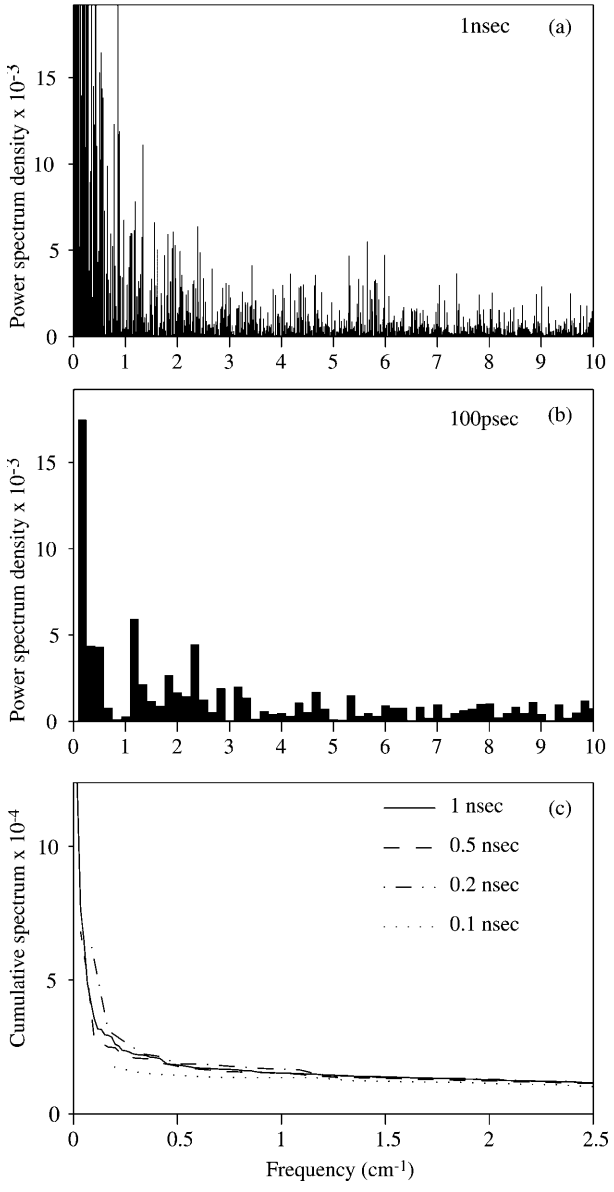

Fig. 2 Power spectral density of volume: **a** from 1 ns trajectory, **b** from 100 ps trajectory, and **c** cumulative sum for trajectories with four different time lengths

Volume fluctuation from principal component analysis

Volume profile

In Fig. 3 we present the projected trajectory and the volume profile on a two-dimensional subspace spanned by a pair of principal modes. In Fig. 4 we present volume profiles in one-dimensional subspaces. The volume profile in the two-dimensional space is directly calculated from structures that we obtain by displacing the structure from the average along the two principal modes. For each profile, 11×11 points are used. Projection of the trajectory onto the same two-dimensional subspace is also shown. Note that the values of the volume in the two-dimensional profile are smaller than those during the dynamics. This is because contributions of only two modes are included in the profile, while those during the dynamics include contributions from all modes. The origin of the two-dimensional space corresponds to the average structure of the dynamics. Its volume is 25,450 $Å^3$. (Here we parenthetically comment that the average structure is obtained by averaging the position of each atom in the molecule in the Cartesian space over the whole trajectory with a precaution that prior to the averaging the translational and rotational degrees of freedom of the molecule are suppressed.)

From these plots we see that the trajectory has undergone a large-scale migration in the subspace spanned by the first three modes. During this migration the protein moved from a region with a small volume (×) to a region with a large volume (○).

From Fig. 3d we see that along the direction of the 100th mode there are almost vertical linear changes of the volume, i.e, the first-order term of Eq. 3 is very small and the second-order term nearly vanishes. So, in Eq. 7 the higher-order terms need not be included except for the first several modes.

Volume fluctuation

For the calculation of the volume fluctuation from principal component analysis, we need to calculate first and second derivatives of the volume (see Eq. 7). This is done numerically from the calculated volume profiles.

