Harmonic and anharmonic aspects in the dynamics of BPTI: A normal mode analysis and principal component analysis

ARTICLE *in* PROTEIN SCIENCE · JUNE 2008

Impact Factor: 2.85 · DOI: 10.1002/pro.5560030608 · Source: PubMed

3 AUTHORS:

Steven John Hayward
University of East Anglia
49 PUBLICATIONS **2,719** CITATIONS
SEE PROFILE

Akio Kitao
The University of Tokyo
95 PUBLICATIONS **2,546** CITATIONS
SEE PROFILE

Nobuhiro Go
Japan Atomic Energy Agency
238 PUBLICATIONS **9,917** CITATIONS
SEE PROFILE

Protein Science (1994), *3*:936-943. Cambridge University Press. Printed in the USA.

Harmonic and anharmonic aspects in the dynamics of BPTI: A normal mode analysis and principal component analysis

STEVEN HAYWARD, AKIO KITAO, AND NOBUHIRO GŌ

Department of Chemistry, Faculty of Science, Kyoto University, Kitashirakawa, Sakyo-ku, Kyoto 606, Japan

(RECEIVED January 20, 1994; ACCEPTED March 30, 1994)

Abstract

A comparison is made between a 200-ps molecular dynamics simulation in vacuum and a normal mode analysis on the protein bovine pancreatic trypsin inhibitor (BPTI) in order to elucidate the dual aspects of harmonicity and anharmonicity in the dynamics of proteins. The molecular dynamics trajectory is analyzed using principal component analysis, an effective harmonic analysis suited for comparison with the results from the normal mode analysis. The results suggest that the first principal component shows qualitatively different behavior from higher principal components and is associated with apparent barrier crossing events on an anharmonic conformational energy surface. The higher principal components appear to have probability distributions that are well approximated by Gaussians, indicating harmonicity. Eliminating the contribution from the first principal component reveals a great deal of correspondence between the 2 methods. This correspondence, however, involves a factor of 2, as the variances of the distributions of the higher principal components are, on average, roughly twice those found from the normal mode analysis. A model is proposed to reconcile these results with those from previous analyses.

Keywords: anharmonicity; collective motions; molecular dynamics; normal mode analysis

The study of protein dynamics is primarily motivated by the view that a protein's function is not determined solely by its static structure but also by the relative movement of the constituent parts of that structure. A number of experimental works support this view, not least the cryocrystallographic studies on ribonuclease A (Cusak, 1992; Rasmussen et al., 1992; Tilton et al., 1992) and the numerous X-ray studies of proteins in liganded and unliganded states (Schulz, 1991), where it appears that rather large-scale motions of protein domains as relatively rigid bodies occur upon the binding of ligands.

In order to gain a detailed understanding as to the nature of protein dynamics, computational methods are indispensable given the complexity of protein molecules. Of the computer methods used, molecular dynamics (McCammon & Harvey, 1987) and normal mode analysis (Brooks & Karplus, 1983; Gō et al., 1983; Levitt et al., 1983) have become the most popular. However, in their approach, these 2 methods are in fact quite far apart. Molecular dynamics, being the raw integration of Newton's equation, should give an accurate simulation of protein motion over a certain time interval but does not directly say anything of the higher physics involved in determining the nature of protein dynamics, leaving that to be deduced from the analysis of the collective trajectories of the protein atoms. On the other hand, the normal mode analysis method assumes from the outset that a harmonic analysis at a single conformational energy minimum can give relevant information on a protein's dynamics. Theoretically this is based upon the well-packedness of the atoms in the interior of the protein, giving it analogous properties to that of a crystal.

Experimentally, the strongest evidence for the accuracy of the normal mode analysis comes from the normal mode refinement method for protein crystallography (Kidera & Gō, 1990, 1992). Using this method, it is possible to separate the external and internal contributions to the Debye-Waller factors. For human lysozyme, a comparison between this internal part and the results from a normal mode analysis on that protein showed a quantitative agreement (Kidera et al., 1992), suggesting that a harmonic potential surface determined at a single energy minimum can indeed give accurate results for protein motion over the whole range of conformational fluctuations at a temperature of 300 K. Paradoxically, though, other experimental (Austin et al., 1975) and computational (Elber & Karplus, 1987; Noguti & Gō, 1989a, 1989b, 1989c, 1989d, 1989e) work has shown that a protein has a multiple-minima conformational en-

Reprint requests to: Nobuhiro Gō, Department of Chemistry, Faculty of Science, Kyoto University, Kitashirakawa, Sakyo-ku, Kyoto 606, Japan; e-mail: go@qchem.kuchem.kyoto-u.ac.jp.

ergy surface over the range of conformational fluctuations at 300 K. A careful comparison, therefore, between a normal mode analysis and a molecular dynamics simulation should help to shed light on these dual aspects. Such a comparison has been made possible by doing a principal component analysis of the molecular dynamics trajectory (Kitao et al., 1991). This method determines a new set of axes, in a multidimensional space, for the description of the distribution of points corresponding to atomic coordinates at each time step along the protein's trajectory. Collective motions occurring during a simulation can be effectively described by these new variables, termed principal component variables. Significant for our purposes though, is that in the case of a truly harmonic conformational energy surface, the principal component analysis is identical to the normal mode analysis provided the mean-square fluctuations have converged during the simulation. In the case of an anharmonic energy surface, the principal components can be thought of as the normal modes of a harmonic approximation to that surface, and, in analogy to normal mode analysis, one can define an effective frequency associated with each principal component.

