# Efficient Characterization of Collective Motions and Interresidue Correlations in Proteins by Low-Resolution Simulations[†]

I. Bahar,[‡,§] B. Erman,[‡] T. Haliloglu,[‡] and R. L. Jernigan*[,§]

*Polymer Research Center and School of Engineering, Bogazici University, and TUBITAK Advanced Polymeric Materials Research Center, Bebek 80815, Istanbul, Turkey, and Molecular Structure Section, Laboratory of Experimental and Computational Biology, Division of Basic Sciences, National Cancer Institute, National Institutes of Health, MSC 5677, Room B-116, Building 12B, Bethesda, Maryland 20892-5677*

ABSTRACT: A low-resolution model is used together with recently developed knowledge-based potentials for exploring the dynamics of proteins. Configurations are generated using a Monte Carlo/Metropolis scheme combined with a singular value decomposition technique (SVD). The approach is shown to characterize the cooperative motions in good detail, at least 1 order of magnitude faster than atomic simulations. Trajectories are partitioned into modes, and the slowest ones are analyzed to elucidate the dominant mechanism of collective motions. Calculations performed for bacteriophage T4 lysozyme, a two-domain enzyme, demonstrate that the structural elements within each domain are subject to strongly coupled motions, whereas the motions of the two domains with respect to each other are strongly anticorrelated. This type of motion, evidenced by the synchronous fluctuations of the domain centroids by up to ±4.0 Å in opposite directions, is comparable to the movements observed by recent spin-labeling experiments in solution. The potential of mean force governing these fluctuations is shown to be anharmonic. The $\beta$-sheet region at the N-terminal domain and the helix E in the C-terminal domain are identified as regions important for mediating cooperative motions and, in particular, for the opening and closing of the active-site cleft between the domains. Residues Leu66−Phe67 in the central helix C stop the propagation of correlated motions between the domains. There is a correlation between the groups involved in highly cooperative motions revealed by simulations and the highly protected regions during unfolding measured by pulsed H/D exchange and 2-D NMR.

A multitude of conformational substates, each of them constituting a local minimum on the highly structured energy landscape, exist in the neighborhood of the native state. These deviate only slightly from the X-ray structure. Fluctuations between such substates are commonly observed in molecular dynamics (MD)[1] simulations *(1)*, while larger-scale conformational changes, including those that eventually lead to unfolding, are less commonly accessible to simulations. Atomic models used in such simulations necessitate the adoption of time steps of the order of femtoseconds, which do not permit attaining time scales longer than nanoseconds. A typical protein MD simulation samples only a limited portion of the overall conformational space, as evidenced by the projection of the trajectory onto the subspace spanned by the three dominant eigenvectors of the displacement covariance matrix *(2)*. Another difficulty with MD simulations is that cross-correlations between the displacements of different atoms in a given protein cannot be precisely captured *(2)*. Such limitations motivate the quest for simplified models and more efficient computational tools.

Numerous recent studies of protein conformations and interactions have aimed at less detailed coarse-grained models. Some of these efforts have been directed toward clarifying the principles governing the structural preferences of proteins *(3)*. The analysis of interresidue contact preferences in X-ray-elucidated structures, in particular, has contributed significantly to our understanding of the dominant forces stabilizing native structures *(4, 5)*. These potentials are conveniently used in coarse-grained models, which presently constitute one way to comprehensively assess the global structural features of proteins *(6−9)*. However, those coarse-grained models have not yet found application to the conformational dynamics of proteins, which is the intention of the present paper.

In addition to the model and parameters, the details of the simulation and information retrieval methods are important for the efficient characterization of a process. Here we make use of the singular value decomposition (SVD) technique, which is particularly useful for characterizing the nonlinear dynamics of multivariate systems *(10)* and identifying the dominant modes of motion in systems whose cooperative dynamics cannot be fully explored within reasonable computation time using conventional simulation techniques. The basic approach in the SVD method is to project a complex motion of *m* degrees of freedom from an *m*-dimensional space into a lower-dimensional space which

most efficiently describes the collective dynamics. The base vectors of the new, reduced space are taken along the dominant directions of motion, also referred to as left singular vectors (LSV), or the principal molecule-optimal dynamic coordinates *(11)*. Reconstruction of MD trajectories in such a reduced space has proven to be useful for elucidating the cross-correlations between atomic displacements and for unraveling the collective dynamics of secondary structural elements and/or domains *(12, 11)*. Romo et al. *(12)* showed, for example, that a mutant of myoglobin, F46V metmyoglobin, alternates between two configurations characterized by the swinging of His64 between open and closed states at the entrance to the heme binding pocket of the protein. Likewise, Garcia and Harman *(11)* applied this same method to analyze the MD trajectories of the cyclic 3′,5′-adenosine monophosphate (cAMP) receptor protein (CRP) complexed with $(cAMP)_2$ to demonstrate the occurrence of a subunit transition from the open to the closed conformation in noncrystalline environments.

Here we will apply the SVD technique to a coarse-grained protein model whose geometric and energetic characteristics were determined in our previous work *(13, 9, 14)*. Calculations will be performed for bacteriophage T4 lysozyme (T4L), an enzyme of 164 residues produced in cells of *Escherichia coli* after infection with T4 bacteriophage, having a similar catalytic activity to that of hen egg white lysozyme. Systematic mutational studies *(15−19)*, H/D exchange coupled with 2-D NMR *(20)*, and site-directed spin-labeling experiments *(21)* have been carried out on T4L, as well as MD simulations with full atomic representation and explicit hydration *(22, 23)*, which provide a wealth of information for testing the performance of our approach.

The structure of T4L was first determined by Matthews and Remington *(24)* and the diffraction data were later refined to 1.7 Å resolution *(25, 26)*. T4L is composed of two domains connected by a 21-residue α-helix (C), as shown in Figure 1. In the N-terminal domain, hereafter referred to as lobe I, there is a β-sheet composed of four strands, three of which form a slightly distorted antiparallel pleated sheet and the fourth connecting a 12-residue helix (B) to helix C. The N-terminus of the chain is folded into an α-helical structure (helix A), which interacts closely with the C-terminal domain. The C-terminal domain (lobe II), on the other hand, contains five helices, D−H. We note that helix A is close in space to lobe II, and may be equally viewed as part of the C-terminal domain (see Figure 1). There is a deep opening between the two lobes. This is the active-site cleft for oligosaccharide binding *(27)*. A hinge-bending-type domain motion opening or closing up the cleft was indicated in mutation experiments *(28, 29, 19)* and in MD simulations *(23)*, which was suggested to be an integral part of the catalytic function of the protein *(19, 21)*.

Our approach is similar to that of Romo *et al. (12)* and that of Garcia and Harman *(11)*, in that we also examine the distribution of conformational substates and also characterize the dominant modes of fluctuations with the SVD technique. However, one main difference is that we analyze a coarse-grained model *(9, 14)*, as opposed to the full atomic description of those studies, and we do Monte Carlo instead of MD simulations. Significantly, this is the first study of conformational fluctuations with the use of a coarse-grained protein.

In the *virtual bond model* presently adopted, each residue is represented by two interaction sites, one on the backbone
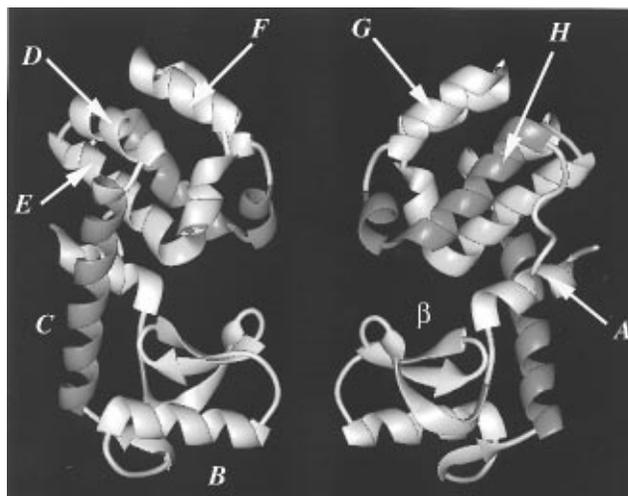


FIGURE 1: Ribbon representation of bacteriophage T4 lysozyme (T4L) (front and back views). The protein consists of two lobes connected by a long α-helix (helix C, Lys60−Arg80). Residues in lobe I form the N-terminal domain, except for the helix A at the N-terminus, which will be shown to be coupled to the C-terminal domain and exhibit strongly anticorrelated motions with respect to residues in lobe I. The C-terminal domain consists of α-helices and loops exclusively, while the N-terminal domain has an α/β structure. Lobe I contains two α-helices, A (Ile3−Glu11) and B (Ile39−Ile50), and a β-sheet region (β) composed of three sequentially contiguous strands, I (Arg14−Asp20), II (Tyr24−Ile27), and III (His31−Thr34), and a separate short strand (Gly56−Ile58) connecting helix B to helix C. Lobe II includes five α-helices of $n \geq 9$ residues, D (Ala82−Ser90), E (Ala93−Met106), F (Thr115−Gln113), G (Trp126−Ala134), and H (Arg137−Thr155). Helix H is broken at residue Thr142, its first turn (Arg137−Gln141) being rather part of the loop between helices G and H. Residues Glu108−Gly113, located in the region between helices E and F, are folded in a $3_{10}$ helix. The figure is drawn using Midas software with the T4L crystal structure obtained *(25)* at 1.7 Å resolution. It will be shown below that the β-sheet region undergoes motions highly coupled, but in an opposite sense, with respect to helix E and the N-terminal part of helix H, and to a lesser extent with respect to helix G. Anticorrelated motions are also observed between these three regions in lobe II and helix B in lobe I.