From the trajectory we calculate other terms needed to obtain the volume fluctuation, i.e., $\langle q_i q_j q_k \rangle$ and $\langle q_i q_j q_k q_l \rangle - \langle q_i q_j \rangle \langle q_k q_l \rangle$. In the summation of the second term (the third-order term) of Eq. 7 we consider only the first 500 modes for the first derivatives and only the first 10 modes for the second derivatives. In the third term (fourth-order term) we consider only the first 10 modes from the summation owing to the fact that the second derivatives for higher modes are very small (see Fig. 4) and, in the case of the third-order term, that non-Gaussian correlations $\langle q_i q_j q_k \rangle$ become very small for higher modes.

The volume fluctuation obtained from Eq. 7 is 364 $Å^3$. This value is very close to the value (352 $Å^3$) directly

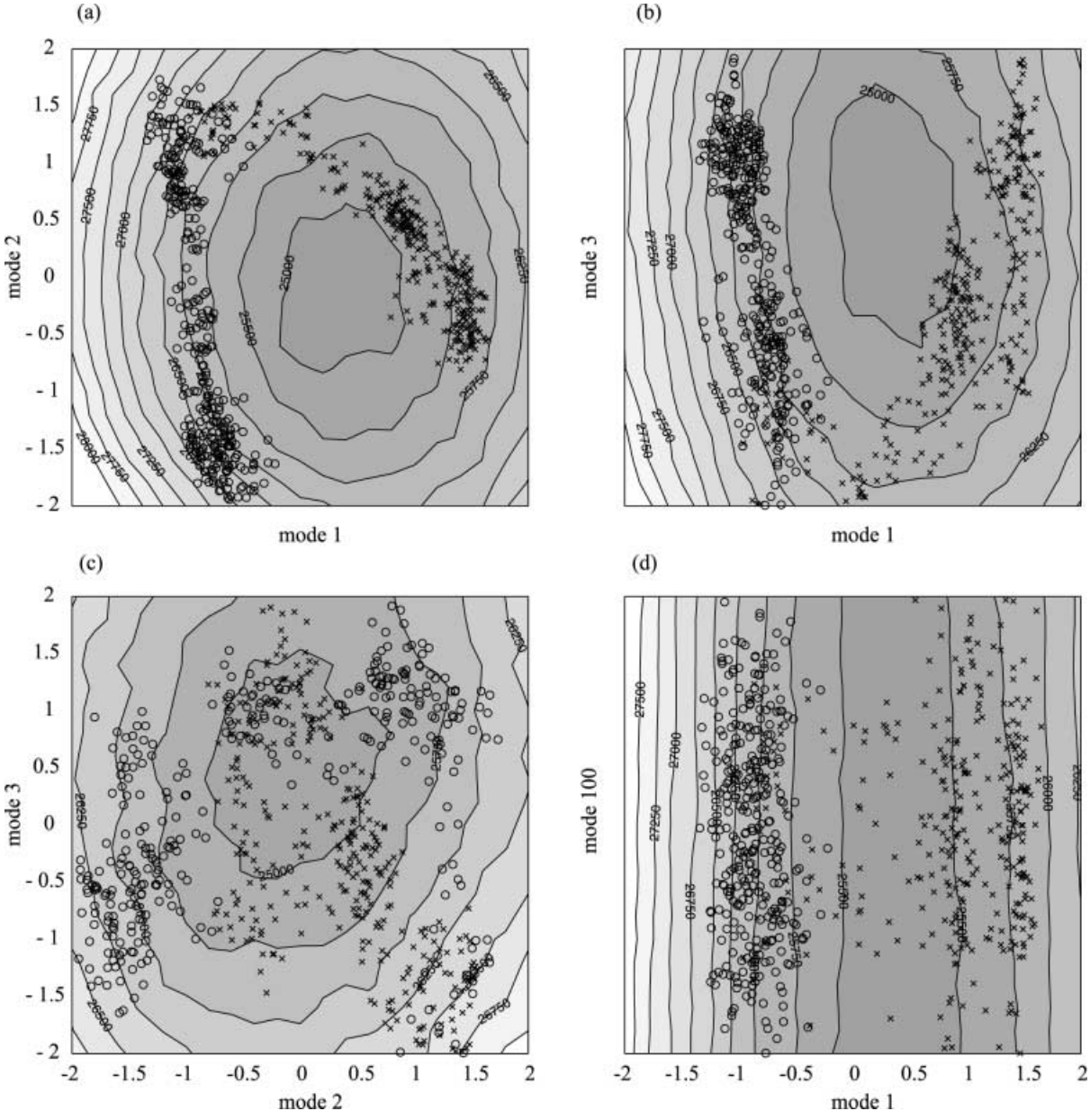

Fig. 3 Projected trajectory and volume profile onto two-dimensional subspaces spanned by a pair of principal modes: **a** mode 1 and mode 2, **b** mode 1 and mode 3, **c** mode 2 and mode 3, **d** mode 1 and mode 100. Projected trajectory is represented by × for 0–500 ps of the trajectory and ○ for 500 ps to 1 ns of the trajectory. *Abscissa* is scaled by r.m.s. fluctuation $\Lambda_m^{1/2}$ along each axis

obtained from molecular dynamics. This result justifies the above treatment, i.e., only the second derivatives of the first 10 modes are included in the calculation. In Fig. 5 the contribution of each mode to the first term (the second-order term) of the volume fluctuation is shown (only the first 500 modes are represented). The second-order term is related to the first derivative of the volume. The cumulative sum of these contributions is 272,547 $Å^6$. The first few modes have a dominant contribution to the volume fluctuation. The first mode presents the largest contribution, which is 43% of the total contribution of principal modes to the second-order term. This result is in agreement with what has been shown in a previous study (Kitao et al. 1998). In principal component analysis the first mode contributes 29% to the total mass-weighted mean-squared fluctuation.

