PROTEINS: Structure, Function, and Genetics 11:205–217 (1991)

Collective Motions in Proteins: A Covariance Analysis of Atomic Fluctuations in Molecular Dynamics and Normal Mode Simulations

Toshiko Ichiye and Martin Karplus
Department of Chemistry, Harvard University, Cambridge, Massachusetts 02138

ABSTRACT **A method is described for identifying collective motions in proteins from molecular dynamics trajectories or normal mode simulations. The method makes use of the covariances of atomic positional fluctuations. It is illustrated by an analysis of the bovine pancreatic trypsin inhibitor. Comparison of the covariance and cross-correlation matrices shows that the relative motions have many similar features in the different simulations. Many regions of the protein, especially regions of secondary structure, move in a correlated manner. Anharmonic effects, which are included in the molecular dynamics simulations but not in the normal analysis, are of some importance in determining the larger scale collective motions, but not the more local fluctuations. Comparisons of molecular dynamics simulations in the present and absence of solvent indicate that the environment is of significance for the long-range motions.**

Key words: molecular dynamics, normal modes, collective motions

INTRODUCTION

A knowledge of the correlation of the atomic motions in proteins is essential to our understanding of their biological functions, including substrate–enzyme interactions and oxygen binding in hemoglobin.[1,2] It is important, for example, to be able to identify low-frequency, collective motions, such as hinge bending, and to determine the influence of neighboring atoms on the dynamics of active site residues. The magnitudes of fluctuations about the average positions of atoms have been estimated from X-ray diffraction studies of temperature factors[3,4] as well as molecular dynamics simulations.[1,2] The simulation studies, which covered time scales of 20 to 100 psec, are in general agreement with X-ray results, indicating that the observed temperature factors are mainly due to fluctuations within the time scales of the simulations. In fact, simulation methods have shown that the dominant contributions to the calculated rms fluctuations occur on a picosecond time scale and experimental techniques such as nuclear magnetic resonance, fluorescence depolarization, Mössbauer and inelastic neutron spectroscopy have confirmed that such high-frequency fluctuations do occur in proteins.

A detailed analysis of the time dependence of the atomic fluctuations in the bovine pancreatic trypsin inhibitor (BPTI) has shown that they are composed of subpicosecond local fluctuations and more collective motions of 1 to 10 psec, or longer.[5] If the fluctuations of an atom are large, the collective motions make the major contribution. It was suggested that atoms with similar long time behavior belong to the same collective group. This suggestion has been verified[6,7] by examining the time-correlation functions of the center-of-mass of atoms with similar atomic autocorrelation functions. Normal mode analyses of BPTI[8–10] have shown that the atomic fluctuations are dominated by delocalized low frequency modes.

In this paper we focus on the correlations of atomic flutuations, particularly those arising from collective motions. The analysis of equal-time covariances and cross-correlations (normalized covariances) of atomic fluctuations provides information about the correlation of atomic motions which complements the analysis of time-correlation functions. (In what follows, correlations will refer to equal-time correlations unless otherwise specified.) The covariance and cross-correlation matrices also provide a more detailed means of comparing different simulations than do the fluctuations (i.e., the diagonal elements of the covariance matrix) because the correlations as well as the magnitudes of the atomic motions are included. In molecular dynamics simulations, the covariances are essential for entropy calculations[11,12] and may be used for determining quasi-harmonic modes.[12,13] Correlated motions are also of interest for the interpretation of diffuse scattering.[14]

We report here a study of covariances and cross-correlations in three molecular dynamics simula-

Received March 14, 1991; accepted for publication March 14, 1991.

Address reprint requests to Dr. Martin Karplus, Department of Chemistry, Harvard University, 12 Oxford Street, Cambridge, MA 02138.

Present address of Toshiko Ichiye: Department of Biochemistry/Biophysics, Washington State University, Pullman, WA 99164.

tions and a normal mode analysis of BPTI. The molecular dynamics simulations are for BPTI in vacuum, in a van der Waals solvent composed of atoms with the size and particle density of water, and in a static crystal field.[15] The first two were used in previous studies of time correlation functions.[5-7] The normal mode calculation is of BPTI in vacuum.[8] A study of the cross-correlations of the low-frequency normal modes and quasiharmonic modes (calculated from the solvent simulation) and their relation to the molecular dynamics simulations is presented elsewhere.[6,7]

The second section outlines the methods employed in the simulations and describes the approaches used in the analysis. The results are presented in the third section. A discussion of the results is then given, followed by a concluding section.

METHODS

The methodologies used for the molecular dynamics and normal modes analyses have been described previously.[1] The molecule studied here is the bovine pancreatic trypsin inhibitor (BPTI), which has 454 heavy atoms plus 4 internal waters. Details of the simulations have been presented (see references below), so we describe them only briefly here.

The three molecular dynamics simulations analyzed here are of BPTI in vacuo (LC), in a van der Waals solvent (LCS), and in a fixed crystal field (LCC).[15] In all of the simulations, the hydrogen atoms of the protein and internal waters are treated as part of the heavy atom to which they are bonded by use of the extended atoms model and all bond lengths are constrained by means of the SHAKE algorithm.[16,17] This reduces the number of degrees of freedom in the simulation and eliminates the high-frequency bond-length vibrations which have little effect on the collective motions.[17] For the simulation of BPTI in a solvent environment, the protein was surrounded by 2647 van der Waals particles with radii and well depths corresponding to those in ST2 water,[18] using periodic boundary conditions. For the simulation of BPTI in a crystal field, the protein was surrounded by 2120 crystal image particles which were kept fixed at the original crystal coordinates. To analyze the atomic displacements in a molecular frame, the effects of overall translation and rotation of the molecule were removed from the atomic coordinates of the LCS[5] and LCC[19] simulations, by rotating the trajectory with respect to the backbone atoms. The analysis of the three molecular dynamics simulations covered a time period of 25 psec; the coordinates of the LC, LCS, and LCC trajectories were sampled at 0.04 psec intervals. The average temperature, as measured by the kinetic energy of the protein, was equal to 305.2 K in the vacuum run, 305.7 K in the solvent run and 305.2 K in the crystal run.

For the normal modes (NM)[8] analyzed here, the second derivatives of the full potential were calculated from the potential energy function for a defined X-ray crystal structure. This approach does not take into account effects of anharmonicity or the environment. The hydrogen atoms which are polar and capable of forming hydrogen bonds were treated explicitly (122 protein and internal water hydrogens) while the nonpolar hydrogens were treated by the extended atom model. Since the total number of atoms is 580 and the bond lengths were not constrained, there are 1740 degrees of freedom. The calculations used the 287 lowest nonnegative internal vibrational modes, which cover the frequency range between 3.1 and 148 cm^{-1} and correspond to oscillation periods of 11 to 0.22 psec. The choice of modes was based on the fact that the atomic fluctuations are dominated by 50 lowest modes and that a test calculation showed that there is very little difference in the covariances calculated from 287 modes and 887 modes. Because the protein was constrained to remain very near the X-ray structure, there are a number of negative modes in the unconstrained system. They have been shown to be highly localized and to correspond to high frequencies; consequently, they are not included.[8]

The temperature for the normal modes is a multiplicative factor determining the magnitude of the mean-square fluctuations and was set equal to 300 K. The cross-correlation matrix is thus independent of temperature, while the covariance matrix elements varied linearly with the temperature.