α-carbon and the second on the side-chain centroid or functional group *(13)*. Empirical energy parameters corresponding to this low-resolution model were recently extracted *(9, 14)* from Brookhaven Protein Data Bank (PDB) *(30, 31)* structures. Here, the motions of the protein will be explored both on a *local* and on a *global* scale. The local scale is chosen at the residue level. First, the performance of the low-resolution model and parameters will be tested by comparing the theoretical α-carbon fluctuations with those derived from experimental temperature factors. Additionally, dihedral angle fluctuations will be explored, revealing the regions enjoying relatively high conformational mobility on a local scale. On a global scale, on the other hand, the mechanism of collective movements is explored. A strongly anticorrelated motion is shown to be effective between the two domains of T4L, involving up to ±4.0 Å displacements of the respective mass centers in an opposite sense. Structural elements appearing to play a key role in the coordination of this cooperative motion are identified, mainly the β-sheet region in the N-terminal domain and the helix E near the active-site cleft in the C-terminal domain. An important point worth noting is that all results are obtained here within computational times at least 1 order of magnitude shorter than conventional, full atomic simulations.
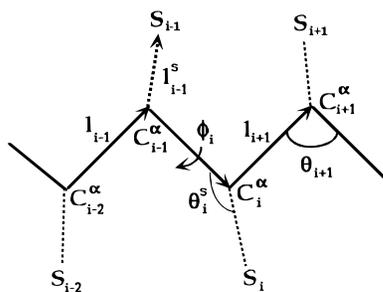
FIGURE 2: Schematic representation of the virtual bond mode. A segment between backbone units $C^\alpha_{i-2}$ and $C^\alpha_{i+1}$ is shown. Side chain attached to the $i$th α-carbon is $S_i$. $\mathbf{l}_i$ is the $i$th virtual bond connecting $C^\alpha_{i-1}$ to $C^\alpha_i$. $\phi_i$ is the rotational angle of the $i$th virtual bond, defined by the relative positions of the four backbone atoms $C^\alpha_{i-2}$, $C^\alpha_{i-1}$, $C^\alpha_i$, and $C^\alpha_{i+1}$. $\theta_i$ is the bond angle between virtual bonds $i$ and $i+1$. $\theta_i^s$ is the bond angle between $\mathbf{l}_i$ and $\mathbf{l}_i^s$ where $\mathbf{l}_i^s$ is the side-chain virtual bond vector pointing from $C^\alpha_i$ to $S_i$. The side-chain virtual bonds are shown as dashed lines. $\phi_i^s$ is the torsional angle about bond $\mathbf{l}_i$ with reference to the four consecutive atoms $C^\alpha_{i-2}$, $C^\alpha_{i-1}$, $C^\alpha_i$, and $S_i$.

## OVERVIEW OF MODEL AND SIMULATION CHARACTERISTICS

A given configuration is represented by $N-1$ backbone virtual bonds $\mathbf{l}_i$ connecting α-carbons $i-1$ and $i$, and $N$ side-chain virtual bonds $\mathbf{l}_i^s$ connecting backbone and side-chain interaction sites of each residue, as illustrated in Figure 2. The atoms used for defining the side-chain interaction centers for each type of amino acid are presented in our previous work *(13)*. The set of generalized coordinates $\{\theta_2, \theta_3, ..., \theta_{N-1}, \phi_3, \phi_4, ... \phi_{N-1}\}$, where $\theta_i$ is the bond angle between $\mathbf{l}_i$ and $\mathbf{l}_{i+1}$, and $\phi_i$ is the torsional angle of bond $\mathbf{l}_i$, characterizes a backbone configuration, all backbone bond lengths being of length $3.81 \pm 0.03$ Å. Side-chain conformation, on the other hand, is conveniently expressed by the set $\{\mathbf{l}_i^s, \theta_i^s, \phi_i^s\}$, $\theta_i^s$ being the angle between $\mathbf{l}_i$ and $\mathbf{l}_i^s$, and $\phi_i^s$ the torsion angle defined by $\mathbf{l}_{i-1}$, $\mathbf{l}_i$, and $\mathbf{l}_i^s$.

The crystal structure of T4L (PDB entry 3LZM) is utilized as the starting configuration. A series of low-resolution conformations, or substates in the neighborhood of the native state, are generated using a Monte Carlo (MC) scheme coupled with a Metropolis algorithm. Two different schemes are adopted for generating substates, which permit us to improve the accuracy and verify the reproducibility of the results. In the first, at each "step", hereafter referred to as "MC step", all $\phi_i$, $\phi^s_i$, and $\theta_i$ values are simultaneously subjected to differential changes, while in the second, the position vectors of residues are perturbed, one by one, using a random number generator subroutine. The change $\Delta x$ in the generalized coordinate $x$ is chosen according to the formula $\Delta x/x = k(x)(2r-1)$, where $r$ is the random variable uniformly distributed in the range $0 \le r \le 1$, and $k(x)$ is a constant damping factor that may be adjusted to mimic the response at a given temperature. The new configuration is accepted or rejected depending on the energy difference with respect to the preceding configuration, following a Metropolis criterion. The energy at each step is found using database-extracted residue-specific potentials composed of two major contributions: (i) nonlocal interactions including all side-chain−side-chain (S−S), side-chain−backbone (S−B), and backbone−backbone (B−B) pairs separated by at least five virtual bonds *(9)* and (ii) bond rotation and bond angle distortion energies *(14)* for pairwise-coupled virtual bonds.

Global characteristics are extracted from three independent runs, each of length $\ge 90\,000$ MC steps, whereas runs of 3000 MC steps are sufficient to obtain reproducible results on a local scale. The rms deviation between the instantaneous interresidue distances and those of the X-ray structure generally increase to about 2.0 Å within the first 1000 steps and then exhibit steady fluctuations remaining in the range $4.0 \pm 2.0$ Å throughout the simulations. These may be viewed as substates in the vicinity of the native state, within the resolution of the present coarse-grained simulations. In simulations, occasional departures from the crystal structure followed by the folding back to the original state are observed, as illustrated in Figure 3. Details of the model, simulation method and parameters are presented in the Appendix.

## RESULTS ON THE LOCAL-SCALE DYNAMICS OF T4L

*Fluctuations of Backbone Atoms: Comparison with Experimental Temperature Factors.* In Figure 4, we present the mean-square fluctuations $\langle \Delta \mathbf{R}_i \cdot \Delta \mathbf{R}_i \rangle$ in the positions of backbone sites observed in simulations as a function of residue index. This has been presented as an average over the two categories of simulations, perturbing the angles and the positions. For comparison, we also display as the dotted curve the experimental results from experiments *(25)*. Experimental $\langle \Delta \mathbf{R}_i \cdot \Delta \mathbf{R}_i \rangle$ values are evaluated from X-ray temperature factors $B_i$ using the relation $\langle \Delta \mathbf{R}_i \cdot \Delta \mathbf{R}_i \rangle = 3B_i/8\pi^2$. Here $B_i$ refers to the average over the $i$th residue backbone atoms. There is good agreement between experimental and theoretical results despite the use of a low-resolution model and a computationally simple and fast method. In fact, the theoretical curve is obtained here by using only 2 h of CPU time on a Silicon Graphics R4400 workstation, which is about 2 orders of magnitude faster than MD simulations.

The major secondary structural elements of the protein are indicated by the boxes in the lower portion of the figure. Examination of the theoretical curve in connection with these secondary structure regions shows that minima generally coincide with α-helices. A broad minimum is observed at the 21-residue helix C that connects the two domains. This region exhibits rather small amplitude fluctuations in both experiments and simulations. Likewise, the helix E located at the cleft between the two domains is distinguished by a well-defined minimum. A strong cooperativity between adjacent residues is inferred from the smooth shape of the minimum at helix E. Other regions of reduced flexibility indicated by simulations are helices B and A in the N-terminal domain and the inner residues of all helices in the C-terminal domain. We note that the minima for helices D−G and for the C-terminal part of helix H cannot be clearly distinguished in the experimental curve, in contrast to results from simulations.

Among the regions exhibiting relatively large amplitude fluctuations, on the other hand, we distinguish the loop region between helices B and C, where experiments and simulations are in perfect agreement, as well as the loop between strand III of the β-sheet region and helix B. Simulations indicate a higher mobility at the β-sheet region compared to experimental results. Residues in β-strands II and III and the tight turns preceding these strands are indeed found in
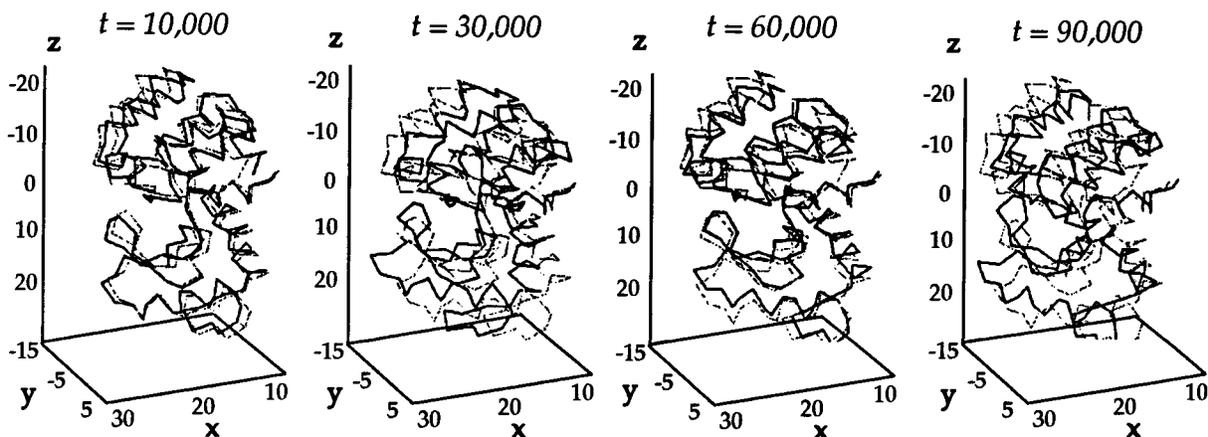
FIGURE 3: Time evolution of T4L backbone configurations. The instantaneous α-carbon traces at $t = 1 \times 10^4$, $3 \times 10^4$, $6 \times 10^4$, and $9 \times 10^4$ MC steps are displayed in black. The native configuration is also shown in each case for comparison, in gray. We note that a significant perturbation in the structure occurs at $t = 3 \times 10^4$ MC steps and is restored at $t = 6 \times 10^4$ MC steps.
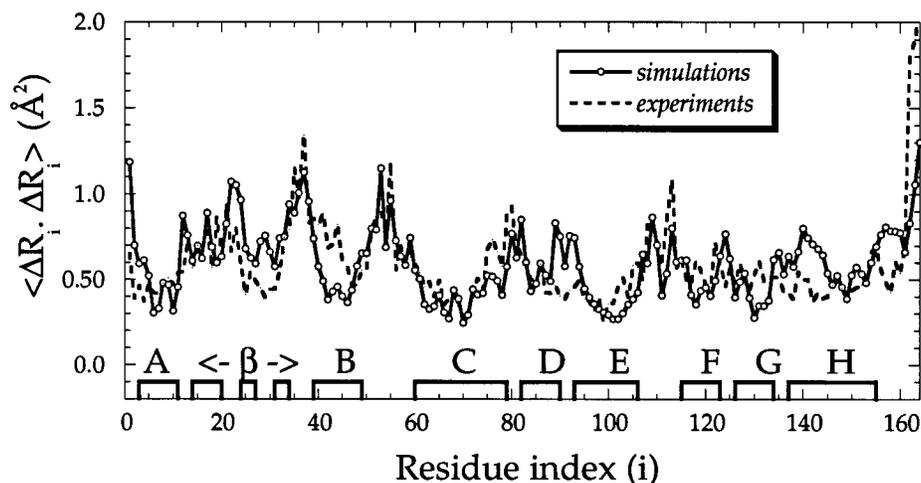


