ELSEVIER

Polymer 43 (2002) 431–439

polymer

www.elsevier.com/locate/polymer

Collective deformations in proteins determined by a mode analysis of molecular dynamics trajectories ☆

Pemra Doruker[a], Ivet Bahar[a], Canan Baysal[b], Burak Erman[b,*]

[a]*Center of Computational Biology & Bioinformatics, and Department of Molecular Genetics & Biochemistry, School of Medicine, University of Pittsburgh, PA 15213 USA*

[b]*Laboratory of Computational Biology and Faculty of Engineering and Natural Sciences, Sabanci University, Orhanli, Tuzla 81474, Istanbul, Turkey*

Abstract

A systematic method of representing and analyzing the intramolecular strains in proteins is proposed. For illustrative purposes, the method is applied to the N-terminal fragment of the human T-cell glycoprotein CD4. The method is based on the singular value decomposition (SVD) of molecular dynamics (MD) trajectories. The slowest three modes of motion that carry information along the protein molecule over large length scales are analyzed, so as to characterize the collective motions and the resulting strains along the three principal axes of the protein. Strong cooperative motions of different types, mainly wave-like, wagging, wiggling, breathing, bending and shearing motions, and rigid body rotations are distinguished. The mean-square fluctuations of C^{α}-atoms induced by the three dominant modes are found to exhibit a closer correlation with the experimental temperature factors in the presence of solvent.

Keywords: Biomolecular process; T-cell glycoprotein; HIV

1. Introduction

A systematic analysis of the type and strength of cooperative motions in proteins is of fundamental importance for clarifying the structure–function relationships underlying most biomolecular processes. Recent studies demonstrate that it is possible to gain insights into the mechanism of cooperative motions by decomposing the molecular dynamics (MD) trajectories into collective anharmonic modes of motion and focusing on the effect of a few dominant modes [1–6]. In the present study, a systematic method of representing and analyzing the cooperative dynamics of proteins is developed. The method is based on the singular value decomposition (SVD) of MD trajectories of atomic fluctuations. SVD has in fact proven useful in various disciplines [3,5], and its recent application to protein dynamics [3,4,6] revealed its potential utility for elucidating the collective motions which are not readily accessible by conventional MD simulations.

In search of cooperative dynamics, the slowest modes of motion that carry information along the protein molecule over large length scales are considered. A large length scale is commensurate with the overall dimensions of the protein. A small length scale, on the other hand, is defined at the residue or sub-residue level. Events taking place at the small length scale may be classified as *local* dynamics.

The specific protein investigated in the present study is the N-terminal fragment of the human T-cell glycoprotein CD4. [7,8] Human CD4 is the receptor for HIV, with the first domain carrying the critical sites for the HIV binding [7,8]. The X-ray elucidated structure of the investigated fragment is shown in Fig. 1. This consists of 177 residues arranged in two domains, both of which fold into Greek key β barrels. The tertiary fold of domain I (residues 1–98), which comprises nine antiparallel strands, is similar to the variable domain of immunoglobulin light chains. Domain II (residues 99–176) is composed of seven antiparallel strands, and resembles immunoglobulin constant domains. Overall, the structure is rod-shaped, as seen from Fig. 1. There is a large hydrophobic interface between the two closely associated domains. In fact, MD simulations of the same structure showed the high mobility of the residues in the binding region of the molecule [9]. Our analysis departs from that of Ptaszek et al. [9] in that significantly longer trajectories are generated

☆ This paper was originally submitted to *Computational and Theoretical Polymer Science* and received on 28 December 2000; received in revised form on 8 March 2001; accepted on 14 March 2001. Following the incorporation of *Computational and Theoretical Polymer Science* into *Polymer*, this paper was consequently accepted for publication in *Polymer*.

* Corresponding author. Tel.: +90-216-483-9505; fax: +90-216-483-9550.

E-mail address: erman@sabanciuniv.edu (B. Erman).