In the present analysis, the equal-time covariances and cross-correlations of the atomic fluctuations obtained from the various simulations were examined. For the displacement vectors $\Delta\mathbf{r}_i$ and $\Delta\mathbf{r}_j$ for atoms i and j, respectively, the covariance $c(i,j)$ is given by:

$$c(i,j) = <\Delta\mathbf{r}_i \cdot \Delta\mathbf{r}_j> \quad (1)$$

in which the angle brackets denote an ensemble average. The diagonal elements of the covariance matrix ($i=j$) are the mean-square (MS) atomic fluctuation, $<\Delta\mathbf{r}^{2_i}>$. The cross-correlation, or normalized covariance, is given by:

$$C(i,j) = \frac{c(i,j)}{[c(i,i)c(j,j)]^{1/2}} = \frac{<\Delta\mathbf{r}_i \cdot \Delta\mathbf{r}_j>}{<\Delta\mathbf{r}_i^2>^{1/2}<\Delta\mathbf{r}_j^2>^{1/2}} \cdot \quad (2)$$

For completely correlated motions, $C(i,j) = 1$, and, for completely anticorrelated motions, $C(i,j) = -1$. Complete correlation means the motions have the same phase as well as the same period. Deviations from 1 (or -1) imply either that the motions of i and j are less correlated (or anti-correlated) or that they deviate from motion along a straight line. An example of the latter is that if two atoms have fluctuations of the same period and phase but the displacements are oriented at an angle of 90°, $C(i,j) = 0$. Thus, examining $C(i,j)$ is useful for identifying

which atoms make up a group with correlated motions. However, $C(i,j)$ will identify the correlated relative motion of groups only if the angle between the directions of motion of the two groups is close to 0° or to 180°. A spherical breathing mode is difficult to identify from the cross-correlations because they are positive for atoms on the same side, negative for atoms on opposite sides, and 0 for atoms at 90°.

The covariances are calculated from the molecular dynamics trajectories by estimating the ensemble average over a set of discrete time points t_n ($n = 1,2, \ldots ,N$); that is, for the correlation of two atoms i and j,

$$c(i,j) = \frac{1}{N} \sum_{n=1}^{N} \Delta \mathbf{r}_i(t_n) \cdot \Delta \mathbf{r}_i(t_n) \tag{3}$$

where $\Delta \mathbf{r}_i(t_n)$ is the displacement of atom i from its average position at time t_n. For the normal modes analysis, the covariances are calculated by the expression (Levy et al., 1984):

$$c(i,j) = k_B T \sum_{k=1}^{K} \frac{\mathbf{a}_i(k) \cdot \mathbf{a}_j(k)}{\omega(k)^2} \tag{4}$$

in which $\mathbf{a}_i(k)$ is the vector of the projection of the kth normal mode with frequency $\omega(k)$ on the Cartesian components of the displacement vector for atom i, K is the number of normal modes included in the calculation, k_B is the Boltzmann constant, and T is the absolute temperature.

The expressions given in Eqs. (3) and (4) are for covariances of atoms i and j. To examine the covariances of *residues* i and j in the results presented below, we replace the atomic displacements $\Delta \mathbf{r}_j$ and projections $\mathbf{a}_i$ with the average vector of the appropriate atoms for a given residue. For backbone averages, the atoms included are N, C_α, and C, and for the side chain averages, all other atoms except the carbonyl O are included.

RESULTS

We have calculated cross-correlations of atomic fluctuations for each of the four simulations of BPTI. Since the LCS run is thought to present the most realistic simulation (see Discussion), we present a detailed analysis of its behaivor. Results for the other simulations are described when they differ from the LCS.

Correlated Motions and Their Analysis

To indicate the range of the correlated motions, we plot the cross-correlations $C(i,j)$ for the backbone atoms of the residues as a function of the distance $r(i,j)$ in Figure 1a–d. For the LCS, LC, LCC, and NM simulations, the results are similar, in that the cross-correlations decrease to zero at about 10 Å, go negative with a minimum in the neighborhood of 15 Å and increase to zero at about 25 Å. In all cases there is considerable scatter in the values for $C(i,j)$ at any given distance $r(i,j)$; the scatter is particularly marked in the LCS and LC simulations. The average value goes to zero between 8 and 10 Å; this is to be compared with a value of about 6Å estimated from the diffuse scattering by insulin.[14] Most of the nearest neighbor correlations range from about 0.4 to 0.9 in all of the simulations; exceptions are some values between 0 and 0.4 for the LCS simulation. The LCS simulation differs from the others in that the cross-correlations become positive for larger distances, reaching values of 0.5 at the maximum backbone distance of 30 Å. There is an overall reduction of long-range correlation in the NM and LCC calculations.

From Figure 1 it is clear that there is come correlation among the motions of widely separated regions of the molecule. To obtain a more detailed representation that permits easy identification of the regions in which there are significant correlations, we present a three-dimensional map of the cross-correlations between residue pairs in Figure 2a for the LCS simulation; the contour maps for the four simulations are shown in Figure 3a–d. The abscissa and the ordinate are residue numbers, i and j, and the height represents the degree of correlation, $C(i,j)$. Since the matrix is symmetric, only half of the contour map is displayed. The diagonal elements which appear horizontally in Figure 2a and from the lower left to the upper right in Figure 3 correspond to self-correlations, which are by definition equal to 1 [Eq. 1]. The nearest neighbor correlations (i.e., residue i with residue $i+1$) are also large and positive (between 0.2 and 0.9), which is expected since these are adjacent residues. As the residue pairs get further apart in sequence number (i.e., away from the diagonal), the correlation tends to decrease. The minimum value is about -0.7, and the maximum value away from the diagonal is about 0.6.

An examination of the maps shows significant correlations between residues that are far apart in sequence number. To interpret these results, it is necessary to relate these correlations to the structure of the protein. In particular, residues which are far apart in sequence number are not necessarily far apart in actual distance because of the folding of the protein chain. Figure 2b is a three-dimensional map and Figure 4 is a contour map of the distance matrix for the LCS average structure in which the abscissa and the ordinate are residue numbers, i and j, and the contours correspond to $\Delta R(i,j)$, the distance between the residues. The diagonal of this plot is equal to 0 and increasing distance is in the $-Z$ direction. Comparisons of Figures 2a and b and of 3 and 4 show that there is considerable similarity for the major features, but that there are also some significant differences.

We first consider the regions of secondary structure in the distance map (Fig. 4). Such a contour plot