Teeter and Case (1990), in their comparison between a normal mode analysis and a molecular dynamics simulation in vacuum on the small protein crambin, found a rather good correspondence between the cumulative frequency distribution determined from the normal mode analysis and the cumulative effective frequency density determined from a quasiharmonic analysis (equivalent to principal component analysis as described above, but no use was made of the eigenvectors) of the molecular dynamics trajectory. In addition, in comparing the RMS fluctuations (RMSF) of the residues, the 2 approaches agreed well, except over the floppy C-terminal region, where the normal mode analysis underpredicted the RMSF values of the residues.

For the polypeptide melittin (Kitao et al., 1991), where the same methods as here were used, the RMSF values for the alpha-carbon atoms derived from the normal mode analysis underpredicted those from the molecular dynamics trajectory in vacuum.

Despite these studies, the nature of a protein's conformational potential energy surface remains unclear. In this work, we make full use of principal component analysis in analyzing a 200-ps molecular dynamics trajectory in vacuum of the small globular protein bovine pancreatic trypsin inhibitor (BPTI) and compare the results with a normal mode analysis.

Materials and methods

Molecular dynamics simulation

The molecular dynamics (MD) simulation was carried out using PRESTO (Morikami et al., 1992) and was run on a Fujitsu VP2600 computer. Coordinates of BPTI were taken from 4PTI in the Protein Data Bank (Deisenhofer & Steigemann, 1975; Markquart et al., 1983). Amber-OPLS united atom parameter set (Weiner et al., 1986; Jorgensen & Tirado-Rives, 1988) was used. With this parameter set, BPTI has 568 "atoms." A cutoff was not applied to the nonbonded interactions and a relative dielectric constant proportional to the distance between the atoms was used. Hydrogen atoms not united with heavy atoms were constrained using SHAKE (Ryckaert et al., 1977). A temperature of 300 K was chosen, the temperature being maintained by the Berendsen method (Berendsen et al., 1984) using a relaxation time of 40 fs for the velocity rescaling. A time step of 2 fs was chosen for the integration of the Verlet algorithm. After a 30-ps equilibration period, the coordinate and velocity trajectories were collected every 12 fs for 200 ps. The first 100 ps of this trajectory were used in a previous work for comparison with a trajectory in water (Hayward et al., 1993).

Principal component analysis

Principal component analysis is performed by diagonalizing the variance–covariance matrix $\boldsymbol{r}$ of the mass-weighted internal displacements. This matrix is given by the expression

$$\boldsymbol{r} = \frac{1}{K}\,\boldsymbol{x}\boldsymbol{x}^{\mathrm{T}}, \tag{1}$$

where $\boldsymbol{x}$ is a matrix whose elements x_{it} give the mass-weighted internal displacements (Cartesian coordinate times square root of mass) of each atomic coordinate i at a given time step t, $\boldsymbol{x}^{\mathrm{T}}$ is the transpose of $\boldsymbol{x}$, and K is the total number of time steps, which was 8,333, i.e., every 24 fs over the whole 200-ps trajectory. The diagonal elements of $\boldsymbol{r}$ are the variances, and the off diagonal elements, the covariances. Diagonalization of $\boldsymbol{r}$ gives a diagonal eigenvalue matrix $\boldsymbol{R}$ and a matrix $\boldsymbol{U}$, the columns of which give the eigenvectors, or principal components. The matrix $\boldsymbol{U}$ is normalized to the identity matrix, $\boldsymbol{I}$. The elements of $\boldsymbol{R}$ are the variances or mean-square fluctuations of the principal components. The trajectory of the jth principal component X_{jt} is given by

$$X_{jt} = \sum_i x_{it} U_{ij}. \tag{2}$$

From the mean-square fluctuations of the principal components it is possible to define an effective frequency, ω_i^{eff}, of each principal component as

$$\omega_i^{eff} = \sqrt{\frac{k_B T}{R_i}}, \tag{3}$$

where R_i denotes the mean-square fluctuation of the ith principal component. In defining the effective frequency, one is determining the frequencies of harmonic modes that have the same mean-square fluctuations as the principal components. In the case of an anharmonic energy surface, the principal components can be thought of as the normal modes (see below) of a harmonic approximation to that surface. In the case of a truly harmonic conformational energy surface, the principal component analysis is identical to the normal mode analysis provided the mean-square fluctuations have converged.