The contribution of the third-order term to the volume fluctuation is negative, –189,913 $Å^6$. This contribution is again dominated by a contribution from very-low-frequency modes. Finally, for the fourth-order term a value of 49,988 $Å^6$ is found.

Conformational changes during the dynamics

We have seen that lysozyme shows a large volume fluctuation during the dynamics. We now look more closely at the structure of lysozyme during the dynamics

in order to identify conformational changes that involve such large volume changes. In Fig. 3a we have seen that the trajectory has undergone a transition from a small volume state to a large volume state in the two-dimensional space spanned by the principal modes 1 and 2. Therefore as representative small-volume and large-volume conformations we take the structure defined by $q_1 = 1$, $q_2 = 0.5$ and $q_1 = -1$, $q_2 = -1.5$, respectively.

In Fig. 6 we show the change in the accessible surface area for each residue and the displacement of C_α atom positions between the two structures. We see clearly a large increase of this surface area for residues located in the range 110–115. At the same time, large displacements are also observed in the same part of the protein. Residues 56–58 present also a relatively large change in their accessible surface area, but displacements of C_α atom positions are relatively small. Residues 47–51 show relatively large displacements but with small changes in their accessible surface area. Residue 47–51 and 110–115 are located in the edges of the cleft, explaining their relatively large displacements. When we look at the 3D representation (Fig. 7) of lysozyme, we see that the side chains of residue 110–115 are more exposed to the solvent in the large-volume structure. We see also that residues 47–51 are moving further apart from the rest of the protein. A study of lysozyme by normal mode analysis (Gibrat and Go 1990) has also revealed these motions.