FIGURE 4: Mean-square fluctuations $\langle \Delta \mathbf{R}_i \cdot \Delta \mathbf{R}_i \rangle$ in the positions of backbone sites of T4L observed in simulations (solid curve) and experiments (dashed curve) as a function of residue index. Experimental values are evaluated using the X-ray temperature factors $B_i$ (25) in $\langle \Delta \mathbf{R}_i \cdot \Delta \mathbf{R}_i \rangle = 3B_i/8\pi^2$. The different secondary structural elements are indicated by the boxes on the lower abscissa.

simulations to exhibit strongly coupled, but anticorrelated, motions with residues of the N-terminal domain, as will be elaborated below. The N-terminal part of helix H is also found to be rather flexible, hence the peak around Tyr139−Asn144, which also deviates from experiment. We note that these residues are located at the opening of the cleft between the two domains (Figure 1), and their motions are important for the opening and closing up of the oligosaccharide binding region.

*Rotational Mobility of Backbone Bonds.* We examine the changes in the rotational angles of backbone virtual bonds of T4L in order to identify the regions having higher or lower conformational flexibility on a local scale. A good measure of the rotational flexibility of the ith $C^{\alpha}-C^{\alpha}$ bond is the autocorrelation function $\langle \cos \Delta\phi_i(\tau) \rangle$. Here the brackets refer to the average cosine of the change $\Delta\phi_i(\tau) = \phi_i(t + \tau) - \phi_i(t)$ in $\phi_i$ taken over all times $t$. Clearly, $\langle \cos \Delta\phi_i(\tau) \rangle$ decays from 1 to 0, as time proceeds; and residues that enjoy a higher rotational freedom undergo a faster loss of correlation.

Figure 5 illustrates the results obtained for a given time interval ($\tau = 1000$ MC steps) averaged from eight independent runs carried out by varying the *xyz* positions. This interval $\tau$ was selected so as to emphasize the differences between the rotational behavior of different residues. In fact, $\langle \cos \Delta\phi_i(\tau) \rangle$ values vary in the broad range [0.03, 0.96], the
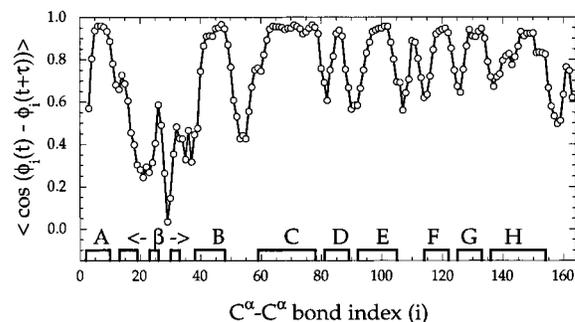


FIGURE 5: Time-delayed autocorrelations $\langle \cos \Delta\phi_i(\tau) \rangle$ for dihedral angle fluctuations $\Delta\phi_i(\tau)$ observed for backbone virtual bonds $2 \leq i \leq N$ with a time interval $\tau = 1000$ MC steps. We note the high rotational flexibility of the residues in the $\beta$-sheet, as opposed to the negligibly weak loss of correlation of inner residues in all helices.

lowest value being that of the virtual bond between Gly28 and Ile29. These residues are in the $\beta$-sheet region of lobe I, at the tight turn between strands II and III, already pointed out above. We can also distinguish that the inner portions of the three $\beta$-strands I−III are less flexible compared to the loops. The high flexibility of the $\beta$-sheet region conforms with previous atomic simulations *(23)*, in which the hairpin loop between residues 19 and 24 was noticed to act as a mobile "flap".

In the other extreme case of residues almost fully conserving their dihedral angles, we see those participating in helices, consistent with the results displayed in Figure 4. Even the short $3^{10}$ helix between Glu108 and Gly113, which is not included within the secondary structures marked on the abscissa, exhibits restricted rotational freedom. Thus, residues in $\alpha$-helices generally resist changes in dihedral angles during the observed simulation period, except for a few at the helix termini and at the N-terminal half of helix H. It is interesting to note that the restricted rotational mobilities of the helices are unaffected by the degree of solvent exposure or burial. For example, the central helix C, which is almost entirely exposed to solvent, exhibits a negligibly weak loss of correlation.

*Correlations between Dihedral Angle Fluctuations.* An interesting feature in Figure 5 is that there are no sharp peaks, minima or maxima for individual residues, in general. Instead, a strong coupling between the rotational motions of adjacent bonds is observed. Furthermore, residues belonging to the same secondary structural unit appear to undergo highly cooperative rotations. Correlations between bond rotations are further explored with the same simulations by examining the time evolution of the torsional cross-correlation function

$$C_{ij}(\tau) = \langle \Delta\phi_i(\tau)\,\Delta\phi_j(\tau)\rangle/[\langle(\Delta\phi_i(\tau))^2\rangle\,\langle(\Delta\phi_j(\tau))^2\rangle]^{1/2} \quad (1)$$

The $C_{ij}(\tau)$ are useful in providing information on the degree of correlation existing between the rotations of bonds $i$ and $j$, occurring within a time interval $\tau$ (32). It varies in the range $[-1, 1]$, the lower and upper limits corresponding to fully anticorrelated (coupled but in an opposite sense) and correlated rotations, while $C_{ij}(\tau) = 0$ refers to uncorrelated rotations.

Two important results from $C_{ij}(\tau)$ calculations for T4L are the following: First, a strong cooperativity between the rotations of the first and second neighboring bond pairs is observed, which efficiently localizes the motion and minimizes atomic fluctuations. The rotations of particular bonds, such as those in the $\beta$-sheet region, are accompanied by coupled, compensating rotations at their first or second neighbors along the chain. This strong cooperativity serves to maintain the overall integrity of the secondary and tertiary structure throughout simulations. Second, in parallel with the cooperative nature of autocorrelations along the chain, cross-correlations between residues appear in blocks, i.e., between secondary structural units as a whole. For example, similarly to previous MD observations (23), a coupling is observed between the rotations of bonds 24−34 of the $\beta$-sheet region and bonds 142−152 of helix H, although these elements belong to different lobes. Also, bonds 37−43 near the N-terminus of helix B are observed to be engaged in anticorrelated rotations with bonds 82−90 (helix D). These observations already manifest the existence of anticorrelations between the motions of the two lobes, which will become more explicit from the SVD of longer trajectories.

## GLOBAL-SCALE DYNAMICS OF T4L

*Application of SVD Technique.* The basic aim in the application of SVD to protein dynamics is to extract useful information from simulations on the collective modes of motions. This is achieved by diagonalizing the so-called configurational fluctuation matrix $\mathbf{A}$. For an $n$-step simulation, $\mathbf{A}$ is composed of the $n$ column vectors of the time-evolved coordinates of $\alpha$-carbons, expressed in terms of fluctuations with respect to native state positions. Thus, for a protein of $N$ residues, $\mathbf{A}$ is a rectangular matrix of size $3N \times n$. For clarity, we may express $\mathbf{A}$ in terms of $N$ submatrices $\mathbf{D}_i$ of size $3 \times n$ each, as

$$\mathbf{A} = \begin{bmatrix} \mathbf{D}_1 \\ \mathbf{D}_2 \\ \vdots \\ \mathbf{D}_N \end{bmatrix} \quad (2)$$

Each submatrix $1 \leq i \leq N$ represents the time evolution of a given residue, i.e.,

$$\mathbf{D}_i = \begin{bmatrix} \Delta x_{i1} & \Delta x_{i2} & \dots & \Delta x_{in} \\ \Delta y_{i1} & \Delta y_{i2} & \dots & \Delta y_{in} \\ \Delta z_{i1} & \Delta z_{i2} & \dots & \Delta z_{in} \end{bmatrix} \quad (3)$$

Here $\Delta x_{ij}$, $\Delta y_{ij}$, and $\Delta z_{ij}$ represent the deviations in the $x$-, $y$-, and $z$-components of the position vector of residue $i$ with respect to its native-state position, at MC step $j$. Information on the local motion of the $i$th residue is determined from the SVD analysis of $\mathbf{D}_i$. Identification of global motions on the other hand, necessitates having the SVD of the complete configurational fluctuation matrix $\mathbf{A}$.