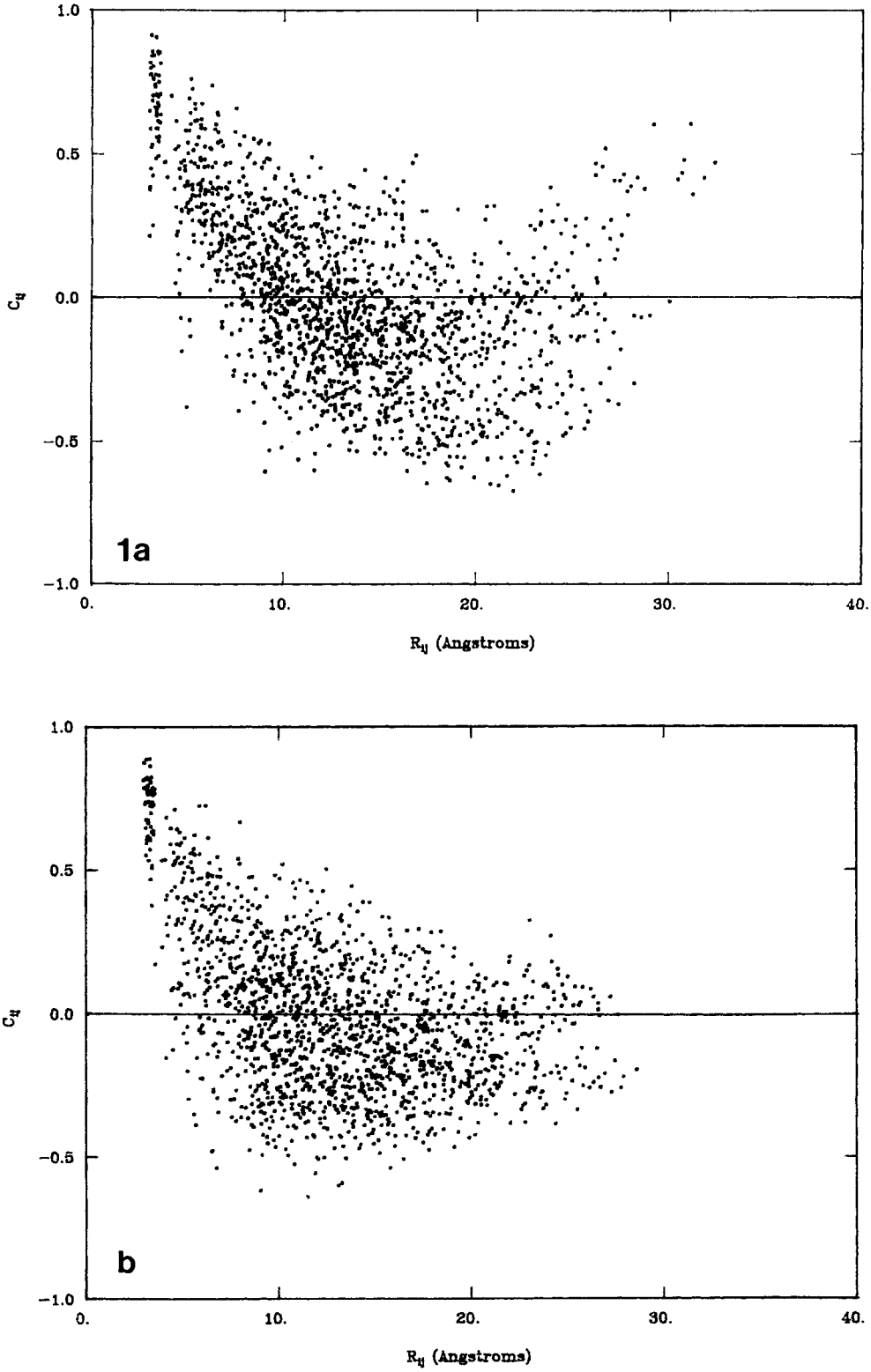
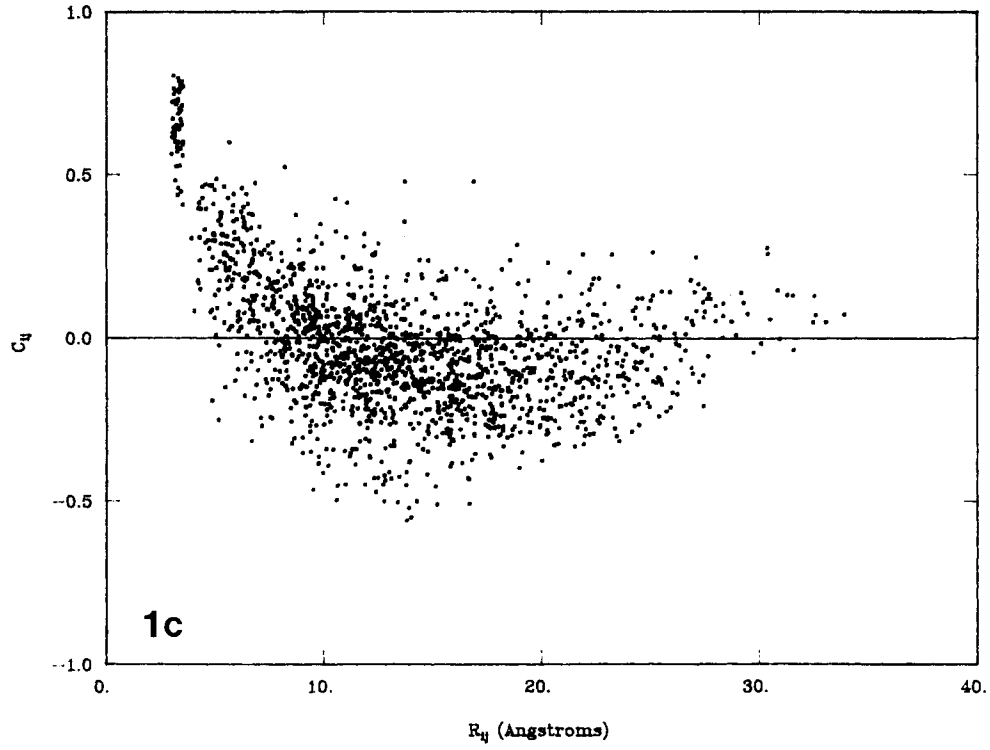


Fig. 1. Cross-correlation as a function of interatomic distance: (**a**) LCS, (**b**) LC., (**c**) LCC, (**d**) NM stimulations.



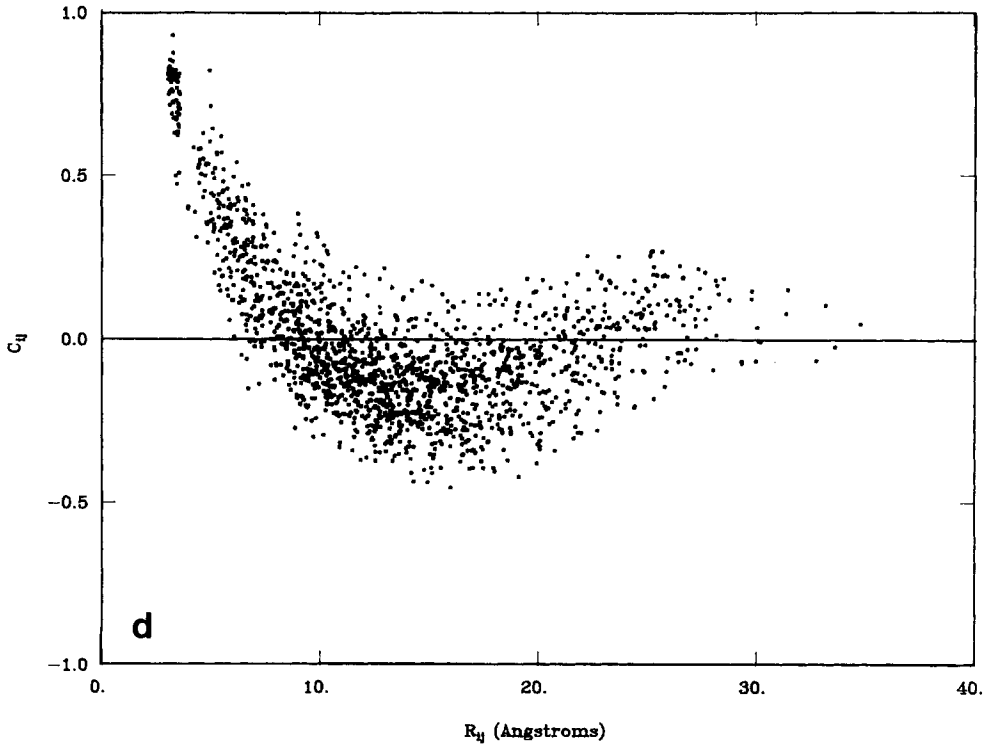


Fig. 1, c, d. Legend appears on page 208.