If the conformational energy surface along a principal component is harmonic, then the probability density will be given by the Gaussian

$$p(X_i) = \frac{\omega_i^{eff}}{\sqrt{2\pi k_B T}} \exp\left(-\frac{\omega_i^{eff\,2} X_i^2}{2k_B T}\right). \tag{4}$$

Normal mode analysis

Normal mode analysis was performed in Cartesian coordinate space with the same parameter set used for the MD. The atomic

Cartesian coordinates are mass weighted in the same way as in the principal component analysis. The potential energy was minimized using PRESTO's conjugate gradient energy minimization routine. The Hessian matrix was then calculated and diagonalized to obtain the eigenvalues and the eigenvector matrix $\boldsymbol{W}$. The eigenvalues are the angular frequencies squared and so define the frequencies of oscillation of the normal modes. The normal modes themselves are given by the columns of $\boldsymbol{W}$, which is normalized to identity matrix. The mean-square fluctuation of the ith atomic coordinate is given by

$$\langle x_i^2 \rangle = \frac{k_B T}{m_i} \sum_j \left(\frac{\boldsymbol{W}_{ij}}{\omega_j} \right)^2. \tag{5}$$

Comparison of principal component and normal mode vectors

To quantify the similarity between the principal component vectors and the normal mode vectors, the inner products are calculated. To do this, first the average conformation from the MD and the minimized structure of the normal mode analysis are brought into their best-fit positions, reorientating the normal mode and principal component vectors accordingly. If $\boldsymbol{w}_i$ represents the ith normal mode vector and $\boldsymbol{u}_j$ represents the jth principal component vector, then their inner product is defined as

$$g_{ij} = \boldsymbol{w}_i \cdot \boldsymbol{u}_j. \tag{6}$$

Because both sets of vectors are orthonormal, the following condition must hold:

$$\sum_{j=1}^{3N-6} g_{ij}^2 = 1. \tag{7}$$

The quantities g_{ij}^2, then, quantify the similarity between the ith normal mode vector and the jth principal component.

Spectral density

All of the quantities defined above for analysis of the MD results are derived from the mean-square fluctuations of atomic fluctuations. Quantities determined from the normal mode analysis and the principal component analysis contain no information on the actual dynamics if the dynamics are not purely harmonic. The spectral density is defined as

$$F(\omega) = \frac{1}{2\pi k_B T} \sum_{i=1}^{3N} \lim_{t \to \infty} \frac{1}{2t} |D_i(\omega)|^2, \tag{8}$$

where

$$D_i(\omega) = \int_{-t}^{t} v_i(t) \exp(-i\omega t)\, dt \tag{9}$$

and $v_i(t)$ is the mass-weighted velocity of the ith atomic component at time t. $F(\omega)$ is a quantity that contains dynamics and, provided there is no external motion, is directly comparable to the frequency density determined from the normal mode analysis or effective frequency density from the principal component analysis.

Results

The total mass-weighted mean-square fluctuation (normalized by dividing by the total mass) calculated over the whole 200-ps MD trajectory has a value of 1.47 Å^2. Calculated over just the first 100 ps, the value is 1.03 Å^2. The normal mode analysis gives the considerably lower value of 0.40 Å^2 for this quantity. The principal components are arranged in order of their mean-square fluctuations, i.e., the first has the largest, the second the second largest, and so on. It has been shown previously (Kitao et al., 1991; Hayward et al., 1993) that the first few principal components make an overwhelming contribution to the total mean-square fluctuation in comparison to the higher principal components, i.e., over the first 100 ps, the first principal component contributes 23% to the total mass-weighted mean-square fluctuation, the second, 13%, the third, 8.6%; over the whole 200-ps trajectory, the first principal component alone makes the overwhelming contribution of 43%, compared to 7% for the second and 4.9% for the third. Figure 1, constructed using Equation 2, shows the 200-ps MD trajectory projected onto the plane defined by the first 2 principal components. Because the first and second principal components together contribute 50% to the total mass-weighted mean-square fluctuation of the protein, Figure 1 must give a fairly accurate representation of the character of the motion. The trajectory begins on the right-hand side of the figure and ends on the left-hand side. A clear clustering effect is visible, indicative of the expected multiple-minima nature of the conformational energy surface for a protein in vacuum. The cluster on the left-hand side of the figure corresponds to the trajectory from 130 to 200 ps. It is clear that the first principal component corresponds to the transitions visible in this figure.

Table 1 shows the values for the inner products, calculated using Equation 6, between the first 10 principal components and the first 10 normal mode vectors. Using Equation 7, one can calculate the extent to which each principal component can be expressed by a set of normal modes. The right-hand column of

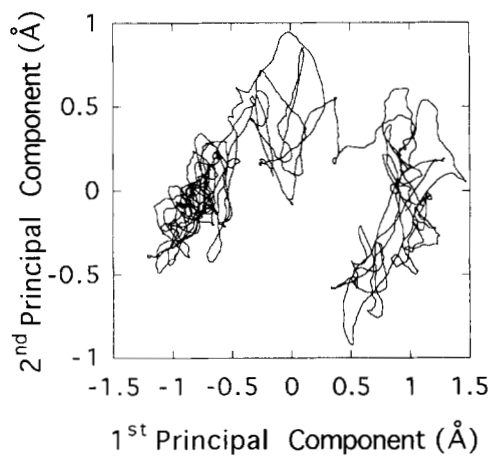

Fig. 1. Projection of 200-ps trajectory onto plane defined by first 2 principal components.