The utility of SVD may be assessed by considering its application to $\mathbf{D}_i$ as an example. Upon SVD, $\mathbf{D}_i$ is readily transformed (33) as

$$\mathbf{D}_i = \mathbf{U}_i\Lambda_i\mathbf{D}_i' \equiv [\mathbf{u}_{i1}\ \ \mathbf{u}_{i2}\ \ \mathbf{u}_{i3}] \begin{bmatrix} \lambda_{i1} & 0 & 0 \\ 0 & \lambda_{i2} & 0 \\ 0 & 0 & \lambda_{i3} \end{bmatrix} \mathbf{D}_i' \quad (4)$$

where $\mathbf{D}_i'$ is another representation of $\mathbf{D}_i$, in a coordinate system whose basis vectors are $\mathbf{u}_{i1}$, $\mathbf{u}_{i2}$, and $\mathbf{u}_{i3}$. These vectors are also referred to as the principal axes or left singular vectors (LSVs) of $\mathbf{D}_i$. $\mathbf{u}_{i1}$ defines the direction along which the $i$th residue undergoes the largest amplitude motion, $\mathbf{u}_{i3}$ refers to the direction of the smallest motion perpendicular to $\mathbf{u}_{i1}$, and $\mathbf{u}_{i2}$ completes the orthonormal system. The three singular values $\lambda_{i1}$, $\lambda_{i2}$, and $\lambda_{i3}$, organized in descending order, provide a measure of the amplitude of motion along the three respective directions. Calculations performed for T4L using eq 4 showed that $\lambda_{i1}$ values plotted against residue index $i$ yield a distribution curve that is almost indistinguishable from the $\langle\Delta\mathbf{R}_i\cdot\Delta\mathbf{R}_j\rangle$ curve displayed in Figure 4. This reveals that the mean-square fluctuation amplitudes observed in simulations reflect almost exclusively the displacement of residues along their preferred direction $\mathbf{u}_{i1}$, the contribution from displacements along the two normal directions $\mathbf{u}_{i2}$ and $\mathbf{u}_{i3}$ being negligibly small.

Similarly to the transformation of $\mathbf{D}$, a new matrix $\mathbf{V}^{\mathrm{T}}$ is obtainable from the SVD of $\mathbf{A}$, i.e., $\mathbf{A} = \mathbf{U}\,\Sigma\,\mathbf{V}^{\mathrm{T}}$. $\mathbf{V}^{\mathrm{T}}$ reflects the time evolution of the *collective coordinates* of the protein in a new $3N$-dimensional space, the axes of which (columns of $\mathbf{U}$) are automatically selected along the principal directions of motion. The first column of $\mathbf{U}$, also called the first molecule-optimized dynamic principal axis, is a basis vector found from a least-squares fit to the dominant direction of motion. The singular value $\sigma_i$ ($i$th element of $\Sigma$) reflects the extent of overall molecular motion along the $i$th principal

axis. A common practice for assessing the mechanism of global motions is to observe the trajectory projected over a few, one to three, dominant principal directions. Since the $i$th row of $\mathbf{V}^{\mathrm{T}}$ represents the time-evolved collective coordinates along the $i$th principal direction, examination of the first few rows of $\mathbf{V}^{\mathrm{T}}$ directly yields information on the trajectory along the dominant principal directions. More interestingly, one may reconstruct $\mathbf{A}$ on the basis of a few ($p$) largest singular values, in order to gain an understanding of the contribution of the dominant, slowest modes to the overall dynamics of the molecule, having eliminated the uninteresting modes *(34)*. In this case, it suffices to recalculate $\mathbf{A}$, or its reduced form $\mathbf{A}'$, from $\mathbf{A}' = \mathbf{U} \Sigma \mathbf{V}^{\mathrm{T}}$ after assigning the value zero to all singular values $\sigma_i$ in the range $p < i < 3N$.

SVD analyses of the T4L trajectories obtained by both simulation methods indicate that the first three singular values amount to 65%, approximately, of the whole spectrum, i.e., $(\sigma_1 + \sigma_2 + \sigma_3)/\Sigma\sigma_i \approx 0.65$, the summation of the denominator being over the complete set of $3 \times 164$ singular values. On the other hand, in the case of correlations between atomic fluctuations, the effect of the dominant singular values is even stronger, as the ratio $\sigma_k^2/\Sigma\sigma_i^2$ should then be taken into consideration for an assessment of the role of the $k$th mode. In this case, the contribution of the first mode alone was found to increase to about 0.80. In light of these observations, we direct our attention to the effect of the first three dominant modes of motion on the dynamics of T4L. In the following, results are presented from simulations where the *xyz* positions are perturbed.

*Structural Elements Undergoing Correlated, Uncorrelated, and Anticorrelated Fluctuations.* The correlation between the fluctuations of residues $i$ and $j$ driven by the first dominant mode of motion is given by

$$C'_{ij} \equiv \langle [\Delta \mathbf{R}_i'(t) \cdot \Delta \mathbf{R}_j'(t)]/[|\Delta \mathbf{R}_i'(t)||\Delta \mathbf{R}_j'(t)|] \rangle \quad (5)$$

Here the prime refers to the configurations reconstructed using the first mode only, the angle brackets represent the average over all time steps $t$, and the denominator displays the magnitudes of fluctuation. $C'_{ij}$ is the $ij$th element of the cross-covariance matrix $\mathbf{C}'$. The cross-covariances between all residue pairs may be conveniently displayed in the form of a correlation map. The results for T4L are presented in Figure 6. The lower triangular part displays the correlation map for residue pairs $(i, j)$ whose cross-covariance lies in the range $0.4 \leq C'_{ij} \leq 1.0$. Five equally distant contour levels are shown. The regions enclosed by the innermost contours refer to residue pairs whose fluctuations are very strongly ($C'_{ij} \geq 0.9$) correlated. The upper triangular part of the map, on the other hand, displays the pairs subject to anticorrelated fluctuations, $-0.4 \geq C'_{ij} \geq -1.0$. Again, innermost regions refer to residue pairs undergoing strongly coupled, but this time in an opposite sense, fluctuations. Regions not appearing in either the lower or upper triangular parts of the map refer simply to uncorrelated or weakly correlated residue pairs.

First, let us consider the lower triangular portion of the map, i.e., examine the regions exhibiting positively correlated fluctuations. In the first place, we see that the diagonal and near-diagonal regions are generally occupied, i.e., near neighboring residues along the backbone undergo correlated motions, as expected from chain connectivity. A strong
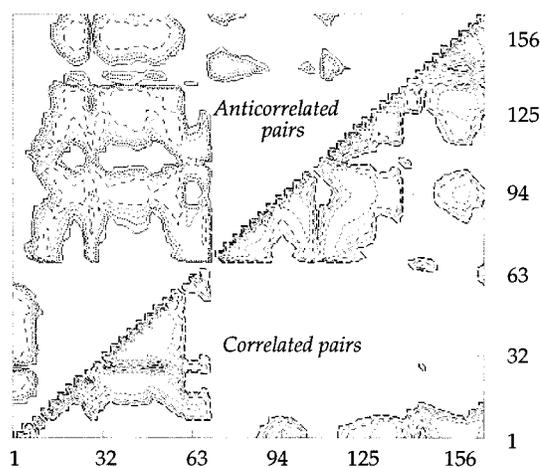


FIGURE 6: Cross-covariance maps reflecting the coupling between the motions of main-chain units $C^{\alpha}_i$ and $C^{\alpha}_j$, driven by the first singular mode of the collective motion of all residues. $i$ and $j$ are the residue indices indicated on the two axes. The results are obtained by using eq 1 from the trajectories reconstructed on the basis of the first SV only. The contours in the diagonal and lower triangular parts refer to the residue pairs whose cross-covariance $C'_{ij}$ lies in the range $0.4 \leq C'_{ij} \leq 1.0$. These are the pairs whose dominant mode of motion is strongly coupled and synchronous. The upper triangular part refer to the pairs subject to anticorrelated fluctuations in atomic positions, $-0.4 \geq C'_{ij} \geq -1.0$. We note that no pair of residues from within lobe I appears in the upper triangular portion, indicating that the motions within lobe I are strongly coherent. Likewise, there are no anticorrelated pairs within the C-terminal part of the chain. See Figure 1 and its caption for the description of structural elements undergoing strongly correlated/anticorrelated motions.

exception is apparent in the neighborhood of L66, between K65 and F67, where a discontinuity is observed along the main diagonal. These residues may be viewed as setting the boundaries of the two intradomain motions along the chain contour. In fact, we note in Figure 6 that residues 11 $\leq i \leq 66$ form a unified, self-correlated block with extensive intradomain correlations, while those in the range $i \geq 67$ form an independent second group of correlated intradomain fluctuations. The former range includes the entire lobe I, except the A-helix at the N-terminus. Helix A is closely coupled to lobe II, and in particular to helices E, G, and H, as may be verified from the contours at the bottom-right part of the map. The second block ($i \geq 67$), on the other hand, comprises several strongly correlated pairs of structural units. For example, helix D is coupled to helices E and F; helix H, on the other hand, is coupled to helices E and G. We note that in this respect helix E plays a mediating role in propagating the correlated motions within lobe II.

Second, we turn our attention to structural units exhibiting *anticorrelated* motions, shown in the upper triangular part of Figure 6. The most striking observation is that anticorrelated motions are generally *interdomain* motions; i.e., they involve residue pairs belonging to different domains. This is visible from the high occupancy of the upper left quadrant of the map. Again K65−Phe67 may be distinguished as setting the boundary between anticorrelated fluctuations. In the N-terminal part of the molecule, only helix A exhibits anticorrelated motions with respect to residues belonging to lobe I. The distinct behavior of helix A is understandable in view of its position in space nearer to the C-terminal domain of T4L. This may be seen from Figure 1. A closer examination indicates that the anticorrelated motion of helix A with respect to lobe I is effectively stopped at residues

Asp10 and Glu11, located at its C-cap. These two residues are positively correlated with the N-terminal part of helix C (up to Phe67), and simultaneously, negatively correlated with the remaining part of helix C and overall helix E. Helices C and E play in fact a key role in monitoring the concerted fluctuations of the two lobes. Within lobe II, on the other hand, we may observe some regions exhibiting intradomain anticorrelated fluctuations. In particular, the tight turn at Thr142 disrupting the helix H emerges as a region with a unique behavior. This residue and its near neighbors (Glu141, Asn144, and Arg145) are relatively decoupled from other residues in the same lobe. This region, together with the carboxy terminus of the molecule, exhibits an anticorrelated motion with respect to the interdomain helix C and the $3^{10}$ helix 108−113 connecting helices E and F. The hairpin loop Met106−Glu108 at the C-cap of helix E is also distinguished by uncorrelated or weakly correlated motions, in general. Thus, the motions of the units in lobe II are not as coherent as those of lobe I, which is readily seen in Figure 6 by the uninterrupted block of correlated residue pairs in lobe I in contrast to several interruptions in lobe II.