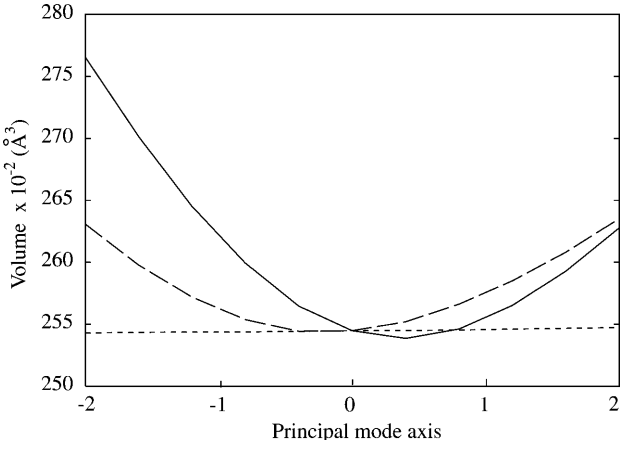

Fig. 4 Volume change is plotted against the scaled principal component variables. Change due to the first mode (*solid line*), second (*broken line*), and 100th (*dotted line*). *Abscissa* is scaled by r.m.s. fluctuation $\Lambda_m^{1/2}$ along each axis

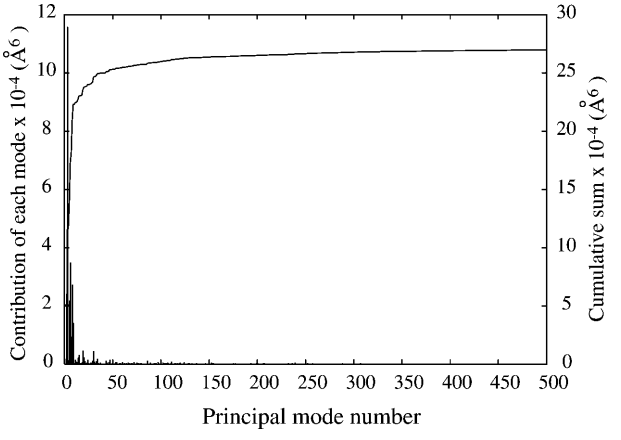

Fig. 5 Contribution (shown by *bars*) of each principal mode to volume fluctuation. Cumulative sum of these contributions is shown in the *solid line*

Discussion

Comparison with results from normal mode analysis

Previously, volume fluctuations in several different proteins have been studied by normal mode analysis (Yamato et al. 1993; Kobayashi et al. 1997). In these studies, a linear approximation of volume as a function

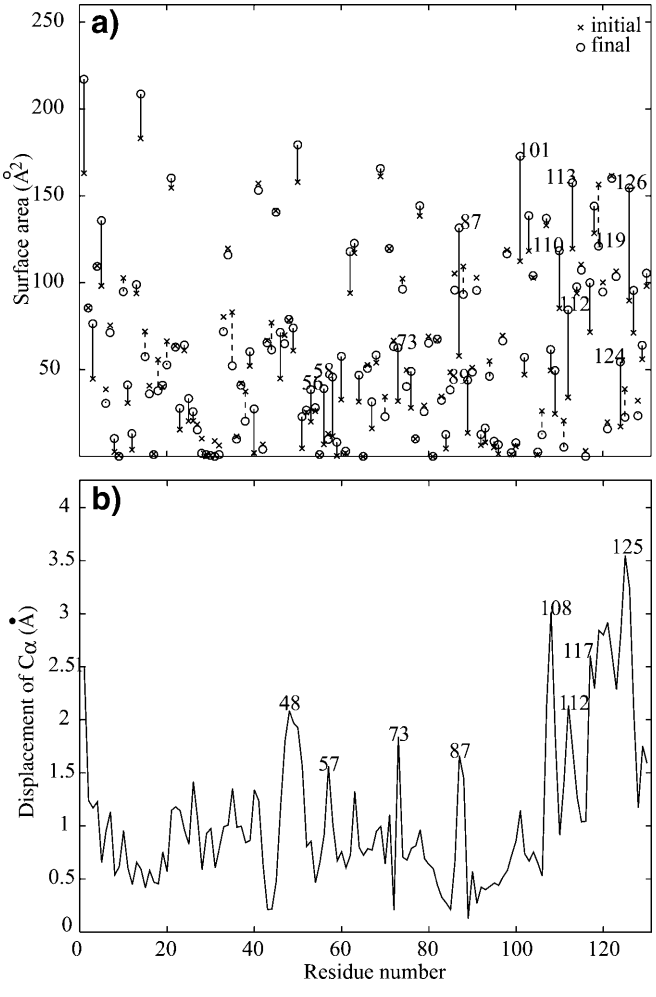

Fig. 6 **a** Change of the accessible surface area of each residue: (×) structure with a small volume, (○) structure with a large volume. *Solid line* indicates an increase of the surface area, *dotted line* a decrease. **b** Displacement of C_α atom positions between the two structures