Table 1. *Inner products between first 10 normal modes (NM) and first 10 principal components (PC)*

	NM no.										
PC no.	1	2	3	4	5	6	7	8	9	10	Percentage[a]
1	−0.18	0.13	0.25	−0.08	0.27	0.17	0.25	0.09	0.38	0.08	30.1
2	−0.23	0.33	−0.34	0.02	0.30	0.02	−0.07	0.01	0.04	−0.10	39.4
3	0.30	0.27	−0.41	0.06	−0.16	0.04	0.09	−0.11	0.03	0.05	38.9
4	0.02	0.15	−0.15	−0.24	−0.06	0.07	−0.01	−0.04	−0.04	−0.05	11.3
5	0.40	−0.07	−0.06	0.35	−0.32	−0.06	0.20	0.05	0.01	−0.10	44.3
6	0.18	0.04	0.01	0.34	0.24	0.06	0.17	0.13	−0.15	0.07	28.8
7	0.01	0.39	0.00	0.06	0.05	−0.10	0.08	−0.34	0.13	−0.12	20.1
8	0.05	−0.03	0.06	0.05	0.01	0.16	−0.09	0.02	−0.03	−0.11	5.6
9	0.13	−0.04	0.00	0.10	0.26	−0.17	0.09	−0.06	−0.04	0.03	13.9
10	−0.07	0.13	−0.08	−0.03	−0.25	0.21	0.03	0.12	0.03	−0.07	16.5

[a] Degree to which each of the first 10 principal components can be expressed in the subspace of the first 10 normal modes.

Table 1 shows the degree to which each principal component can be expressed in the subspace defined by the first 10 normal mode vectors.

By taking their inner product, one finds that the first principal component derived from the whole 200-ps trajectory can be expressed to 66% by the first principal component derived from the first 100-ps trajectory. In the case of the second principal component, the corresponding value is 74%.

Figure 2 shows the probability densities for the first, second, third, fourth, 10th, and 100th principal components derived directly from the MD using Equation 2 in comparison with the Gaussian distributions derived using Equation 4. The first principal component shows peaks clearly associated with the clusters in Figure 1. Surprisingly, the second and higher principal components show considerable agreement with the Gaussian distributions, a result similar to that of Amadei et al. (1993) in their analysis of a 900-ps MD trajectory of lysozyme in water. Despite this general agreement, the character of the expected multiple-minima energy surface still features in these figures, e.g., the peak between −0.3 Å and 0 Å seen in the probability density for the second principal component can be identified with the cluster on the left-hand side of Figure 1.

Figure 3 shows the frequency of each mode plotted against mode number for both the principal components and normal modes. In this figure the scale for the normal mode frequency is a factor of 1.4 greater than the scale for the principal component frequency. One can see considerable agreement over the first 700 modes. However, one exception to this agreement occurs for the first modes, the ratio of the first normal mode frequency to the first principal component frequency being 4.0.

Using Equation 5, the RMSF values of the C^{α} atoms can be calculated from the normal mode analysis. Figure 4A shows these in comparison with the RMSF values for the C^{α} atoms calculated from the MD. One can see that the normal mode analysis predicts some of the features in the pattern of peaks and troughs in the MD RMSF values, but considerably underpredicts

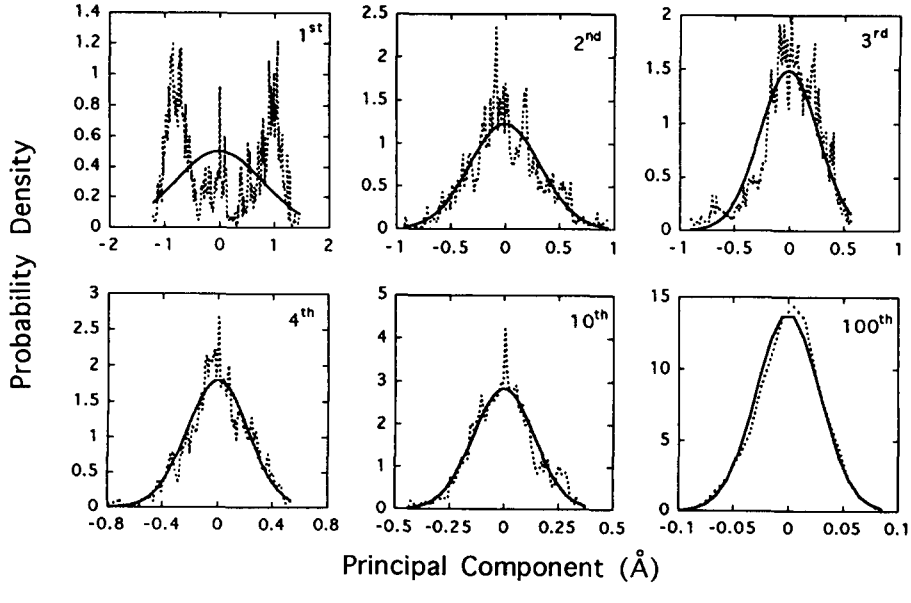

Fig. 2. Probability densities along first, second, third, fourth, 10th, and 100th principal components calculated directly from trajectory (broken line), in comparison with Gaussian distributions of same mean-square deviation as each principal component (see Equation 4; continuous line).