Finally, the structural units acting as nuclei for the anticorrelated motion of the two domains may be identified from the innermost regions of the contours in the upper triangular part of Figure 6. These are the $\beta$-sheet region and helices A and B in lobe I, and in lobe II the helix E, the N-terminal part of helix H, and to a lesser extent, helix G. More specifically, Ile17−Asp20 in strand I, Tyr25−Ile29 in strand II, Ala42 in helix B, Leu7 in helix A, Arg95 and Ala98−Met102 in helix E, and Thr152 in helix H are found in the present simulations to be the residues exhibiting the strongest correlated/anticorrelated motions with other residues belonging to the same/opposite domain.

*Concerted Motion of the Two Lobes.* The mechanism of the global motion of the two domains of T4L may be elucidated by analyzing the time evolution of their centroids, as driven by a few slowest modes of motion. For this purpose, we reconstructed the trajectories of the centroids using the first three dominant modes. The vectors $\Delta\mathbf{R}_i'(t)$, $3 \leq i \leq 59$, are averaged for defining the fluctuation $\Delta\mathbf{r}_I(t)$ of the centroid of the first lobe at a given time. Likewise, that of the second lobe, denoted as $\Delta\mathbf{r}_{II}(t)$, is estimated from $\Delta\mathbf{R}_i'(t)$ values of $\alpha$-carbons $80 \leq i \leq 156$. The terminal residues whose X-ray coordinates are relatively less precise and the helix C bridging the two lobes are not included in these two groups. The resulting trajectories are displayed in Figure 7. Panels a−c in Figure 7 represent the time evolution of the Cartesian components of the vectors $\Delta\mathbf{R}_I(t)$ and $\Delta\mathbf{R}_{II}(t)$. The bold and plain curves refer to lobes I and II, respectively, as indicated. A strong anticorrelation between the fluctuations of the two centroids is observed. In fact, the two curves in each of the panels of Figure 7 are observed to evolve in opposite senses, with a striking symmetry covering most of the fine details of the respective time series.

Figure 7 demonstrates the presence of a concerted interdomain motion, corresponding to the opening and closing of the separation between the centroids of the two lobes. This highly concerted anticorrelated motion of the centroids is a direct manifestation of the dominant modes of motion disclosed by the SVD analysis. We note that the distance between the centroids of the two lobes departs by up to $\pm$ 4.0 Å from the equilibrium value observed in the native state.
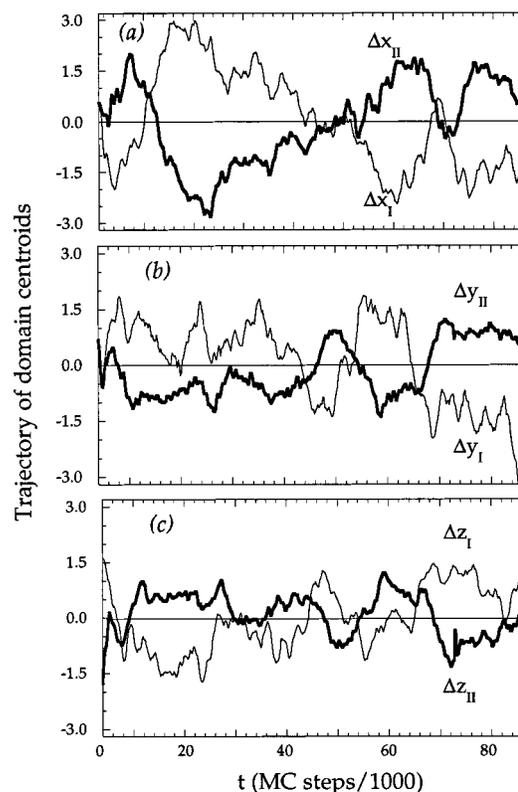


FIGURE 7: Time evolution of the position of the centroids of lobes I and II. Here the trajectories are reconstructed on the basis of the first SV. The *x*-, *y*-, and *z*-components of the fluctuation vectors $\Delta\mathbf{R}_I(t)$ and $\Delta\mathbf{R}_{II}(t)$ are shown. The two domains are observed to undergo fluctuations in opposite directions. Lobe I leads the concerted motion, with lobe II being delayed by 2000 MC steps, approximately.

Largest amplitude fluctuations, up to $\pm$ 3.2 Å, are observed in the *x*-direction. The fluctuations in the *y*- and *z*-directions are about 2 times smaller in amplitude. Thus the anticorrelated motion of the 2 domains does not correspond to an opening and closing of the cleft by the displacements of the two domains along the longitudinal axis of the molecule but involves both normal and shear components as may be verified from an examination of the orientation of the molecule with respect to the Oxyz frame in Figure 3.

Another extremely interesting feature in Figure 7 is a consistent short time delay discernible in the motion of center II compared to center I suggesting that the concerted interdomain motion is driven by lobe I and that lobe II follows in its opposing motion. The previous observation (Figure 6) that the residues within lobe I obey a more coherent motion than those of lobe II is also consistent with the dominant effect of lobe I in exerting control of the concerted motion.

*Time Dependence of Auto- and Cross-Covariances of Domain Fluctuations.* The time evolution of correlations between the fluctuations of the two lobes is presented in Figure 8. The upper part displays the time-delayed *auto*-covariances $\langle\Delta\mathbf{R}_I(t_0)\cdot\Delta\mathbf{R}_I(t_0 + t)\rangle$ and $\langle\Delta\mathbf{R}_{II}(t_0)\cdot\Delta\mathbf{R}_{II}(t_0 +t)\rangle$ obtained for the centroid of each lobe, separately, as a function of the time delay *t*. The angle brackets denote the average over all starting times $t_0$. Time along the abscissa is expressed in units of MC steps. The lower curve, on the other hand, represents the evolution of the *cross*-covariance between the two lobes, $\langle\Delta\mathbf{R}_I(t_0)\cdot\Delta\mathbf{R}_{II}(t_0 + t)\rangle$. We note that the intercept of the cross-covariance curve at $t = 0$ is $-0.70$,
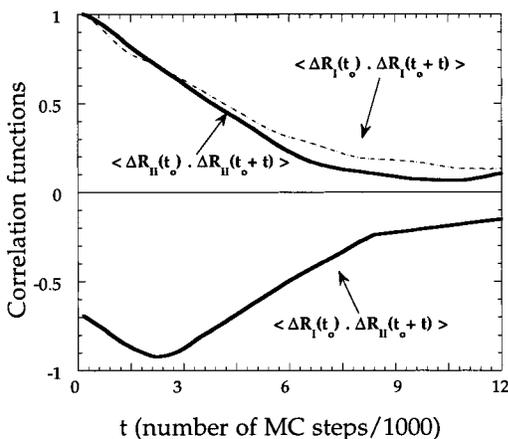
FIGURE 8: Time-delayed auto- and cross-covariances associated with the fluctuations of the centroids I and II. The curves in the upper part represent the auto-covariances $\langle \Delta \mathbf{R}_I(t_0) \cdot \Delta \mathbf{R}_I(t_0 + t) \rangle$ and $\langle \Delta \mathbf{R}_{II}(t_0) \cdot \Delta \mathbf{R}_I(t_0 + t) \rangle$ evaluated as a function of the time interval $t$, averaged over various starting times $t_0$. The lower curve represents the time dependence of the cross-covariance $\langle \Delta \mathbf{R}_I(t_0) \cdot \Delta \mathbf{R}_{II}(t_0 + t) \rangle$.

approximately. This confirms that the two lobes undergo coupled but opposite sense fluctuations at equilibrium. It is interesting to note that the degree of coupling becomes even stronger after a short time interval during which the cross-covariance curve reaches a value near −1.0. Thus, anticorrelated fluctuations become more pronounced after an initial adjustment period of 2000 MC steps. In fact, this time interval could be associated with the short time shift between the fluctuations of the two domains already noticed in Figure 7.

The auto-covariances of the two lobes shown in the upper part of Figure 8 almost coincide with each other. This confirms that their dynamics obeys the same characteristic behavior. The time dependences of the auto-covariances may be fit by a single-exponential decay with time constant of $4.5 \times 10^3$ MC steps, with a correlation coefficient of 0.98. We note that the relaxation time of this process is significantly slower than that of dihedral angle correlations. In fact, as illustrated in Figure 5, a substantial loss in dihedral angle autocorrelations was observed at 1000 MC steps.