0032-3861/02/$ - see front matter
PII: S0032-3861(01)00424-4

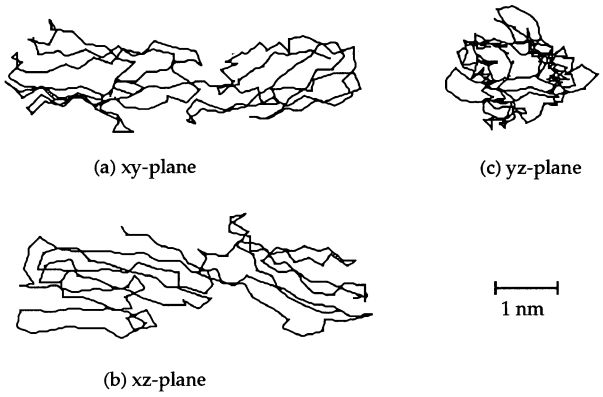

Fig. 1. Projection of X-ray structure (trace of C^{α}-atoms) of CD4 on the three principal planes: (a) front view, (b) top view, and (c) end view. Domain I and II are located on the left and right halves of the front view.

here, and the results are analyzed by SVD technique with the aim of exploring the collective motions of the molecule. Thus, we concentrate here on the contribution of different nonlinear modes to the cooperative motions of the overall molecule.

2. Theory

2.1. The fluctuation trajectory matrix and its modal decomposition

The dynamics of the protein will be expressed in terms of the time-evolved coordinates of backbone carbon atoms C^{α}. In the coarse-grained scale as in the present one, each C^{α}-atom represents a repeat unit (or residue) of the protein. The repeat unit is thus interchangeably used with the corresponding C^{α} atom, which also carries an amino acid side group. The coordinates of the C^{α}-atoms are expressed in the principal coordinate system *xyz*, as outlined in the Appendix A. The position vectors $\mathbf{R}_i$ of C^{α} atoms in the frame *xyz* will be referred to as the *principal coordinates* of the molecule. The x and *z*-axes define the major and minor principal directions, respectively. The *x*, *y* and *z* directions are alternatively named as the longitudinal, lateral, and transverse directions, respectively. The structure of CD4 backbone projected along the three principal planes is presented in Fig. 1. The view in part (a), projected on the *xy*-plane will be referred to as the *front view*. The view projected on the *yz*-plane in part (b) will be referred to as the *top view*, and the view projected on the *yz*-plane in part (c) is the *end view*, or equivalently, the cross-section of the protein.

The thermal fluctuations of the position vectors $\mathbf{R}_i$ with respect to their native states, undergone in MD simulations of equilibrated CD4 structure, will be analyzed. The fluctuations are organized in a matrix $\Delta\mathbf{R}$ called the *fluctuation trajectory matrix*

$$\Delta\mathbf{R} = \begin{bmatrix} \Delta R_1(t_1) & \Delta R_1(t_2) & \cdot & \cdot & \cdot & \Delta R_1(t_n) \\ \Delta R_2(t_1) & \Delta R_2(t_2) & \cdot & \cdot & \cdot & \Delta R_2(t_n) \\ \Delta R_3(t_1) & \Delta R_3(t_2) & \cdot & \cdot & \cdot & \Delta R_3(t_n) \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \Delta R_m(t_1) & \Delta R_m(t_2) & \cdot & \cdot & \cdot & \Delta R_m(t_n) \end{bmatrix} \quad (1)$$

for a protein of *m* residues. Here, $\Delta\mathbf{R}_i(t_j)$ represents, for example, the change in the position vector of the *i*th α carbon in the *j*th time step with respect to its position in the equilibrated molecule. The organization of the time-evolved coordinates in the matrix form of Eq. (1) permits the formal separation of space and time-dependent components as

$$\Delta\mathbf{R} = \begin{bmatrix} & \text{time} \rightarrow \\ \text{space} & \\ \downarrow & \end{bmatrix} \quad (2)$$

The fluctuation trajectory matrix may be readily decomposed into the product of three matrices using the SVD technique [10] as:

$$\Delta\mathbf{R}_{[3m\times n]} = \mathbf{U}_{[3m\times 3m]}\mathbf{\Lambda}_{[3m\times 3m]}\mathbf{V}^{\mathrm{T}}_{[3m\times n]} \quad (3)$$

Here, $\mathbf{U}$ is the matrix of left singular vectors (LSV) of $\Delta\mathbf{R}$, which reflects the time-averaged space-dependent features. $\mathbf{V}$ is the matrix of right singular vectors (RSV). The *i*th row of *V* represents the time evolution of $\Delta\mathbf{R}_i$ in the space spanned by the LSVs, shortly referred to as the singular space. $\mathbf{\Lambda}$ is the diagonal matrix of the singular values λ_i of $\Delta\mathbf{R}$. These are organized in descending order $\lambda_1 \geqq \lambda_2 \geqq \cdots \geqq \lambda_{3m}$.