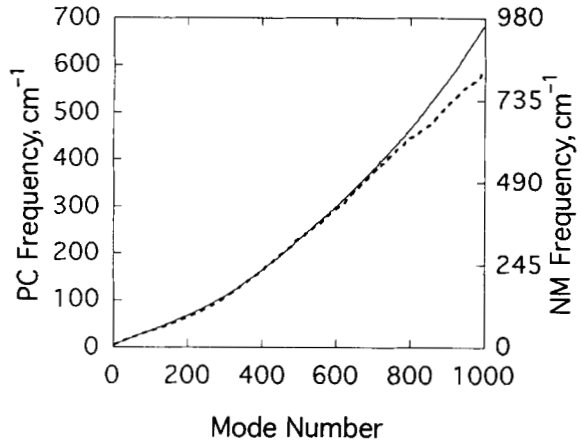

Fig. 3. Mode frequency plotted against mode number, for principal component (continuous line) and normal mode (broken line).

their actual values. Figure 4B shows the MD RMSF values of the C^{α} atoms in comparison with the contribution from just the first principal component. This shows that the transitions in the direction of the first principal component are alone sufficient to determine the overall pattern of peaks and troughs in the total RMSF values for the C^{α} atoms. Figure 4C shows the result from the normal mode analysis for the RMSF values of the C^{α} atoms, now multiplied by the factor of 1.4, in comparison with the contribution from all principal components excluding the first one. The good correspondence between the two is clearly visible.

Figure 5 shows the spectrum, calculated using Equation 8, over the whole 200-ps trajectory in comparison with the frequency density derived from the normal mode analysis and the effective frequency density from the principal component analysis. The spectral density showed no change when calculated over just the first 100 ps. In the previous paper (Hayward et al., 1993), the vacuum spectral density contained broad peaks centered on 120 cm^{-1} and 200 cm^{-1}. By performing the Fourier transforms using the IMSL fast Fourier transform routine on the velocity trajectory at 12-fs intervals instead of an original 18-fs interval, these peaks were seen to shift to frequencies higher than 600 cm^{-1}, the maximum frequency displayed in Figure 5, showing that they were aliased signals of motions with frequencies much higher than the sampling frequency. The normal mode analysis predicts the spectral density quite accurately. The effective frequency density calculated from the principal component analysis accurately predicts the upper extent of the spectral density. The fact that the frequency density from the normal mode analysis runs slightly below that of the effective frequency density from the principal component analysis can be explained by Figure 3. Because the frequency densities are given by the inverse derivative of the curves in Figure 3, the difference in scale means that, for a given frequency interval, there are fewer normal modes than principal components. In addition, the peaks in Figure 5 are a result of a minimum in the derivative in the curves of Figure 3 at approximately the 100th mode. The difference in scale, then, explains why the peak in the normal mode frequency density occurs at a slightly higher frequency than for the effective frequency density.