*Potential of Mean Force Governing the Concerted Motion of the Lobes.* The distribution of the fluctuation $\Delta R_{I,II}(t)$ in the instantaneous separation of the lobe centroids is presented in normalized form in Figure 9. $\Delta R_{I,II}(t)$ is given by

$$\Delta R_{I,II}(t) = |\mathbf{R}_{I,II}(t)| - |\bar{\mathbf{R}}_{I,II}| =$$
$$|\bar{\mathbf{R}}_{I,II} + \Delta \mathbf{R}_{II}(t) - \Delta \mathbf{R}_I(t)| - |\bar{\mathbf{R}}_{I,II}| \quad (6)$$

where $\mathbf{R}_{I,II}(t)$ is the instantaneous separation between the centroids of the two lobes and $\bar{\mathbf{R}}_{I,II}$ is the separation vector pointing from A to B in the crystal structure, given by $\bar{\mathbf{R}}_{I,II} = [-6.14\ 0.26\ 23.82]$ in our reference frame Oxyz. The filled circles represent the probabilities $P(\Delta R_{I,II})$ of the successive bins in the range $[-4.0, 4.0]$ of $\Delta R_{I,II}(t)$ visited throughout the simulations. The dotted curve is obtained by direct interpolation of the data. The best-fitting unimodal distribution curve is also shown by the solid curve. The latter exhibits a peak around $\Delta R_{I,II} = -0.5$ Å. Thus, the most probable interdomain distance (peak in the probability distribution curve) is shifted to a slightly compressed state relative to the crystal structure, but the mean distance (average over the distribution curve) coincides with that of
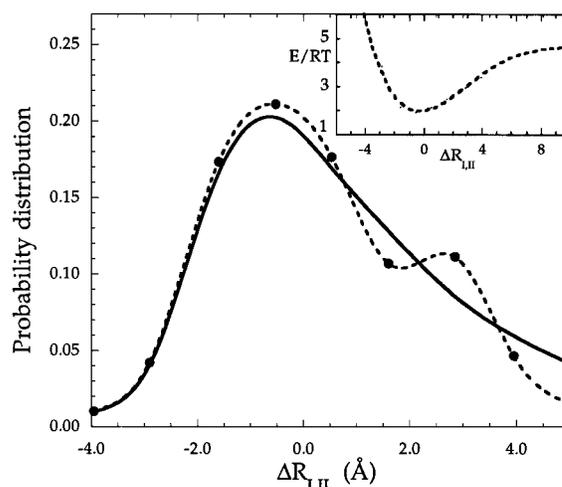


FIGURE 9: Probability distribution of fluctuations $\Delta R_{I,II}$ in the distance between the centroids of the lobes I and II. The dots represent the results from simulations joined by a dotted curve to guide the eye, and the curve is the best-fitting unimodal distribution function. The inset displays the effective potential of mean force.

the crystal structure. The difference between the mean and most probable values is a consequence of the anharmonicity of the potential of mean force describing the relative positions of the two domains.

In the inset, we display the potential of mean force associated with the anticorrelated fluctuations of the two domains. The latter is directly found from the relation $E/RT = -\ln P(\Delta R_{I,II}) + c$, where $R$ is the gas constant, $T$ is the absolute temperature, and $c$ is taken as zero here. $P(\Delta R_{I,II})$ denotes the probability of occurrence of a change $\Delta R_{I,II}$ in the distance between the two centroids. The curve displayed in the inset is obtained by fitting the results with a fourth-order polynomial, which permits extrapolation to a longer distance range. A soft anharmonic potential of mean force is obtained, with a minimum at a relatively close distance between the lobes—thus favoring the closed configuration of the cleft—but permitting the opening of the cleft upon the action of a relatively weak (∼4RT) potential of mean force.

## DISCUSSION

*Limitations and Advantages of the Low-Resolution Model.* The basic limitation of a coarse-grained model is obviously its inaccuracy on the atomic scale. Roughly, bonds of ∼1.5 Å on the atomic level are replaced here by virtual bonds of ∼3.8 Å, and it is natural to expect a commensurate decrease in the resolution of the structures. For example, a distance rms deviation of 4.0−6.0 Å from native structure is observed here, as opposed to a value of 1.5−2.0 Å in atomic MD simulations. Basically, the α-carbon trace of the protein, with a single residue-specific interaction site attached to each α-carbon, is simulated. Various atomic details are overlooked. However, a large number of studies have demonstrated the utility of residue-level approaches *(35)*. The resolution of the backbone, and the specificity of the side-chain sites adopted here, were shown in previous comparisons with experiments and in threading tests to give a realistic account of secondary structure propensities and residue-specific tertiary contact preferences *(9, 14)*. Now, the present results go a step further and suggest that information on the dynamic properties at the tertiary structure level may be

extracted from such coarse-grained simulations, provided that the trajectories are filtered with a powerful technique such as SVD. The present geometric and energetic parameters are robust enough to maintain the overall tertiary fold throughout these simulations, despite occasional departures from native-state coordinates, as illustrated in Figure 3. Possible noise and errors are minimized simply by concentrating on the dominant mode of motion. For example, the correlation map displayed in Figure 6 gives clear information on the types and strengths of correlations between different structural elements of the protein in the native state, as driven by the slowest mode of motion. MD simulations carried out for the same protein *(22, 23)* did successfully indicate that a hinge-bending motion was operative between the two lobes, in conformity with results suggested from experiments. However, these motions were not regular, and anticorrelations between the two domains could not be detected. The results of Arnold et al. *(23)* (their Figure 6) stand in strong contrast to our results (our Figure 6). This difference could be due to the absence of frictional damping in the present MC simulations, compared to MD simulations or Langevin dynamics, in which solvent interactions weaken the intramolecular correlations.

*Comparison with H/D Exchange and 2-D NMR Observations.* Throughout the present study, the importance of helix E (Ala93−Met106) to the overall stability of the molecule has been emphasized several times. This is pointed out to be a structural element that mediates the correlated motions of the two lobes. The participation of helix E among the structural elements driving the concerted motion of the two lobes suggests that helix E might be a structural element formed at early stages of folding and that it remains folded to maintain the coherence of the structure. The fact that this helix shows a high degree of protection from solvent exchange during refolding in pulsed hydrogen−deuterium exchange and two-dimensional NMR experiments *(20)* also is consistent with this view that it is formed at an early stage of folding and is a stabilizing agent in subsequent stages of folding.

The element exhibiting the next highest protection after helix E was pointed out *(20)* to be the β-sheet region. We note that Ile17−Asp20 in strand I and Tyr25−Ile29 in strand II were listed above among residues exhibiting remarkably strong correlations/anticorrelations with other residues in the same/opposite domain. It is interesting to note that the present correlations occur despite the high rotational mobility of the loop regions preceding or succeeding these strands (see Figure 5). The explanation is that bond rotations must be highly cooperative to localize efficiently the structural perturbations and minimize distortion of the secondary structure. And these rotations were observed to occur within time scales about 1 order of magnitude shorter than the global motions. Upon elimination of the fastest modes by SVD and examination of the slowest, collective modes of motions, it was indeed possible to discern the involvement of the β-strands in the cooperative dynamics of the protein. Other regions showing high protection were parts of helices A and C, which may be understood both on the basis of these making substantial tertiary contacts with helix E and the β-sheet region as well as also being highly constrained on a local scale. Our simulations suggest that helix B, the N-terminal part of helix H, and helix G might be the next group of secondary structural elements to follow after the

folding of the β-sheet and helices E, A, and C.

Another observation from pulsed hydrogen−deuterium exchange/two-dimensional NMR experiments is the occurrence of a concerted conformational motion during the folding process that embodies the whole molecule. This contrasts to the sequential folding mechanism observed for hen egg white lysozyme *(36)*, in which the two structural domains are separately and sequentially folded. Although T4L and hen egg white lysozyme exhibit some structural similarities, another significant difference pointed out by Lu and Dahlquist *(20)* is the rapid formation of the β-sheet region in T4 lysozyme; whereas the homologous β-sheet of hen egg white lysozyme shows no evidence of early structure. The presently observed strong coupling between the two lobes, and the participation of even spatially distant structural elements in anticorrelated motions, conform to the experimental findings of Lu and Dahlquist about the concerted folding of the two domains.

*Thermal Fluctuations and Correlation with Mutation Studies.* The mean-square deviations $\langle \Delta \mathbf{R}_i \cdot \Delta \mathbf{R}_i \rangle$ in the positions of backbone sites throughout the simulations were shown in Figure 4 to be in good accord with experimental thermal fluctuations. In general, the analysis of crystallographic thermal factors is complicated by the difficulty of separating the contributions from crystal packing effects *(37)*. The average thermal motions in the crystal structure of wild-type T4L show, indeed, a highly irregular distribution *(17)*. However, the thermal B factors of the backbone atoms averaged on a per-residue basis exhibit a smoother distribution, and these are quite satisfactorily accounted for by the present coarse-grained simulations.

It has been pointed out that regions of restricted mobility may generally be associated with disruptive mutation sites *(15, 18)*. It might be interesting in this respect to compare the minima of the theoretical fluctuation amplitude curve (Figure 4) with sites whose substitution has been pointed out to be deleterious in previous mutation studies. For example, substitutions at Met6, Leu33, Leu66, Leu91, Arg96, Cys97, Ala98, Met102, Val103, Glu105, Trp126, Trp138, Arg145, Ala146, Val149, Phe153, Gly156, Thr157, Trp158, and Ala160 were found to be substantially destabilizing by Alber et al. *(15)*. We note that except for Trp138 and residues $i \geq 156$ at the carboxy terminus of the chain, all these residues correspond to minima in the theoretical $\langle \Delta \mathbf{R}_i \cdot \Delta \mathbf{R}_i \rangle$ curve. The chain terminus is relatively flexible in our simulations, and the destabilizing effect of mutations at this region cannot be attributed to the rigidity of the structure. On the other hand, it is interesting to note that the theoretical curve has minima at some destabilizing residues such as Val149 and Phe153, which cannot be distinguished as regions of low mobility in the experimental curve.

Of the 2015 single amino acid substitutions carried out by Poteete and co-workers *(16)*, only a small fraction (10−15%) were found to be deleterious. The 12 positions with the highest sensitivity to mutation were reported to be Glu11, Gly30, Tyr161, Asp10, Trp138, Val149, Gly28, Ser136, Thr26, Ala98, Asp20, and Ile58. Again most of these residues are at local minima in the theoretical $\langle \Delta \mathbf{R}_i \cdot \Delta \mathbf{R}_i \rangle$ curve (Figure 4). We note that Tyr161 in particular exhibits a minimum, although it is located at the relatively flexible C-terminal end of the protein. On the other hand, Gly28 and Gly30, located at the tight turn between β-strands II and

III, exhibit some increased mobility relative to the neighboring residues in the $\beta$-sheet region. Their sensitivity to mutation can, however, be explained by the fact that the substitution of any larger size residue could not accommodate the tight turn between strands II and III and could significantly affect the structure and dynamics of these two elements, having a significant influence on the collective motion of the overall molecule.