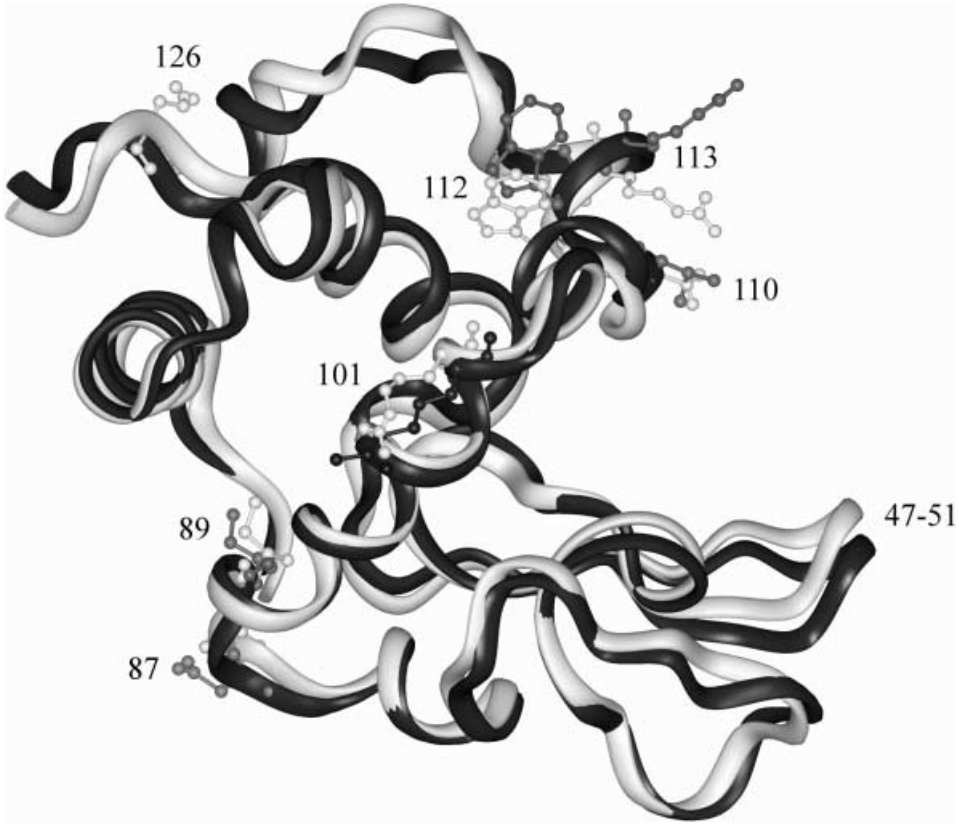

Fig. 7 Superposition of the two structures obtained by displacement along mode 1 and mode 2 (*gray*: small-volume structure; *black*: large-volume structure)

of a normal mode variable was made. Thus, the volume was expanded in powers of normal mode variables and approximated to the first order. If we make the same approximation in our principal component analysis and therefore retain only the first term (the second-order term) in Eq. 7, we obtain a value of 552 $Å^3$ for the r.m.s. volume fluctuation. This value differs considerably from the one obtained directly from the trajectory. This result indicates clearly that the normal mode analysis is not enough for the discussion of volume fluctuation. However, the spirit of the normal mode analysis can be revived for the treatment of volume fluctuation by extending it to the principal component analysis so as to be able to treat effects due to anharmonicities.

Timescale

The comparison of the power spectrum obtained from the MD (Fig. 2) and the contributions of various principal modes (Fig. 5) reveals the same characteristics, i.e., the low-frequency region contributes the most to the volume fluctuation. This result is in agreement with previous studies with normal mode analysis, where the low-frequency region was also found to be important for the volume fluctuation. However, in the principal component analysis, as well as in the power spectrum obtained from the dynamics, we see that the lowest frequency modes are more enhanced compared to the results observed in the normal mode analysis (Kobayashi et al. 1997).