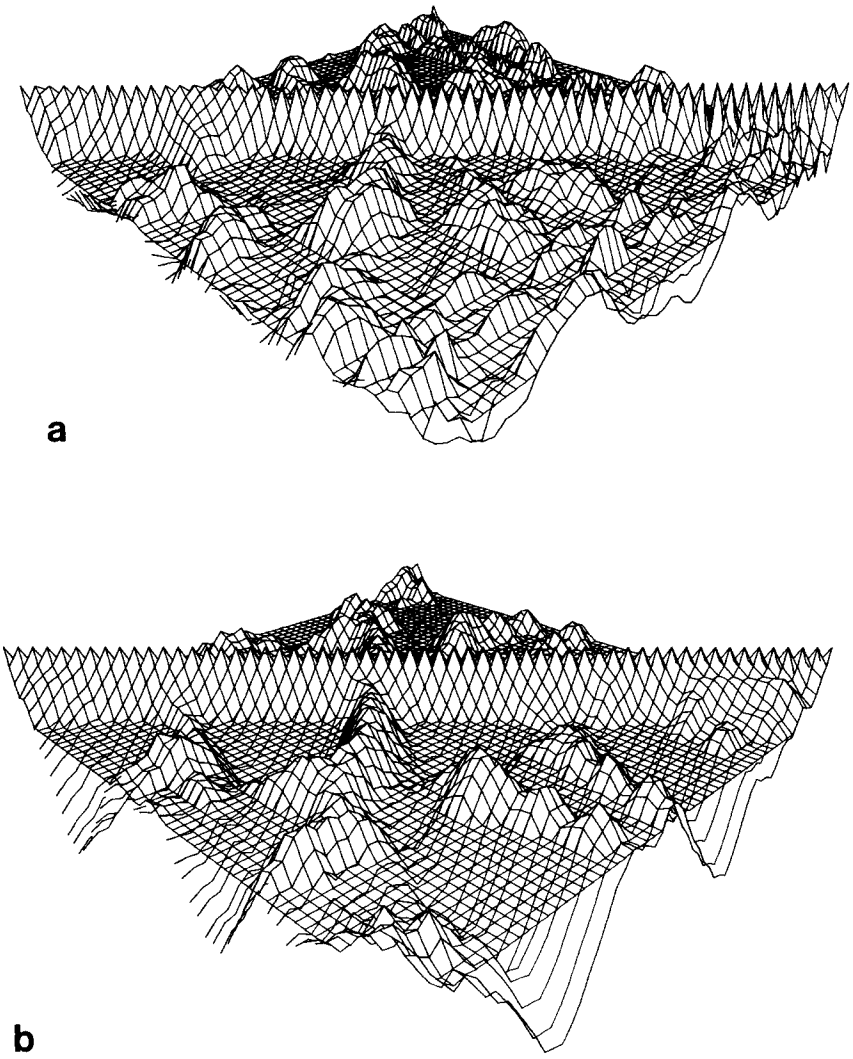


Fig. 2. Three-dimensional cross-correlation and distance plots, averaged by residue over backbone atoms, for the LCS simulation. The *x* and *y* directions correspond to residue number, with the origin at the left hand corner. (**a**) Cross-correlation matrix of atomic fluctuations; the *z* direction represents the value of the cross-correlation (the cross-correlations are equal to 1 along the diagonal and are less than one everywhere else; the plane corresponds to zero with values above positive and below negative). (**b**) Interatomic distance matrix; the *z* direction corresponds to the value of the interatomic distance (the interatomic distances are equal to 0 along the diagonal and increasing distances are plotted in the negative z direction; a plane at 12 Å is shown for reference).

was first used for the protein ribonuclease-S by Phillips,[20] who found clear patterns for the α-helices and β-sheets. To aid in the description, Figure 5a shows the α carbon positions of BPTI based on the structure of Deisenhofer and Steigemann[21] and Figure 5b presents the pattern of main chain hydrogen bonds. The minimum (under 6 Å) extending from (29,24) to (36,16) perpendicular to the diagonal, corresponds to the two strands of antiparallel β-sheet (residues 16 to 24 and 36 to 29). The minimum (under 6 Å) parallel to the diagonal from (47,47) to 56,56) for residues separated by about four or five residues corresponds to the C terminal α-helix (residues 47 to 56). The 3_{10} helix (residues 3 to 7) is also visible, although it is not apparent in the average LCC structure.[6] Furthermore, the strands made up of residues 5 to 13 and 38 to 48 are close to the β-sheet, giving rise to minima perpendicular to the diagonal from (16,13) to (24,5), (41,24) to (48,18) and (43,36) to (48,29), as well as being close to each others, with a minimum from (38,13) to (45,4). The disulfide bonds between residues 5 and 55, 14 and 38, and 30 and 51 appears as minima (between 5 and 7 Å) on the plot.

The contour plot of the cross-correlation matrix for the LCS simulation (Fig. 3a) shows that most of the regions of positive correlation in the cross-correlation plot for the LCS run correspond to minima in the interatomic distance plot. In particular, there are positive values up to 0.5 between residue pairs of the β-sheet and between residue pairs of the α-helix. In the helix region, residues 47 to 56, there is a strong positive correlation (0.35 to 0.55) between residues i and $i+3$ or $i+4$, which are hydrogen

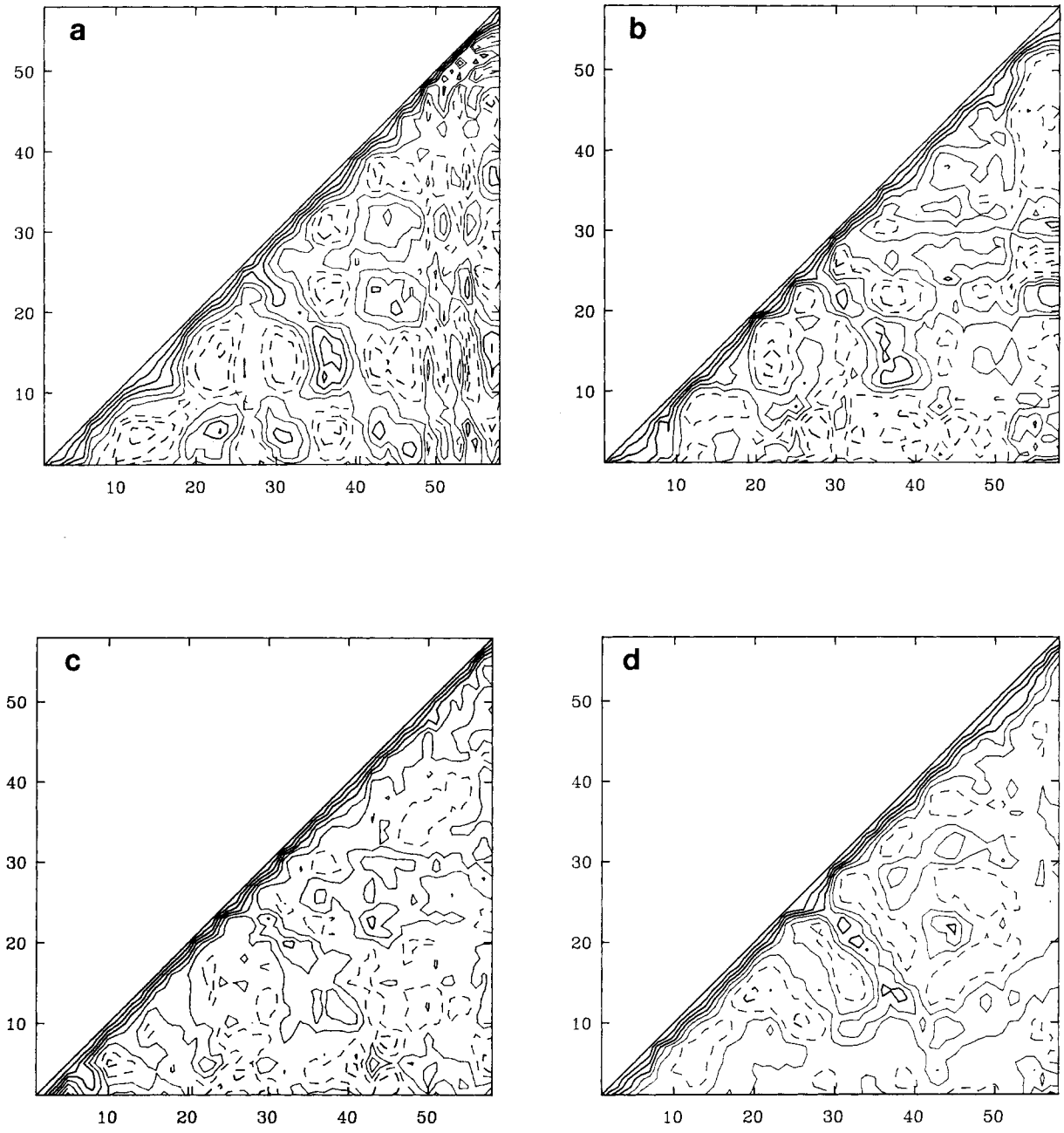


Fig. 3. Contour diagrams of cross-correlations and covariances of atomic fluctuations, averaged by residue over backbone atoms. Cross-correlations: (**a**) LCS, (**b**) LC, (**c**) LCC, (**d**) NM. Solid lines represent values greater than or equal to zero and dashed lines represent values less than zero. Contour levels are drawn at intervals of 0.2; they are equal to 1 along the diagonal and otherwise are less than 1.