In general, the regions of restricted mobility, manifested by the minima in Figure 4, correspond to the loci of buried residues; whereas the maxima refer to solvent-exposed residues. Yet not all solvent-exposed residues possess large conformational freedom. A typical example is the pair His31 and Asp70, which exhibit small amplitude motions, in spite of the fact that these residues are located on the surface of the protein. Their reduced mobilities may be attributed partly to the existence of a salt bridge between them. Another explanation of these discrepancies may be the presence of structural waters that are so stable in their positions as to act as if they were part of the more rigid protein structure.

*Concerted Motion of the Backbone.* Recently, Zhang et al. *(19)* invited attention to the adaptability seen in 25 crystal forms of T4L. In different crystal environments, the mutants of T4L were pointed out to display a range of over 50° in the hinge-bending angles between the amino- and carboxy-terminal domains. This wide range was suggested to be an indication of the intrinsic flexibility of T4L and not an artifact due to mutations or crystal contact perturbations. Backbone flexibility was indeed asserted to play an important role in accommodating the sequence variants of T4L *(38)*. Concerted main-chain and side-chain displacements with movements of helical segments as large as 0.8 Å were pointed out to occur in that latter study. Another test of homology modeling on A98V mutants indicates, however, that allowing the side chains to move more flexibly than strictly indicated by a side-chain rotamer library may also compensate for the backbone rigidity *(39)*. Here, we showed that fluctuations of even larger amplitude may take place, up to ±4.0 Å displacements in the centroids of the two lobes, without the breakdown of tertiary structure (Figure 7). This demonstrates the high flexibility of the backbone of T4L and its adaptability to local perturbations in solution, in conformity with X-ray crystallographic measurements of T4L mutants *(19)*. The distance between the centroids of the two lobes are observed in the present study to vary in the range $20.3 \leq |\mathbf{R}_{I,II}(t)| \leq 28.2$ Å, which is comparable to the relative domain movement between the open and closed forms of T4L recently probed by site-directed spin labeling experiments *(21)*.

A repacking of the interface between helices A and C has been pointed out to occur during the conformational change of T4L between open and closed forms. For example in the mutant I3P of T4L *(29)*, Phe4, Phe67, and Phe104 were completely repacked. Hinge points in the vicinity of residues 13, 58−60, and 79−80 are reported for this mutant, while those for mutant M6I were found near residues 7−13 and 70−75 *(29)*. Glu64 and Asp72 were also shown to undergo discrete changes in their side-chain rotamer angles $\chi_1$, suggestive of new conformers stabilized by the opening of the hinge-bending angles beyond ~15° from the wild-type value. Our coarse-grained simulations cannot evaluate such changes in side-chain rotameric states, because of the use of a single side-chain site per residue. MD simulations of

domain motions by Arnold *et al.* *(23)*, on the other hand, indicated small changes over several residues, with hinge loci comprising residues 8−14 and 77−83. Our calculations indicate that residues K65−Phe67 in helix C bridging the two lobes play a key role in the global motion of the molecule. The correlated motions within a given lobe, for example, extend up to these residues but are not transmitted into the second lobe, as may be seen from the correlation map (Figure 6). Not surprisingly, L66 has been observed to be a thermally destabilizing site *(15)*.

*Major Advantages of the Coarse-Grained SVD Approach and Future Prospects.* (i) The virtual bond model and simulation methods are simple and allow for the coarse-grained examination of collective motions within computation times from 1 to 2 orders of magnitude faster than conventional MD simulations. Thus, a systematic analysis of the collective dynamics can be performed for structurally and/or functionally similar proteins, with the aim of identifying their common or distinct mechanisms of motion on a coarse-grained scale.

(ii) The agreement between mean-square displacements and *B* factors displayed in Figure 4 shows the suitability of the short runs of $\sim$ 3000 MC steps for describing local conformational fluctuations in the absence of nonlinear modes. We note that in a recent normal mode analysis by Tirion *(40)*, a one-parameter potential energy function was shown to give a satisfactory account of the *B* factors. Our recent study, motivated by the work of Tirion, demonstrated that an even simpler analytical approach with a single parameter, which does not require a systematic normal mode analysis with energy minimization, does successfully reproduce the *B* factors *(41)*. Local fluctuations near the native state can equally well be interpreted by these alternative methods.

(iii) The information obtained by SVD reflects the effects of global nonlinear collective excitations which were pointed out *(42)* to be responsible for most of the atomic fluctuations. The extraction of the dominant mode thus provides an extremely clear picture of the intrinsic preferences of amino acids on both a local and a global scale, without the perturbations arising from the higher frequency modes.

(iv) Reconstruction of the trajectories on the basis of the dominant LSVs proved useful for identifying structural units subjected to correlated, uncorrelated, or anticorrelated fluctuations, during the cooperative conformational motions of the protein. Here, all of the secondary structural elements within the N-terminal lobe of T4L are shown to undergo highly coupled fluctuations, the exception being helix A, which acts instead as part of the C-terminal domain. As to lobe II, although most structural elements do undergo correlated motions, some units appear relatively uncorrelated and a few anticorrelated. Thus, lobe I is subject to the more coherent motions and is also observed to lead the concerted fluctuations of the two domains.

(v) Cross-covariance maps constructed on the basis of the dominant mode of motion provide important references for the identification of structural elements involved in cooperative conformational motions. The perturbations or noise imparted by the interfering modes of lesser importance are removed altogether. Effects from sequential or spatial contiguity, local packing density, and most pronounced specific interactions and conformational potentials are operative in these correlation maps.

(vi) Insights into the kinetics and mechanism of cooperative motions may be gained by the reconstruction of the trajectories on the basis of the dominant LSVs. Although the present analysis has been limited to the subspace of conformations in the neighborhood of the native structure, the constraints on the Metropolis algorithm may be relaxed, or the perturbations intensified by adjustment of the control parameter $k(x)$, so as to explore the unfolding pathway, as well. Yet it should be recalled that the potentials of mean force employed in the present study are derived from compact, folded structures as the reference state, and their extension to unfolded structures might necessitate the incorporation of solvent−side-chain interactions or the use of solvent-mediated interresidue potentials *(5, 9)*, as recently discussed *(35)*.

## APPENDIX

*Configurational Energies in the Coarse-Grained Model.* The energy of a given configuration is evaluated on the basis of two types of interactions. The former relates to the interactions between the nonbonded sites separated by at least three intervening residues along the protein backbone. These are referred to as *long-range* (L) interaction*s*. The term "long" refers to the distance along the chain sequence, only, and does not refer to the actual distance in space. *Short-range* (S) interactions, on the other hand, are those associated with the nearest (along the main chain) covalently linked units. The total energy of a given configuration $\{\Phi\}$ is thus expressed as a sum of two terms, $E_L\{\Phi\}$ and $E_S\{\Phi\}$, accounting for the long-range and short-range interactions as

$$E\{\Phi\} = E_L\{\Phi\} + E_S\{\Phi\} \tag{7}$$

$E_L\{\Phi\}$ comprises three contributions given by

$$E_L\{\Phi\} = \sum_{i=1}^{N-3}\sum_{j=i+3}^{N} W_{SS}(r_{ij}) + \sum_{i=1}^{N-4}\sum_{j=i+4}^{N} W_{SB}(r_{ij}) + \\ \sum_{i=1}^{N-5}\sum_{j=i+5}^{N} W_{BB}(r_{ij}) \tag{8}$$

Here, $W_{SS}(r_{ij})$ denotes the potential of mean force between side groups $S_i$ and $S_j$, $r_{ij}$ being the distance between these groups in configuration $\{\Phi\}$, $W_{SB}(r_{ij})$ is the potential between the side group and the backbone interaction centers of the $i$th and $j$th residues, and $W_{BB}(r_{ij})$ is that occurring between the two long-range backbone units $C^\alpha_i$ and $C^\alpha_j$. On the other hand, the short-range conformational energy $E_S\{\Phi\}$ is found from

$$E_S\{\Phi\} = \sum_{i=2}^{N-1} E(\theta_i) + \\ \left[\sum_{i=3}^{N-1}\frac{1}{2}[E(\phi^-_i) + E(\phi^+_{i-1})] + \sum_{i=3}^{N-2}\Delta E(\phi^-_i, \phi^+_i)\right] + \\ \sum_{i=3}^{N-1}[\Delta E(\theta_i, \phi^-_i) + \Delta E(\theta_{i-1}, \phi^+_{i-1})] \tag{9}$$

Here, the first summation is the energy associated with the distortion of the backbone virtual bond angles, the second term (in brackets) refers to the torsions of backbone virtual

bonds, and the last term accounts for the coupling between these bond angles and torsions. The notation $E(\phi^-_i)$ and $E(\phi^+_i)$ is adopted for the torsional energies of the virtual bonds preceding and succeeding the $i$th $\alpha$-carbon, respectively. Conforming with the conventional terminology of polymer statistics *(43, 44)*, the energy terms with a single argument in eq 9, $E(\phi^+_i)$, $E(\phi^-_i)$, and $E(\theta_i)$, are referred to as the *first-order* interaction energies, and the remainder, $\Delta E(\phi^-_i, \phi^+_i)$, $\Delta E(\theta_i, \phi^+_i)$, and $\Delta E(\theta_i, \phi^-_i)$ are the *second-order* interaction energies. The latter group corrects for the departure from the additive contribution of the energies associated with the individual degrees of freedom, thus accounting for the important pairwise interdependence of the geometric variables. All energy terms appearing in the right-hand sides of eqs 8 and 9 are residue-specific, except for $W_{BB}(r_{ij})$.