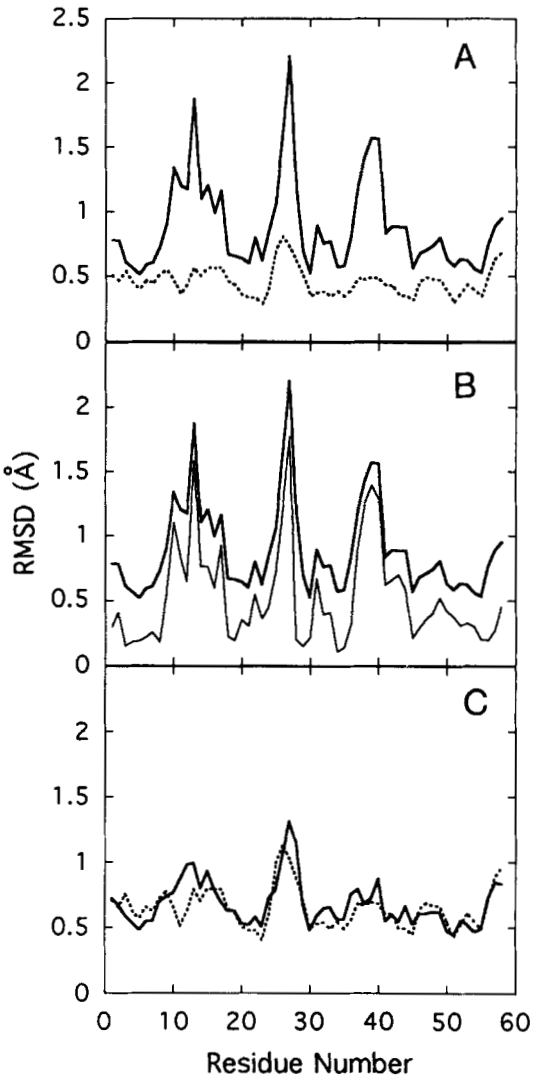

Fig. 4. RMSF values of C^{α} atoms plotted against residue number. **A:** Total RMSF of C^{α} atoms (continuous line), in comparison with result from normal mode analysis (broken line). **B:** Total RMSF of C^{α} atoms (thick line), in comparison with contribution from first principal component only (thin line). **C:** Contribution from all principal components excluding the first one (continuous line), in comparison with results from normal mode analysis multiplied by the factor 1.4 (broken line).

Discussion

The dual aspects of harmonicity and anharmonicity found in previous experimental and computational works on protein dynamics are clearly demonstrated in this work. The anharmonicity is visible in the clustering effect in Figure 1, indicative of the expected multiple-minima nature of the potential energy surface. The harmonic aspect is demonstrated in the correspondence between the results of the normal mode analysis and principal component analysis as shown in Figures 3, 4C, and 5. Interestingly, the anharmonic aspect appears to be embodied to a greater extent in the first principal component than any other. In fact, the first principal component, which clearly represents the transitions between minima (see Fig. 1) and makes, in comparison to

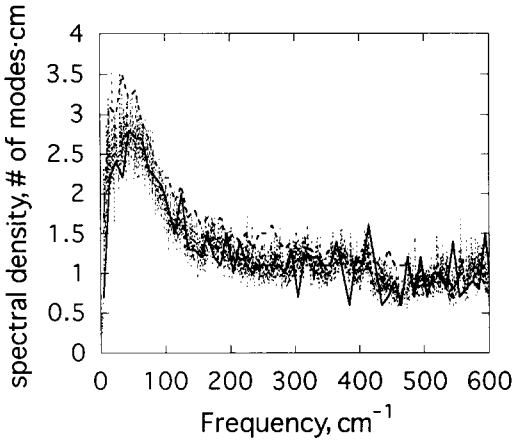

Fig. 5. Spectral density calculated using Equation 8 (fine broken line), in comparison with effective frequency density from principal component analysis (thick broken line) and frequency density from normal mode analysis (continuous line).

other principal components, an overwhelmingly large contribution to the total mean-square fluctuation (see Fig. 4B), displays behavior qualitatively different from all other principal components. This is obvious in Figure 2, where it is the only principal component not well approximated by a Gaussian. It is also demonstrated by the degree of correspondence between the principal component analysis and the normal mode analysis (Figs. 3, 4C), found only when the first principal component is excluded. This correspondence between the normal mode analysis and the principal component analysis is not exact, however, but involves a factor of 1.4, i.e., the RMSF values for the C^{α} atoms from the principal component analysis are, on average, a factor of 1.4 greater than their corresponding values from the normal mode analysis. Kitao et al. (1991), in their comparison of a normal mode analysis with MD simulations on melittin, also found that the RMSF values of the C^{α} atoms from the MD simulation in vacuum were a factor of about 1.4 greater than those from the normal mode analysis. However, in the work on crambin by Teeter and Case (1990), where a comparison between a normal mode analysis and a 100-ps vacuum MD simulation was made, a greater degree of correspondence than here was seen for the RMSF values of the residues (averaged over backbone heavy atoms), although, over the final 8 residues, in the floppy loop region, the values from the MD simulation also exceed those from the normal mode by about the same factor.

These results appear to be in contradiction with those from the normal mode refinement method applied to lysozyme (Kidera et al., 1992). There, a comparison of the RMSF values of the residues determined from the X-ray diffraction experiment showed a quantitative agreement with the normal mode analysis results on that protein. Looking at Figure 4A, we see that here no such agreement exists. However, there are 2 essential differences between this study and that one. In the case of the normal mode refinement experiment, lysozyme was, first, in a crystal environment and, second, solvated by water. As for the latter, in a previous paper, a comparison between an MD simulation of BPTI in water and in vacuum was made (Hayward et al., 1993). The difference between the water and vacuum results found there is not sufficient to account for the discrepancy found here. Below we make the argument that if one takes the crystal environment into account, one may then expect to find a good agreement between the results from the normal mode analysis and those from an MD simulation.

Let us assume that, for proteins in general, the unique motion represented by only the first principal component in the case of BPTI is actually represented by a small number of the lowest frequency principal components. Let us call these modes the anharmonic modes and all others the harmonic modes.