Let us consider the contribution of the *k*th mode to the observed dynamics. The trajectory corresponding to the *k*th mode may be obtained from $\Delta\mathbf{R}(k) = \mathbf{U}\Lambda'\mathbf{V}^T$, where $\mathbf{\Lambda}'$ contains all zeros except the *k*th singular value λ_k. For a given (fixed) time t_j, the fluctuation of the position vector of the *i*th C^{α} atom may be written as:

$$\Delta\mathbf{R}_i(t_j,k) = Q(t_j,k)\mathbf{u}_i(k) \quad (4)$$

where $Q(t_j,k)$ is a coefficient of proportionality, equating to the *ij*th element of $\lambda_k\mathbf{V}^{\mathrm{T}}$. Thus, $Q(t_j,k)$ modulates the instantaneous fluctuations of all residues induced by the *k*th mode, by a constant factor. The vector $\mathbf{u}_i(k)$, on the other hand, is independent of time, and represents the direction of fluctuations of the *i*th C^{α} atom, specifically, for the *k*th mode. The three components of $\mathbf{u}_i(k)$ are proportional to the principal components of the displacement vector or the displacement field vector [11] for the *i*th α-carbon. In the exploratory calculations below, $Q(t_j,k)$ will be chosen as unity and the three components $u_{ix}(k)$, $u_{iy}(k)$, and $u_{iz}(k)$ of $\mathbf{u}_i(k)$, referred to simply as the components of the displacement field vector, will be investigated.

Several important corollaries follow from Eq. (4), for a pure mode, say the kth: (i) Plotting the components of the displacement field vector as a function of the residue index i gives the fluctuation profile of the protein along the backbone. Alternatively, plotting them as a function of position $\mathbf{R}$, irrespective of the residue index, gives the fluctuation profile in space. For a pure mode, the ratio of the fluctuation amplitudes of any two residues is independent of time. Only the amplitude of fluctuation of each residue changes by a constant factor $Q(t_j,k)$ corresponding to the jth time step. (ii) Zero's of the fluctuation profile correspond to stationary residues (or regions) of the protein. Minima indicate constrained regions. (iii) The time decay of the normalized *auto* and *cross*-correlation functions for the fluctuations of residues are all identical, exhibiting the same correlation time τ_k characteristic of the kth mode. τ_k is therefore a collective property of the overall protein.

2.2. The principal components of the displacement field in a protein

In order to explore the displacements of the protein in the coarse-grained sense, one may define average displacements by performing a summation over residues in a given volume element. For the interest of simplicity, we consider three infinitesimal volume elements, each in the form of a slab, normal to the x, y, and z-axes as shown in Fig. 2. The displacement, $\langle u_i(k,x)\rangle$, induced by the kth mode along the ith principal direction, undergone by a slab which is normal to the x-axis of the principal frame is found from

$$\langle u_i(k,x)\rangle = \frac{\sum_{p=1}^{m} u_{px}(k)\delta(x-x_p)}{\sum_{p=1}^{m}\delta(x-x_p)} \tag{5}$$

where the summations are performed over all m residues, and $\delta(x-x_p)$ is the dirac delta function which selects all residues located within a slab at $x \pm \mathrm{d}x$. Proceeding in this manner, one may obtain nine average displacements which may be arranged in a 3×3 matrix $\mathbf{D}(k)$ of displacements, as:

$$\mathbf{D}(k) = \begin{bmatrix} \langle u_x(k,x)\rangle & \langle u_x(k,y)\rangle & \langle u_x(k,z)\rangle \\ \langle u_y(k,x)\rangle & \langle u_y(k,y)\rangle & \langle u_y(k,z)\rangle \\ \langle u_z(k,x)\rangle & \langle u_z(k,y)\rangle & \langle u_z(k,z)\rangle \end{bmatrix} \tag{6}$$

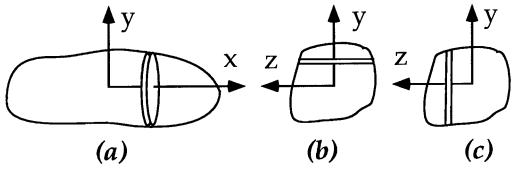

Fig. 2. Schematic representation of three slabs through the molecule, the motions of which will be analyzed: (a) a cross-section normal to the x-axis, the molecule being observed from the front view, (b) a horizontal plane as seen from the end view, (c) a vertical plane as seen from the end view.