bonded, and a somewhat weaker one (up to 0.2) between i and $i+8$. Furthermore, there are strong negative correlations (up to -0.45) between residues i and $i+2$, i and $i+6$ and i and $i+10$, which are on opposite sides of the helix. When the simulation is translated to the frame of the helix, the cross-correlation pattern becomes even more distinct; however, when it is translated and rotated to the frame of the helix, the pattern disappears indicating that it is due to rotation about the long axis of the helix. Moreover, the correlation function of a vector along the helix diameter when translated but not rotated to the frame of the helix shows a decay, indicative of rotation about the long axis. In the β-sheet region, within the triangle bounded by $i=39$ and $j=10$, which defines the cross-correlation of the β-sheet with itself, the negative region (up to -0.6) centered at (23,15), (31,15), (36,23), and C(36,31) is due to anticorrelation of the two ends of the sheet. The correlations are high (0.56) between residues 14 and 38,

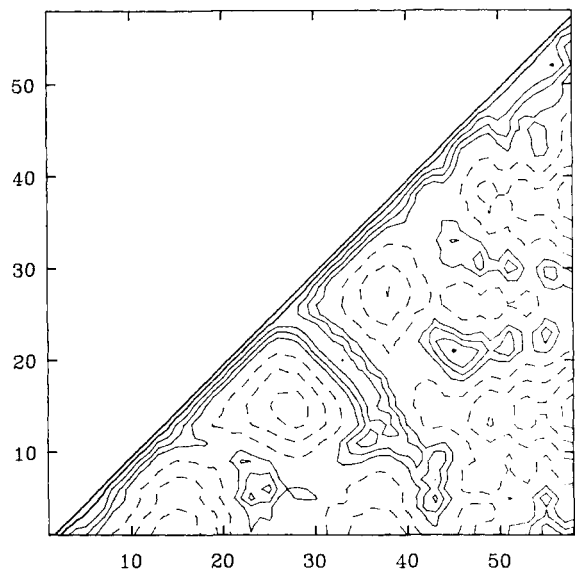


Fig. 4. Contour diagram of interatomic distances, averaged by residue over backbone atoms, for the average LCS structure. The abscissa and ordinate are residue numbers and the contours represent $\Delta R(i,j)$, with solid contour levels every 2 Å between 0 and 10 Å and dashed contour levels every 5 Å between 10 and 30 Å.

which form a disulfide bond at the top of the molecule at one end of the β-sheet. The correlations between residues 5 and 55 and at 30 and 51 are low, even though they also form disulfide bonds. Evidently, the backbone atoms of these residues are constrained by the surrounding protein, in contrast to the relatively free motion of residues 14 and 38, which are more exposed. The correlations of the side chain atoms involved in all the disulfide bonds are high.

There are also correlations between residues which are not directly bonded to each other, either by covalent or hydrogen bonds. There are positive correlations between residues in the β-sheet and the nearby residues 5 to 14 and 38 to 48, both of which are strands running parallel to the sheet (see Fig. 4). Thus, the strands 13 to 5 and 38 to 48 and the β-sheet strands 16 to 24 and 36 to 29, which comprise much of the molecule, move as a flexible unit. There are, in addition, maxima corresponding to correlation of the 3_{10} helix with residues near 31 and 47 due to side chain interactions. The middle region, consisting of the 3_{10} helix, a hydrophobic region (which includes Phe-33, Tyr-21, Phe-22, Tyr-23, and Phe-45) and one side of the helix, is correlated. The loop from residues 24 to 29 at the "bottom" of the molecule has small positive correlations with most of the rest of the molecule, except that it is anticorrelated with the 3_{10} helix. Of most interest is the strong correlation (0.3 to 0.6) of residues 56 to 58 at the "bottom" of the molecule (Fig. 5) with residues 12 to 17 and 35 to 38 at the "top" of the molecule, which are separated by more than 30 Å; this corresponds to the positive correlations in the long distance portion of Figure 1.

In the cross-correlations of fluctuations averaged over backbone atoms versus those for individual atoms, the effects of bond-angle and some torsion vibrations are averaged out. The contour plots of C_α and backbone N cross-correlations are very similar to those for the average backbone fluctuations. Thus, the very local fluctuations appear to make relatively minor contributions.

Contour plots were also made for side chain vs. side chain and side chain vs. backbone correlations (Fig. 6a,b). The side chain vs. side chain plot shows smaller nearest neighbor correlations (some close to zero) and correlations in the β-sheet (between 0.0 and 0.4) than the backbone vs. backbone plot but stronger correlations, both positive and negative, of the α-helix (residues 47 to 56) with itself (between −0.7 and 0.8) as well as with other residues. The 3_{10} helix side chains also appear to have stronger correlations with other residues. The correlations of the side chains involved in disulfide bonds (0.52 to 0.76) are also higher than the backbone values. The side chain vs. backbone plot is interesting because a given side chain may have interactions with several backbone groups, while a given residue backbone will tend to have interactions with only a few adjacent side chain groups. This asymmetry is evident in the skewed appearance of the contour plots.

Dependence on Simulation

The equivalent plots for the other trajectories are quite similar to LCS results with some notable exceptions. The LC run (Fig. 3b) is marked by an overall reduction in the magnitude of the correlations. The positive and negative correlations found in the LCS run between the 3_{10} and α- helices and the rest of the protein are not apparent in the LC run. There is, however, a strong correlation between residues 22, 31, and the C-terminus and a somewhat weaker correlation with 1 and 5, probably due to structural changes associated with shrinkage.

The behavior of the α-helix is different in the LCS and LC simulations. From the time series of a vector along the helix axis, the α-helix moves more in the LC run than in the LCS run. When each of the two simulations is translated to the frame of the α-helix, it appears that while the α-helix rotates about the helix axis in the LCS run, it rotates perpendicular to the axis in the LC run; there is a large region of anticorrelation centered about (49,55) in the latter. When the LC and LCS runs are translated and rotated to the helix frame, the cross-correlations of these regions are more alike, indicating that the deviations are due mainly to differences in overall motion of the helix.