A second point of discussion is the large value we obtained for the volume fluctuation. In fact, we see a big difference between our value of the volume fluctuation and the one obtained in a previous study (Paci and Marchi 1996). As already mentioned, this difference comes mainly from the difference of the time length of the trajectory used (see Fig. 1c). As is clear from Fig. 2c, in turn this difference comes mainly from contributions from the frequency region 0–0.3 cm^{-1}. This very-low-frequency region is related to slow motions, often involving transitions between different minima, which cannot be treated by the normal mode analysis. The motions due to the normal mode analysis simulate those taking place during the short timescale of the picosecond range. Both the 100 ps simulation and the normal mode analysis are not able to treat transitions between local minima and volume fluctuations associated with such transitions.

Comparison with experimental results

Experimentally, the value of the volume fluctuation is obtained by the measurement of isothermal intrinsic compressibility:

$$\beta_{\mathrm{T}} = \frac{1}{k_{\mathrm{B}} T V_0} \langle (\Delta V^2) \rangle \tag{10}$$

where V_0 is the total volume of the protein studied. Most of the available compressibility data and the related volume fluctuations come from sound velocity measurements (Gekko and Noguchi 1979; Gavish et al. 1983; Gekko and Hasegawa 1986; Kharakoz and Sarvazyan 1993). By sound velocity experiments, one obtains the adiabatic compressibility of the solution $\beta_{S,solution} = 1/(\rho u^2)$ where ρ is the density and u the sound velocity. From $\beta_{S,solution}$ one derives $\bar{\beta}_S$, which is the partial molar adiabatic compressibility of the protein. This compressibility is usually separated into two contributions: $\bar{\beta}_S = \beta_S + \Delta\beta_h$ which represent the intrinsic contribution and the hydrational contribution, respectively. To obtain the adiabatic intrinsic compressibility it is necessary to determine the contribution due to hydration. Two different ways are used to estimate this term (Kharakoz and Sarvazyan 1993). One is called the additive method, the second the regression method. Even though these two methods give similar results, they are both based on some strong approximations. From the adiabatic compressibility, $\bar{\beta}_S$, it possible to determine the isothermal compressibility, $\bar{\beta}_T$.

Sound velocity experiments (Gekko and Noguchi 1979; Gekko and Hasegawa 1986) report the isothermal compressibility $\bar{\beta}_T$ in the range of 1.92 to 15×10^{-6} bar^{-1}. This compressibility still includes the hydrational contribution. The authors suggest that the relative volume fluctuation would become about 50% larger when correction due to the hydration term is considered. Kharakoz and Sarvazyan (1993) reported the adiabatic intrinsic compressibility and the isothermal intrinsic compressibility for globular proteins to be in the range of $(13 \pm 3) \times 10^{-6}$ bar^{-1} and $(14 \pm 3) \times 10^{-6}$ bar^{-1}, respectively. From our result, i.e., $V_0 = 27{,}633$ Å^3 and $\langle \Delta V^2 \rangle = (352 \text{ Å}^3)^2$, we obtained β_T to be $10^8 \times 10^{-6}$ bar^{-1}. This is a value much larger than those estimated from experimental data.

This large difference forces us to seriously search for its origin. On the side of simulation, our analysis indicates that the conformational fluctuations such as shown in Fig. 7 are mainly responsible for the large volume fluctuation. There may be room to suspect that such conformational fluctuations are an artifact due to simulation. However, from all our experience we feel that they should be real. On the side of deducing the value for the intrinsic quantity from the experimental observable, one has to resort to some models of hydration which we feel are not established. In the light of the result we obtained in this paper, we propose that the approximations made for $\Delta\beta_h$ be re-examined and suggest that the intrinsic compressibilities of proteins are much larger than the one reported earlier.