In a crystal environment, large-scale motions, such as those represented by the anharmonic modes, are expected to be severely restricted. Evidence for this comes from the NMR experiments on the calcium binding protein calmodulin (Ikura et al., 1992). Calmodulin is very flexible in solution, undergoing extremely large-scale bending hinged at the center of the long central helix. In the crystal, however, this helix is rigid. Further evidence for the effect of the crystal environment on collective motions can be found in the work on hemoglobin crystals, where no cooperativity was found upon the binding of oxygen, suggesting that the concerted conformational changes in structure are unable to occur in the crystal (Mozzarelli et al., 1991). We assume, then, that the motion represented by the anharmonic modes is frozen when a protein is in its crystal environment. In the case of BPTI, we expect the motion represented by the first principal component only to be frozen. However, directly removing the contribution from the first principal component, in the case of BPTI, still leaves the factor of 1.4 between the RMSF values from the normal mode analysis and those from the MD simulation to be accounted for. This discrepancy could be explained by a coupling between modes as found in the work by Horiuchi and Gō (1991) on lysozyme. There, the trajectories of a Monte Carlo simulation and a vacuum MD simulation were projected onto the normal modes to determine the extent to which the normal mode analysis predicted the actual RMSF values. It was found that, for some modes, the normal mode analysis underpredicted their true RMSF values. This was explained by a coupling between the higher and lower modes, whereby movement along the lower mode shifted the minimum point of the potential energy along the direction of the higher mode (see Fig. 6A). However, the picture here is more accurately displayed by Figure 6B. Motion in the direction of the anharmonic modes occurs as a series of barrier crossing events that are relatively weakly coupled to the harmonic modes. Given such a coupling, the freezing of the motion along the anharmonic modes would reduce the RMSF values along all other modes. It is suggested, therefore, that in the crystal environment, the freezing of the motion along the anharmonic modes brings the results into agreement with those from the normal mode analysis. We are currently investigating this possibility.

Although normal mode analysis considerably underpredicts the MD RMSF values, as seen in Figure 4A, the direction of the transition as described by the first principal component occurs to a significant extent in the subspace described by the first 10 normal modes, a result also found in a comparison of a normal mode analysis with MD simulations in water and vacuum on the polypeptide melittin (Kitao et al., 1991). These results support the concept of an important subspace defined by a subset of low frequency normal modes used in normal mode refinement of X-ray data.

One important question concerns the length of the simulation and how our results would have differed had we done, for ex-

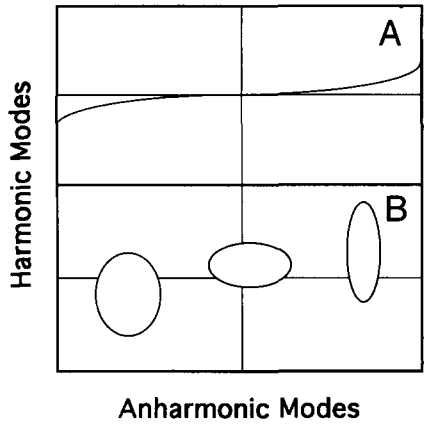

Fig. 6. A: Each point on the curved line represents the minimum in potential energy along the direction of the higher mode variable (here labeled as "Harmonic Modes" variable) for a given value of the lower mode variable (here labeled as "Anharmonic Modes" variable; from Horiuchi and Gō [1991]). **B:** Picture developed here. Ellipses represent potential energy minima, well separated in direction of anharmonic modes. Due to coupling, barrier crossing in the direction of the anharmonic modes causes a slight increase in the RMSF values of the harmonic modes.

ample, a 1-ns simulation. Apart from the first principal component, the plots in Figure 2 indicate that, for these modes, convergence has been achieved. However, in a longer simulation new transitions are expected to occur. Here the transition in the final 100 ps occurred largely in the same direction as the transition in the first 100 ps of the simulation, suggesting the possibility that the results would not change qualitatively for a longer simulation.

One further important question is to what extent the results would have been affected if the simulation had been performed in water. In a previous work (Hayward et al., 1993), a comparison between an MD simulation in vacuum and in water was made. The results of that work show that many of the plots presented here would have been affected by the presence of a water solvent. However, the differences were explained by treating the principal components as harmonic oscillators, the solvent influencing the dynamics through the friction term in the Langevin equation for a damped harmonic oscillator. If that model is correct, then none of the main conclusions would be affected by the presence of a water solvent.

The picture emerging for protein dynamics is one in which a small number of variables are necessary to describe large-scale motions of possible functional significance. In this work a single variable appears to be unique in describing the motion in BPTI and represents apparent barrier crossing of the multiple-minima conformational energy surface. The motion in the direction of this variable obscures the actual correspondence with the normal mode analysis results. It is thought that this motion is frozen in the crystal environment and the correspondence then becomes exact.

Acknowledgments

Computations were done in computer centers at Kyoto University and at the Institute of Molecular Science. This work was supported by grants from the Ministry of Education, Science and Culture, Japan, and from the International Human Frontier Science Program. S.H. is a recipient of a European Community Science and Technology Program Fellowship for Japan.