The cross-correlation plot for the NM calculation (Fig. 3d) shows even smaller magnitudes for the long-range correlations than the LC run. It is simi-

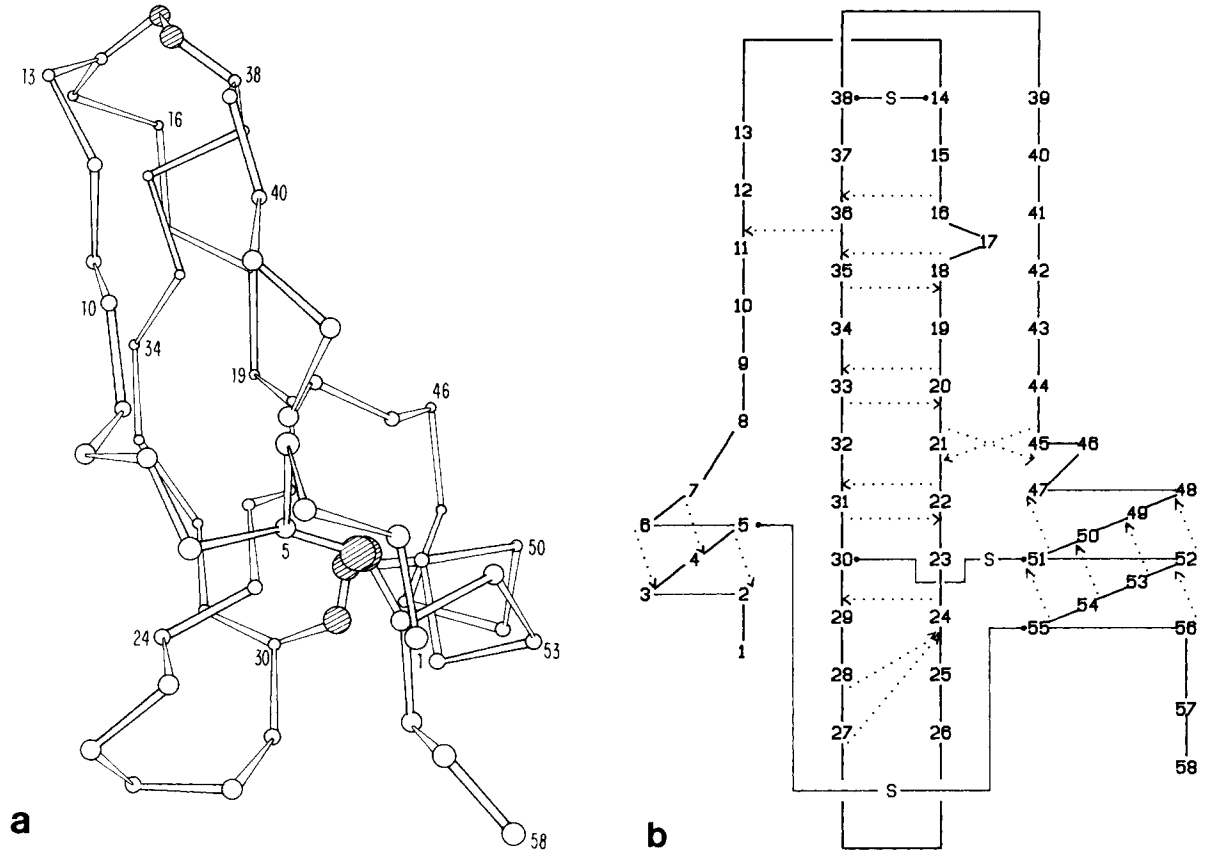


Fig. 5. (**a**) X-ray crystal structure showing the C_α atoms[21] and (**b**) hydrogen bonding pattern of PTI. A hydrogen bond was defined to have a donor-acceptor distance less than 3.1 Å except in the 3-10 helix at residues 3–7, where the distances were between 3.1 to 3.5 Å. Hydrogen bonds are given by dashed lines and disulfide bonds are denoted by solid lines with an "S."

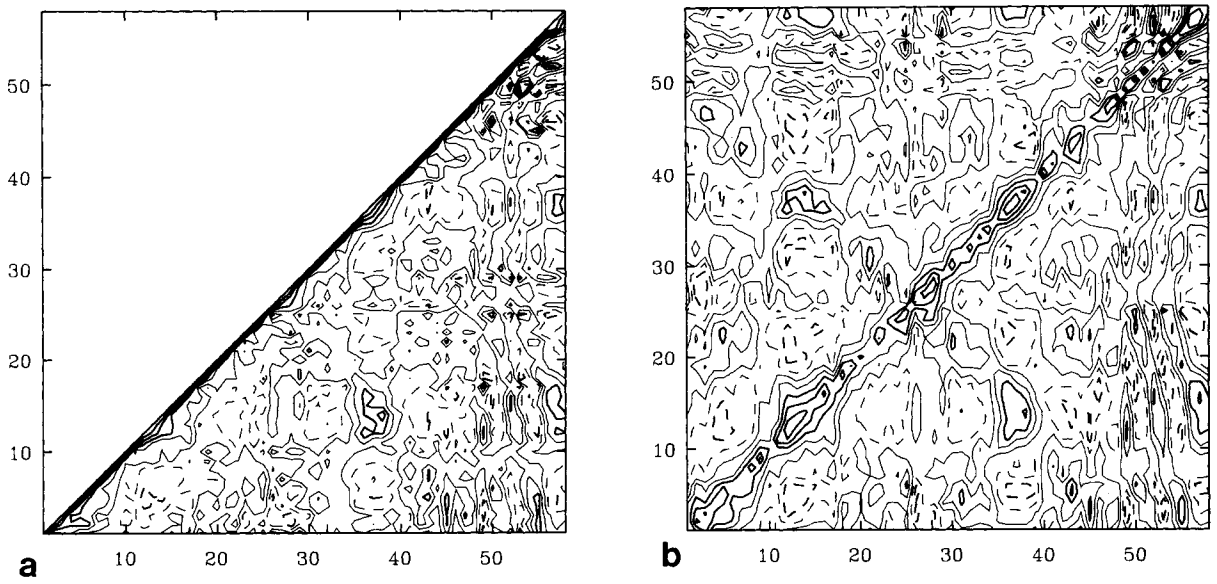


Fig. 6. Contour diagram of (**a**) side chain vs. side chain and (**b**) side chain vs. backbone residue averaged cross-correlations for the LCS simulation; in the latter, the horizontal axis corresponds to the side chain and the vertical axis to the backbone.

TABLE I. Correlation Coefficients (Eq. 5) for Internal Motions

	LCS	LC	LCC	NM
	a. R_c for $C(i,j)$, $i \neq j$			
LCSB	0.96	0.44	0.54	0.67
LCS		0.48	0.56	0.63
LC		0.57	0.62	
LCC				0.62
	b. R_c for $c(i,j)$, $i \neq j$			
LCSB	0.94	0.37	0.46	0.66
LCS		0.43	0.46	0.60
LC			0.48	0.55
LCC				0.47
	c. R_c for $c(i,i) = \langle \Delta r_i^2 \rangle$			
LCSB	0.96	0.62	0.38	0.75
LCS		0.64	0.40	0.69
LC			0.51	0.55
LCC				0.23

TABLE II. Comparison of Structures

	LC	LCC	NM
	a. R_c for $r(i,j)$, $i \neq j$ (Eq. 5)		
LCS	0.98	0.98	0.99
LC		0.96	0.97
LCC			0.98
	b. rms difference of C_α in Å		
LCS	1.71	1.60	1.36
LC		2.32	2.11
LCC			1.47
LCS	LC	LCC	NM
	c. R_{gyr} in Å		
10.5	10.1	10.6	10.6

lar to the LC simulation in that the rotation about the long axis of the α-helix (residues 47 through 56) is not apparent and the two termini do not move as much. In general, the correlation patterns follow the patterns of the interatomic distance plot much more closely than the other simulations, as indicated by Figure 1d.