Besides the most frequently used method for the sound velocity measurement, some other methods like fluorescence (Marden et al. 1986), hole burning (Zollfrank et al. 1991), and crystallography (Kundrot and Richards 1987) also allow the determination of the compressibility. In an experimental study of the heme protein by fluorescence, compressibility softer than that of water was found. This means that the compressibility of proteins measured by fluorescence is at least 10 times larger than the value reported by the sound velocity experiment. In the hole burning experiment the relative volume fluctuation reported for mesoporphyrin was 3%. In an NMR study of basic pancreatic trypsin inhibitor (BPTI) under a hydrostatic pressure of 2000 bar (Li et al. 1998), the lengths of hydrogen bonds were found to decrease on average by 1%. This decrease of lengths appears to suggest a compressibility larger than those observed by sound velocity. In all these experiments, other than the sound velocity measurement, pressure is applied as a hydrostatic pressure, while in the sound velocity measurement no such static pressure is applied.

So far, it appears that the measurements of sound velocity give small values of compressibility, while other methods performed under hydrostatic pressure give large values. One disturbing exception for this tendency is the result from crystallography. The crystallographic structure determination of lysozyme was measured (Kundrot and Richards 1987) first under atmospheric pressure and then under a hydrostatic pressure of 1000 bar. From the structures obtained they derived a value of compressibility three times smaller than that from the sound velocity measurements. This is a result in the opposite direction from those obtained by other methods carried out under hydrostatic pressure. Kharakoz and Sarvazyan (1993) discussed the difference in the values obtained by crystallography and by sound velocity. It appears that essentially they remain puzzled when they say that the difference may be due to the different nature of the methods used. We are also puzzled by the result from the crystallographic study. We would offer one mechanism as a possible reason for the small compressibility observed by crystallography. The crystallographic study is done first by preparing a crystal under atmospheric pressure and then applying the hydrostatic high pressure. Once a crystal is made, the conformational and therefore volume fluctuations of the protein molecules would be very much suppressed. Similar phenomena have been observed. Normal mode analysis agrees better than the MD simulation when the internal r.m.s. fluctuations of the residues are compared with the one obtained by the X-ray normal mode refinement method (Hayward and Go 1995). This result has been explained by the suppression of large-scale anharmonic motions by the crystalline environment. Because compressibility is related to volume fluctuation by Eq. 10, this implies that the compressibility of protein molecules become smaller once they are crystallized. The small value observed by Kundrot and Richards may be due to this mechanism.

Conclusion

We have studied volume fluctuations of human lysozyme during a 1 ns dynamics simulation by two different approaches. In the first, the volume is calculated directly

along the 1 ns MD trajectory. The obtained r.m.s. volume fluctuation is 352 $Å^3$. In the second one, volume fluctuation is derived from a model based on principal component analysis. The principal modes are obtained from the 1 ns trajectory. In the latter model, the volume is expressed as a function of the principal component variables and is expanded and approximated to the second order. The volume fluctuation obtained is 364 $Å^3$. Thus we obtained a good agreement between the two calculations for the volume fluctuation. From the principal component analysis we can identify a small number of conformational changes responsible for the large volume fluctuation, and therefore for the softness of the protein.

The result of volume calculations along the 1 ns trajectory shows that the r.m.s. volume fluctuation depends significantly on the time length of the trajectory. The relative volume fluctuation is 0.49% for a 100 ps trajectory and 1.27% for the 1 ns trajectory. This increase for the amount of volume fluctuation is associated with migration of the state point in the conformational space.

Volume fluctuation is related to compressibility. We point out, from our theoretical approach, that the intrinsic compressibility of proteins may be larger than that deduced from sound velocity measurements. We assume that the difference comes from the way of estimating the hydration term.

Acknowledgements F.T. gratefully acknowledges a travel fellowship from Paul Sabatier University. O.M. gratefully acknowledges a research fellowship from the Japan Society for the Promotion of Science (JSPS) for Young Scientists. This study has been supported by Grants-in Aid to A.K. and N.G. from the Ministry of Education, Japan.