The nonbonded potentials of mean force $W_{SS}(r_{ij})$, $W_{SB}(r_{ij})$, and $W_{BB}(r_{ij})$ and the short-range energies $E(\phi^+_i)$, $E(\phi^-_i)$, $E(\theta_i)$, $\Delta E(\phi^-_i, \phi^+_i)$, $\Delta E(\theta_i, \phi^+_i)$, and $\Delta E(\theta_i, \phi^-_i)$ are taken from our recent studies *(9, 14)*. The nonbonded potentials are given therein at 0.4 Å intervals in the range $2.0 \le r_{ij} \le 12.4$ Å, the lower and higher limits corresponding to the boundaries of the infinite repulsion and negligibly weak interaction regions, respectively. The bond angle distortion and bond torsion energies are given at intervals of $\Delta\theta_i = 10°$ and $\Delta\phi^\pm_i = 30°$ in the ranges $60° \le \theta_i \le 180°$ and $-180° \le \phi^\pm_i \le 180°$.

*Simulation Method.* All coordinate and vector quantities are expressed in a laboratory-fixed frame Oxyz, with the origin at the first $\alpha$-carbon, the $x$-axis along the first virtual backbone bond; the two N-terminal virtual bonds are located in the $xy$-plane, and the $z$-axis completes a right-handed coordinate system, in conformity with the conventional approach of polymer statistics *(43)*. This representation, also repeated for chain-embedded local frames, allows the fast generation of new configurations using serial multiplication of bond-based transformation matrices *(44)*. There is a total of 316 virtual bonds in T4L, consisting of 163 main-chain and 153 side-group bonds excluding glycines, denoted as $l_i$ and $l_i^s$, $2 \le i \le 164$. The energy of the native structure, $E\{\Phi\}_0$, is found from eqs 7−9 to be $-5.3NRT$. This is in accord with typical values obtained for 302 native proteins *(9)*.

In simulations, two methods were adopted for perturbing the molecular structure. In the first, bond dihedral angles were all simultaneously changed by small increments $\Delta x$, following the formula $\Delta x/x = k(x)(2r - 1)$, where $r$ is a random number uniformly distributed in the range $0 < r \le 1$, and $k(x)$ was taken as 0.003 for $x = \phi_i$ and $\theta_i$ and 0.0015 for $x = \phi_i^s$. In the second, atomic coordinates of randomly selected backbone sites were changed one by one by small increments, using the formula $\Delta x = k(x)(2r - 1)$, with $k(x) = 0.8$. These perturbations are verified to generate a trajectory with a distance rms deviation from the X-ray structure remaining below 7.0 Å in simulations. A new conformation $\{\Phi\}$ is accepted if the Metropolis criterion $\exp[-(E\{\Phi\} - E\{\Phi\}_0)/\tau] \le r$ is satisfied, and rejected otherwise, $\tau$ being taken as $0.3RT$. This serves to eliminate substates having an energy more than $4.0RT$ above the native (initial) state. The fraction of accepted moves, and consequently the size of the motions, may be readily controlled by modifying the parameter $\tau$. The acceptance ratio using $\tau = 0.3RT$ was ~55%.

Identification of time in this type of simulation is not completely straightforward, in contrast to conventional MD or Langevin dynamics. Here the MC/Metropolis method was preferred due to its simplicity and efficiency. There is a resemblance of the present approach to lattice dynamics (*45–47*), where a small set of fixed transitions on a lattice are permitted, *e.g.*, a flip of bonds on one side of a square to the opposite side of the square. In the present case as well as in those simulations, the only identification of time is with the number of steps, which can be only qualitatively correct. It is not an exact relationship, especially for lattice dynamics where the individual steps could have different unit times. In one way the present simulation based on changes in virtual bond angles and torsion angles involves a simpler and more consistent set of unit transitions. The number of steps in the present calculations should consequently be more directly relatable to the time scale in MD simulations.

## REFERENCES

1. Elber, R., and Karplus, M. (1987) *Science 235*, 318−321.
2. Clarage, J. B., Romo, T., Andrews, B. K., Pettitt, B. M., and Phillips, J. G. N. (1995) *Proc. Natl. Acad. Sci. U.S.A. 92*, 3288−3292.
3. Dill, K. A., Bromberg, S., Yue, K., Fiebig, K. M., Yee, D. P., Thomas, P. D., and Chan, H. S. (1995) *Protein Sci. 4*, 561−602.
4. Miyazawa, S., and Jernigan, R. L. (1985) *Macromolecules 18*, 534−552.
5. Miyazawa, S., and Jernigan, R. L. (1996) *J. Mol. Biol. 256*, 623−644.
6. Jernigan, R. L. (1992) *Curr. Opin. Struct. Biol. 2*, 248−256.
7. Hinds, D. A., and Levitt, M. (1994) *J. Mol. Biol. 243*, 668−682.
8. Park, B. H., and Levitt, M. (1995) *J. Mol. Biol. 249*, 493−507.
9. Bahar, I., and Jernigan, R. L. (1997) *J. Mol. Biol. 266*, 195−214.
10. Golub, G. H., and Van Loan, C. F. (1989) *Matrix Computations*. Johns Hopkins University Press, Baltimore, MD.
11. Garcia, A. E., and Harman, J. G. (1996) *Protein Sci. 5*, 62−71.
12. Romo, T. D., Clarage, J. B., Sorensen, D. C., and Phillips, G. N. (1995) *Proteins: Struct., Funct., Genet. 22*, 311−321.
13. Bahar, I., and Jernigan, R. L. (1996) *Folding Des. 1*, 357−370.
14. Bahar, I., Kaplan, M., and Jernigan, R. L. (1997) *Proteins: Struct., Funct., Genet.* (in press).
15. Alber, T., Dao-pin, S., Nye, J. A., Muchmore, D. C., and Matthews, B. W. (1987) *Biochemistry 26*, 3754−3758.
16. Rennell, D., Bouvier, S. E., Hardy, L. W., and Poteete, A. R. (1991) *J. Mol. Biol. 222*, 67−87.
17. Matthews, B. W. (1993) *Annu. Rev. Biochem. 62*, 139−160.
18. Matthews, B. W. (1995) *Adv. Protein Chem. 46*, 249−278.
19. Zhang, X.-J., Wozniak, J. A., and Matthews, B. W. (1995) *J. Mol. Biol. 250*, 527−552.
20. Lu, J., and Dahlquist, F. W. (1992) *Biochemistry 31*, 4749−4756.
21. Mchaourab, H., Oh, K. J., Fang, C. J., and Hubbell, W. L. (1997) *Biochemistry 3*, 307−316.
22. Arnold, G., and Ornstein, R. L. (1994) *Proteins: Struct., Funct., Genet. 18*, 19−33.
23. Arnold, G. E., Manchester, J. I., Townsend, B. D., and Ornstein, R. L. (1994) *J. Biomol. Struct. Dyn. 12*, 457−475.
24. Matthews, B. W., and Remington, S. J. (1974) *Proc. Natl. Acad. Sci. U.S.A. 71*, 4178−4182.
25. Weaver, L. H., and Matthews, B. W. (1987) *J. Mol. Biol. 193*, 189−199.
26. Matthews, B. W. (1987) *Biochemistry 26*, 6885−6888.
27. Anderson, W. F., Grütter, M. G., Remington, S. J., Weaver, L. H., and Matthews, B. W. (1981) *J. Mol. Biol. 147*, 523−543.
28. Faber, H. R., and Matthews, B. W. (1990) *Nature 348*, 263−266.
29. Dixon, M. M., Nicholson, H., Shewchuck, L., Baase, W. A., and Matthews, B. W. (1992) *J. Mol. Biol. 227*, 917−933.
30. Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F. J., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T., and Tasumi, M. (1977) *J. Mol. Biol. 112*, 535−542.
31. Abola, E. E., Bernstein, F. C., Bryant, S. H., Koetzle, T. F., and Weng, J. (1987) in *Crystallographic Databases-Information Content Software Systems, Scientific Applications* (Allen, F. H., Bergerhoff, G., and Sievers, R., Eds.). p 107, Data Commission of the International Union of Crystallography, Bonn, Germany, and Cambridge and Chester, England.
32. Haliloglu, T., Bahar, I., Erman, B., G., K. E., and Mattice, W. L. (1996) *J. Chem. Phys. 104*, 4828−4834.
33. Press, W. H., Teukolski, S. A., Vetterling, W. T., and Flannery, B. P. (1992). *Numerical Recipes in Fortran*. Cambridge University Press, Cambridge England.
34. Amadei, A., Linssen, A. B. M., and Berendsen, H. J. C. (1993) *Proteins: Struct., Funct., Genet. 17*, 412−425.
35. Jernigan, R. L., and Bahar, I. (1996) *Curr. Opin. Struct. Biol. 6*, 195−209.
36. Miranker, A., Radford, S. E., Karplus, M., and Dobson, C. M. (1991) *Nature 349*, 633−636.
37. Artymiuk, P. J., Blake, C. C. F., Grace, D. E. P., Oatley, S. J., Phillips, D. C., and Sternberg, M. J. E. (1979) *Nature (London) 280*, 563−568.
38. Baldwin, E. P., Hajiseyedjavadi, O., Baase, W. A., and Matthews, B. W. (1993) *Science 262*, 1715−1717.
39. Lee, C. (1996) *Folding Des. 1*, 1−12.
40. Tirion, M. M. (1996) *Phys. Rev. Lett. 77*, 1905−1908.
41. Bahar, I., Atilgan, A. R., and Erman, B. (1997) *Folding Des. 2*, 173−181.
42. Garcia, A. E. (1992) *Phys. Rev. Lett. 68*, 2696−2699.
43. Flory, P. J. (1969) *Statistical Mechanics of Chain Molecules*, Interscience, New York.
44. Mattice, W. L., and Suter, U. W. (1994). *Conformational Theory of Large Molecules. The Rotational Isomeric State Model in Macromolecular Systems*, John Wiley & Sons, Inc., New York.
45. Covell, D. G. (1994) *J. Mol. Biol. 235*, 1032−1043.
46. Kolinski, A., and Skolnick, J. (1994) *Proteins: Struct., Funct., Genet. 18*, 338−352.
47. Kolinski, A., and Skolnick, J. (1994) *Proteins: Struct., Funct., Genet. 18*, 353−366.