References

Amadei A, Linssen ABM, Berendsen HJC. 1993. Essential dynamics of proteins. *Proteins Struct Funct Genet 17*:412–425.

Austin RH, Beeson KW, Eisenstein L, Frauenfelder H, Gunsalus IC. 1975. Dynamics of ligand binding to myoglobin. *Biochemistry 14*:5355–5373.

Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, Haak JR. 1984. Molecular dynamics with coupling to an external bath. *J Chem Phys 81*:3684–3690.

Brooks B, Karplus M. 1983. Harmonic dynamics of proteins: Normal modes and fluctuations in bovine pancreatic trypsin inhibitor. *Proc Natl Acad Sci USA 80*:6571–6575.

Cusak S. 1992. Frozen stiff. *Curr Opin Struct Biol 2*:411–413.

Deisenhofer J, Steigemann W. 1975. Crystallographic refinement of the structure of bovine pancreatic trypsin inhibitor at 1.5 Å resolution. *Acta Crystallogr B 31*:238–250.

Elber R, Karplus M. 1987. Multiple conformational states of proteins: A molecular dynamics analysis of myoglobin. *Science 235*:318–321.

Gō N, Noguti T, Nishikawa T. 1983. Dynamics of a small protein in terms of low-frequency vibrational modes. *Proc Natl Acad Sci USA 80*: 3696–3700.

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

Horiuchi T, Gō N. 1991. Projection of Monte Carlo and molecular dynamics trajectories onto normal mode axes: Human lysozyme. *Proteins Struct Funct Genet 10*:106–116.

Ikura M, Clore MC, Gronenborn AM, Zhu G, Klee CB, Bax A. 1992. Solution structure of a calmodulin-target peptide complex by multidimensional NMR. *Science 256*:632–638.

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

Kidera A, Gō N. 1990. Refinement of protein dynamic structure: Normal mode refinement. *Proc Natl Acad Sci USA 87*:3718–3722.

Kidera A, Gō N. 1992. Normal mode refinement: Crystallographic refinement of protein dynamic structure. I. Theory and test by simulated diffraction data. *J Mol Biol 225*:457–475.

Kidera A, Inaka K, Matsushima M, Gō 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, Gō N. 1991. The effects of solvent on the conformational and collective motions of protein: Normal mode analysis and molecular dynamics simulations of melittin in water and in vacuum. *Chem Phys 158*:447–472.

Levitt M, Sander C, Stern PS. 1983. The normal modes of a protein: Native bovine pancreatic trypsin inhibitor. *Int J Quant Chem 10*:181–199.

Markquart M, Walter J, Diesenhofer J, Bode W, Huber R. 1983. The geometry of the reactive site and of the peptide groups in trypsin, trypsinogen and its complexes with inhibitors. *Acta Crystallogr B 39*:480–490.

McCammon JA, Harvey SC. 1987. *Dynamics of proteins and nucleic acids*. Cambridge, UK: Cambridge University Press.

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.

Mozzarelli A, Rivetti C, Rossi GL, Henry ER, Eaton WA. 1991. Crystals of haemoglobin with the T quarternary structure bind oxygen noncooperatively with no Bohr effect. *Nature 351*:416–419.

Noguti T, Gō N. 1989a. Structural basis of hierarchical multiple substates of a protein. I: Introduction. *Proteins Struct Funct Genet 5*:97–103.

Noguti T, Gō N. 1989b. Structural basis of hierarchical multiple substates of a protein. II: Monte Carlo simulation of native thermal fluctuations and energy minimization. *Proteins Struct Funct Genet 5*:104–112.

Noguti T, Gō N. 1989c. Structural basis of hierarchical multiple substates of a protein. III: Side chain and main chain local conformations. *Proteins Struct Funct Genet 5*:113–124.

Noguti T, Gō N. 1989d. Structural basis of hierarchical multiple substates of a protein. IV: Rearrangements in atom packing and local deformations. *Proteins Struct Funct Genet 5*:125–131.

Noguti T, Gō N. 1989e. Structural basis of hierarchical multiple substates of a protein. V: Nonlocal deformations. *Proteins Struct Funct Genet 5*:132–138.

Rasmussen BF, Stock AM, Ringe D, Petsko GA. 1992. Crystalline ribonuclease A loses function below the dynamical transition at 220 K. *Nature 357*:423-424.

Ryckaert JP, Ciccotti G, Berendsen HJC. 1977. Numerical integration of the cartesian equations of motion of a system with constraints: Molecular dynamics of *n*-alkanes. *J Comput Phys 23*:327-341.

Schulz GE. 1991. Domain motions in proteins. *Curr Opin Struct Biol 1*: 883-888.

Teeter MM, Case DA. 1990. Harmonic and quasiharmonic descriptions of crambin. *J Phys Chem 94*:8091-8097.

Tilton RF, Dewan JC, Petsko GA. 1992. The effects of temperature on protein structure and dynamics—X-ray crystallographic studies of the protein ribonuclease A at nine temperatures from 98 K to 320 K. *Biochemistry 31*:2469-2481.

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