References

Beglov D, Roux B (1994) Finite representation of an infinite bulk system: solvent boundary potential for computer simulations. J Chem Phys 100: 9050–9063

Ding HQ, Karasawa N, Goddard WA III (1992) Atomic level simulations on a million particles: the cell multipole method for Coulomb and London nonbond interactions. J Chem Phys 97: 4309–4315

Eisenhaber F, Lijnzaad P, Argos P, Sander C, Scharf M (1995) The double cubic lattice method: efficient approaches to numerical integration of surface area and volume and dot surface contouring of molecular assemblies. J Comput Chem 16: 273–284

Gavish B, Gratton E, Hardy CJ (1983) Adiabatic compressibility of globular proteins. Proc Natl Acad Sci USA 80: 750–754

Gekko K, Hasegawa Y (1986) Compressibility-structure relationship of globular proteins. Biochemistry 25: 6563–6571

Gekko K, Noguchi H (1979) Compressibility of globular protein in water at 25 °C. J Phys Chem 83: 2706–2714

Gibrat JF, Go N (1990) 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

Hayward S, Go N (1995) Collective variable description of native proteins dynamics. Annu Rev Phys Chem 46: 223–250

Hayward S, Kitao A, 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

Hoover WG (1985) Canonical dynamics: equilibrium phase-space distribution. Phys Rev A 31: 1695–1697

Jorgensen WL, Chandrasekhar J, Madura JD (1983) Comparison of simple potential functions for simulating liquid water. J Chem Phys 79: 926–935

Kharakoz DP, Sarvazyan AP (1993) Hydrational and intrinsic compressibilities of globular proteins. Biopolymers 33: 11–25

Kidera A, Inaka K, Matsushima M, Go N (1992) Normal mode refinement: crystallographic refinement of protein dynamic structure. II. Application to human lysozyme. J Mol Biol 225: 477–486

Kitao A, Hirata F, Go N (1991) The effects of solvent on the conformation and the collective motions of protein: normal mode analysis and molecular dynamics simulation of melittin in water and in vacuum. Chem Phys 158: 447–472

Kitao A, Hayward S, Go N (1998) Energy landscape of a native protein: jumping-among-minima model. Proteins 33: 496–517

Kobayashi N, Yamato T, Go N (1997) Mechanical property of a TIM-barrel protein. Proteins 28: 109–116

Kundrot CE, Richards FM (1987) Crystal structure of hen egg-white lysozyme at a hydrostatic pressure of 1000 atmospheres. J Mol Biol 193: 157–170

Li H, Yamada H, Akasaka K (1998) Effect of pressure on individual hydrogen bonds in proteins. Basic pancreatic trypsin inhibitor. Biochemistry 37: 1167–1173

Marden MC, Hui-Bon Hoa G, Stetzkowski-Marden F (1986) Heme protein fluorescence versus pressure. Biophys J 49: 619–627

Miyashita O, Go N (1999) Pressure dependence of protein electron transfer reactions: theory and simulation. J Phys Chem B 103: 562–571

Morikami K, Nakai T, Kidera A, Saito M, Nakamura H (1992) PRESTO (protein engineering simulator): a vectorized molecular mechanics program for biopolymers. Comput Chem 16: 243–248

Nosé S (1984) A molecular dynamics method for simulations in the canonical ensemble. Mol Phys 52: 255–268

Paci E, Marchi M (1996) Intrinsic compressibility and volume compression in solvated proteins by molecular dynamics simulation at high pressure. Proc Natl Acad Sci USA 93: 11609–11614

Paci E, Velikson B (1997) On the volume of macromolecules. Biopolymers 41: 785–797

Sugita K, Kitao A (1998) Improved protein free energy calculation by more accurate treatment of nonbonded energy: application to chymotrypsin inhibitor 2, v57a. Proteins 30: 388–400

Weiner SJ, Kollman PA, Nguyen DT, Case DA (1986) An all atom force field for simulations of proteins and nucleics acids. J Comput Chem 7: 230–252

Yamato T, Higo J, Seno Y, Go N (1993) Conformational deformation in deoxymyoglobin by hydrostatic pressure. Proteins 16: 327–340

Zollfrank J, Friedrich J, Fidy J, Vanderkooi JM (1991) Photochemical holes under pressure: compressibility and volume fluctuations of a protein. J Chem Phys 94: 8600–8603