To quantitate the degree of correlation between the backbone cross-correlations and covariances for the four simulations, the linear correlation coefficients, R_c, were calculated. This quantity is defined as:

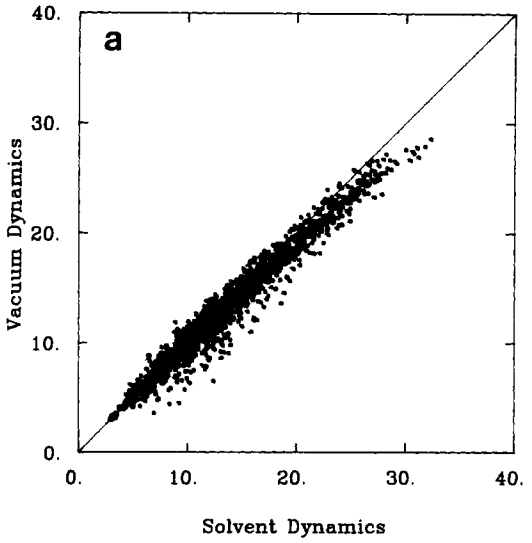
$$R_c = \frac{\overline{pq} - \overline{p}\,\overline{q}}{(\overline{p^2} - \bar{p}^2)^{1/2}\,(\overline{q^2} - \bar{q}^2)^{1/2}} \qquad (5)$$

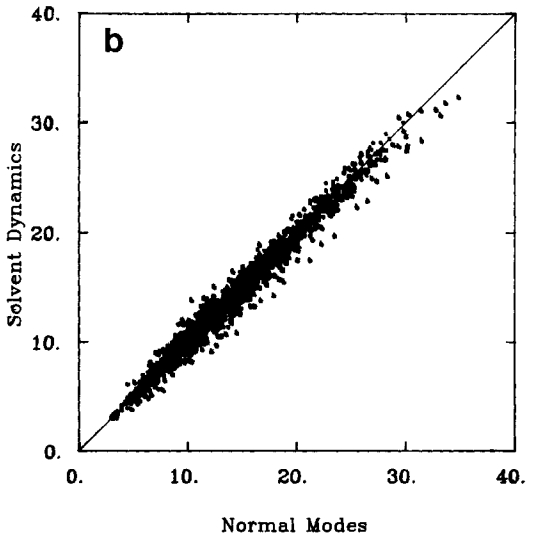
in which p and q are the elements for the two simulations being compared and the overbars refer to an average over the elements of the lower triangular matrix. In Table Ia, the values of R_c for $C(i,j)$ for pairs with $i \neq j$ (1653 points) are presented. The values of R_c show that the simulations are all correlated with each other (the probability of a random sample of 1,000 points from uncorrelated distributions having a larger value of R_c is $<<0.001$ for all the simulations). The LCSB trajectory, which is from the same simulation as the LCS but has overall rotation and translation removed by matching all mass-weighted atoms as opposed to the only backbone atoms, is very similar to the LCS with R_c = 0.96. The differences in the correlation of LCSB and LCS with a given simulation indicate the effects of overall motion on the cross-correlations. The procedure used for LCSB is more correct, and the better agreement of LCSB than LCS with NM may be due in part to the fact that both LCSB and the NM have overall motions removed in the same manner. All of the simulations have the highest correlation with the NM analysis (R_c = 0.62–67), with LCSB most correlated with it.

The value of R_c for the $c(i,j)$ matrices are given in Table Ib. The correlation of the $c(i,j)$ matrices of two simulation is somewhat less than that of the $C(i,j)$ matrices. This is because strongly correlated or anticorrelated regions tend to have large fluctuations (i.e., they are associated with collective motions rather than just local fluctuations) so the apparent differences are increased by including information about the magnitude of the fluctuations. Better agreement of the $C(i,j)$ matrices than the $c(i,j)$ matrices indicates that the simulations are similar in the relative motions of groups more so than in magnitudes of motions. The most notable difference in the trends of the $c(i,j)$ versus the $C(i,j)$ matrix is that the NM simulation has much greater differences in correlation with the rest of the simulations (i.e., a low value of R_c = 0.47 with the LCC and a high value of 0.66 with the LCSB).

The correlation of the diagonal elements of $c(i,j)$ (Table Ic), which are the mean-square fluctuations for the backbone atoms of a given residue, indicates that the fluctuations for the LCS and NM simulations (R_c = 0.69–0.75) are most similar. The LC simulation is most like the LCS simulation, although the converse is not true, and the LCC fluctuations are least like any of the other simulations, being closest to the LC simulation (R_c = 0.51).

The agreement of the correlation matrices is somewhat dependent on the differences in structure since the relative orientation of the motions of the two residues being correlated affects the magnitude of $C(i,j)$. The values of R_c of the interatomic distance matrices are given in Table IIa and the rms deviations between the structures are given in Table IIb. We note that even if R_c = 1 for the interatomic distances, the RMS deviation between the structures need not be zero, since R_c measures deviations from linear dependence, regardless of slope or intercept. For instance, in the case of uniform shrinkage of one structure with respect to a second, the position relative to the center of mass of every atom in one structure, $\mathbf{r}_1$, is related to that of the second $\mathbf{r}_2$,





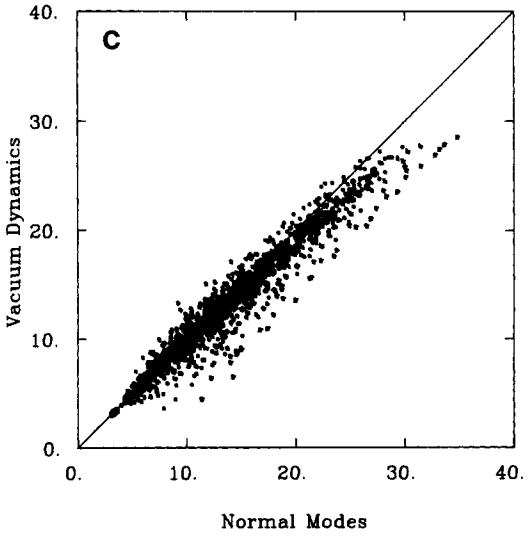


Fig. 7. Scatter plots of interatomic distances. (**a**) Solvent vs. vacuum, (**b**) normal mode vs. solvent, (**c**) normal mode vs. vacuum.

by $\mathbf{r}_1 = b\mathbf{r}_2$ where $b = R_{gyr,1}/R_{gyr,2}$ and R_{gyr} is the radius of gyration. Although $R_c = 1$ for this case, the rms deviation of the two structures is $(R_{gyr,1}-R_{gyr,2})$. Thus, $(R_{gyr,1}-R_{gyr,2})$ for the C_α atoms (Table IIc) indicate about 0.5 Å of the rms differences of the LC structures with the other structures may be accounted for by the differences in the size of the average structures. Another cause of large rms differences in atomic positions with values of $R_c \sim 1$ is if subregions of two structures are very similar but are packed differently in the two structures. Scatter plots of interatomic distances (Fig. 7) show that the close contacts are very similar in all of the average structures but that the atoms separated by large distances are closer together in the LC average structure than any other structure. It is also apparent that the LCS structure is very similar to the NM structure.

DISCUSSION

To determine the significance of the results reported in this paper, it is necessary to consider the nature of and differences among the various simulations. All of the molecular dynamics simulations used the united atom approximation (i.e., all hydrogens treated as part of the extended atom to which they are bonded) while the normal mode calculation used a polar hydrogen force field, in which polar hydrogens are treated explicitly.[8] Studies have shown that the polar hydrogen model gives a better description of hydrogen bonding, particularly when the hydrogen bonds are made or broken. However, the effects on small amplitude motions is not very important. Also, there have been other improvements in the force field parameters since the simulations were made. The close agreement between the normal mode and molecular dynamics results indicates that the behavior studied in this paper is rather insensitive to the force field. This is encouraging since even the best force fields in present use are not highly accurate.

The normal mode calculation is based on the assumption that the atomic motions are harmonic, whereas molecular dynamics simulations include anharmonic effects. It has been shown that the potentials of mean force for atoms in proteins often have multiple wells and are softer than harmonic barriers to large fluctuations.[23] The lack of anharmonic contributions to the second derivatives of the potential energy in the normal mode calculation restricts the motion to a single minimum, but the results are exact for the model. Molecular dynamics simulations allow a more complete coverage of the conformational space but longer simulations would be required to obtain accurate statistics on the slow

collective motions. For the lowest frequency (on the order of 3 cm^{-1}) in the normal mode analysis, a 25 psec simulation corresponds to only 2.5 periods.

Another factor which affects the accuracy of the simulation is the treatment of the environment. The vacuum simulation is least like BPTI under normal experimental conditions in solution or in a crystal and thus the molecule shrinks during the course of the simulation. The effect of shrinkage has been reduced in the solvent and crystal simulations and in the normal mode analysis. In the crystal simulation, this is accomplished by static crystal neighbors, and in the normal modes calculation, by introducing harmonic constraints in the energy minimization. In a real solution or crystal environment, low-frequency modes could be coupled to phonon modes, which are not included in the crystal or normal mode calculations. This may contribute to the reduction in the long-range correlations in these two calculations relative to the other simulations. Despite the use of a van der Waals rather than polar solvent and the possibility of artifacts from the periodic boundary conditions,[22] the solvent simulation is expected to give the most realistic picture of the motions in BPTI.

Overall, the general agreement between the different calculations (different force fields, different models, different environments) indicates that trends seen in the simulations are probably correct despite the shortcomings in each. As already recognized when the first simulations were made,[1] the most important factor determining the internal motions is bond connectivity and packing.

Turning now to the results, the most striking conclusion is that the cross-correlations of the atomic fluctuations are related to the distances between the atoms. The pattern seen for all of the simulations is positive correlations for small separations and negative correlations for separations corresponding to about half of the length of the molecule. This pattern may be interpreted in terms of the structural features of BPTI. Residues which are adjacent in sequence have highly correlated motions due to the covalent bonds. Those within the same secondary structure also have correlated motions, but to a somewhat lesser degree, due to the presence of hydrogen bonds. Larger scale motions of the secondary structures can be seen in patterns of correlation and anticorrelation, such as the rotation of the α-helix and the bending of the β-sheet in the solvent simulation. Although the exact nature of these slower motions cannot be ascertained from the simulation because of the relative short timescale of the simulation, it is evident that secondary structural elements can move as collective groups.

The type of analysis presented here has many possible applications. One is for the identification of collective motions. For instance, for a protein with a hinge-bending mode, the cross-correlation of any pair of the atoms in the same lobe may be dominated by that motion so that atoms within a lobe would be highly correlated and atoms on different lobes would be highly anticorrelated. An application of this type has been made recently to interpret the internal motions obtained in molecular dynamics simulations of the HIV protease dimer.[24] Another application is for choosing atoms to include in a simulation in which only a small region of a protein, such as an active site, is treated by full molecular dynamics and the rest are neglected or constrained by a harmonic force field.[25,26] The degree of correlation of the local fluctuation would be difficult to detect from cross-correlations if the region itself is involved in a large amplitude correlated motion. By calculating cross-correlations over smaller time intervals, more local correlations could to be seen. A complementary method useful for identifying more local correlations when more than one type of collective motion is involved is to compare the time correlation functions of the atoms involved.[5]

CONCLUSIONS

The cross-correlations and covariances of atomic fluctuations provide new insights concerning collective motions in proteins. They demonstrate that secondary structural elements can move as collective groups. That molecular dynamics simulations in different environments and a normal mode analysis give similar results for the correlation of the internal motions suggests that such collective effects are controlled by the structure of the molecule. Environmental effects appear to be important only for the long range, larger scale collective motions. Application of the same type of analysis to simulations of other proteins in different environments would be of interest to determine whether the results obtained for BPTI are of general applicability.

ACKNOWLEDGMENTS

We thank Bernard R. Brooks for providing the normal modes, and S. Swaminathan and Hsiang-ai Yu for providing the rotated trajectories for the solvent and crystal simulations. We also thank Charles L. Brooks, III and B. Montgomery Pettitt for helpful discussions. Supported in part by a grants from the National Science Foundation and the National Institutes of Health.

REFERENCES

1. Brooks, C.L. III, Karplus, M., Pettitt, B.M. "Proteins: A Theoretical Perspective of Dynamics, Structure, and Thermodynamics." Adv. Chem. Phys. LXXI. New York: John Wiley, 1988.
2. McCammon, J.A., Harvey, S. "Dynamics of Proteins and Nucleic Acids." Cambridge: Cambridge University Press, 1987.
3. Ringe, D., Petsko, G.A. Study of protein dynamics by X-ray diffraction. Methods Enzymol. 131:389–433, 1986.
4. Petsko, G.A., Ringe, D. Fluctuations in protein structure from X-ray diffraction. Annu. Rev. Biophys. Bioeng. 13: 331–371, 1984.

5. Swaminathan, S., Ichiye, T., van Gunsteren, W.F., Karplus, M. Time dependence of atomic fluctuations in proteins: Analysis of local and collective motions in bovine pancreatic trypsin inhibitor. Biochemistry 21:5230–5241, 1982.
6. Ichiye, T., Ph.D. The internal dynamics of proteins. Thesis. Harvard University, 1985.
7. Ichiye, T., Karplus, M. To be published.
8. Brooks, B.R., 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.
9. 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.
10. Levitt, M., Sauder, C., Stern, P.S. Protein normal-mode dynamics: Trypsin inhibitor, crambin, ribonuclease and lysozyme. J. Mol. Biol. 181:423–447, 1985.
11. Karplus, M., Kushick, J.N. Method for estimating the configurational entropy of macromolecules. Macromolecules 14:325–332, 1981.
12. Levy, R.M., Karplus, M., Kushick, J., Perahia, D. Evaluation of the configurational entropy for proteins: Application to molecular dynamics simulations of an α-helix. Macromolecules 17:1370–1374, 1984.
13. Brooks, B.R., Karplus, M. To be published.
14. Caspar, D.L.D., Clarage, J., Salunke, D.M., Clarage, M. Liquid-like movements in crystalline insulin. Nature (London) 332:659–662, 1988.
15. van Gunsteren, W.F., Karplus, M. Protein dynamics in solution and in a crystalline environment: A molecular dynamics study. Biochemistry 21:2259–2274, 1982.
16. van Gunsteren, W.F., Berendsen, H.J.C. Algorithms for macromolecular dynamics and constraint dynamics. Mol. Phys. 34:1311–1327, 1977.
17. van Gunsteren, W.F., Karplus, M. Effect of constraints on the dynamics of macromolecules. Macromolecules 15: 1528–1544, 1982.
18. Stillinger, F.H., Rahman, A. Molecular dynamics study of temperature effects on water structure and kinetics. J. Chem. Phys. 57:1281–1292, 1972.
19. Yu, H.-A., Karplus, M. Unpublished results.
20. Phillips, D.C. "Biochemical Society Symposium No. 30 in 1969, British Biochemistry Past and Present." Goodwin, T.W. (eds.). London: Academic Press, 1970: 11.
21. Deisenhofer, J., Steigemann, W. Crystallographic refinement and the structure of the bovine pancreatic trypsin inhibitor at 1.5 Å resolution. Acta Crystallogr. B31:238–250, 1975.
22. Allen, M.P., Tildesley, D.J. "Computer Simulation of Liquids." Oxford: Clarendon Press, 1987.
23. Ichiye, T., Karplus, M. Anisotropy and anharmonicity of atomic fluctuations in proteins: Analysis of a molecular dynamics simulation. Proteins 2:236–259, 1987.
24. Swaminathan, S., Harte, W.E., Jr., Beveridge, D.L. Identification of domain structure in proteins via molecular dynamics simulation: application to HIV-1 protease dimer, J. Am. Chem. Soc. (in press).
25. Brooks, C.L. III, Karplus, M. Deformable stochastic boundaries in molecular dynamics. J. Chem. Phys. 79: 6312–6325, 1983.
26. Brünger, A., Brooks, C.L. III, Karplus, M. Active site dynamics of ribonuclease. Proc. Natl. Acad. Sci. U.S.A. 82: 8458–8462, 1985.