The argument k indicates that the kth mode is retained. The elements of the first column of $\mathbf{D}(k)$ denote the average or collective displacements of a cross-section at a position x along the respective principal directions. The elements of the second column represent the displacements of a slab normal to the y-axis through the molecule. The position of the slab is defined relative to the origin by the value of the argument y. Finally, the elements of the third column represent the three components of the displacement vector of a vertical slab whose position relative to the origin is given by the argument z. The three planes whose fluctuations are represented by the three columns of the tensor of Eq. (6) are shown in the respective parts (a)–(c) of Fig. 2.

2.3. Deformation gradients and strains in the protein

The deformation gradient tensor $\mathbf{u}_{i,j}$ for the protein may be obtained from Eq. (6) as:

$$u_{i,j} = \begin{bmatrix} \dfrac{\partial\langle u_x(k,x)\rangle}{\partial x} & \dfrac{\partial\langle u_x(k,y)\rangle}{\partial y} & \dfrac{\partial\langle u_x(k,z)\rangle}{\partial z} \\ \dfrac{\partial\langle u_y(k,x)\rangle}{\partial x} & \dfrac{\partial\langle u_y(k,y)\rangle}{\partial y} & \dfrac{\partial\langle u_y(k,z)\rangle}{\partial z} \\ \dfrac{\partial\langle u_z(k,x)\rangle}{\partial x} & \dfrac{\partial\langle u_z(k,y)\rangle}{\partial y} & \dfrac{\partial\langle u_z(k,z)\rangle}{\partial z} \end{bmatrix} \tag{7}$$

where $\mathbf{u}_{i,j}$ represents differentiation of the ith component of $\mathbf{u}$ with respect to the jth coordinate. The strain tensor $\boldsymbol{\epsilon}$ may be obtained from the deformation gradient tensor according to the relation $\boldsymbol{\epsilon} = (\mathbf{u}_{i,j} + \mathbf{u}_{j,i})/2$.

3. Calculations

Three MD trajectories have been generated. Two of the trajectories are obtained in vacuo by using the parameter set 37D4 of the GROMOS87 force field [12]. The third MD simulation has been carried out in water using the consistent valance force field (CVFF) implemented within the Molecular Simulations Inc. InsightII package [13]. The independent runs were carried out in order to verify the reproducibility of the results, as well as to assess the effect of solvent. The starting configuration of the protein is taken from the X-ray structure entry 2CD4 [7] available in the Brookhaven Protein Databank [14,15].

In all the simulations, the initial structure is relaxed by conjugate gradients energy minimization. Simulations are carried out at 300 K by coupling to an external heat bath [16] with a relaxation time constant of 0.01 ps. Bond lengths are constrained to their equilibrium values by the SHAKE procedure [17]. In the vacuum simulations 800 ps long trajectories, each following a 200 ps equilibration period, are generated with a time steps of 2 fs. Non-bonded interactions are calculated using the twin range method [16] with short- and long-range cutoff radii of 8 and 12 Å, respectively. The neighbor list is updated every 20 fs using the concept of charge groups. With the simulation of 2CD4 in

water, a cage of water molecules of 6 Å thickness is formed around the protein molecule. This procedure leads to the generation of 2001 water molecules. A 100 ps equilibriation period is followed by a 300 ps trajectory, whereby 2.5 fs time steps are used. Group-based cutoffs are used with a 9.5 Å cutoff distance. A switching function is used with spline and buffer widths set to 1.0 and 0.5 Å, respectively. The neighbor list is updated whenever any atom moves more than one-half the buffer width.

Note that, in previous work MD simulations of 135 and 100 ps duration have been carried out in solution and in vacuo, respectively, starting from the same structure [9]. Simulations in vacuo exhibited stronger fluctuations near the native structure (rms deviation of backbone atoms of about 3.0 Å) compared to those in solution (rms deviation of about 1.5 Å). Yet, the data from both of the simulations were reported to agree moderately well with each other and with the crystal structure [9]. Similarly, the average rms fluctuations in C^{α}-positions is 3.02 and 1.44 Å, in the vacuum and explicit solvent simulations, respectively, indicating that substates around the native state, only, are visited throughout the simulations.

The mean position vectors $\bar{R}_i$ of the C^{α} coordinates throughout simulations are evaluated from the average over-all configurations recorded at 0.25 ps intervals. The results are rearranged in the form of the matrix $\bar{R}$ given by Eq. (A2) to obtain the principal axes x, y and z of the protein as defined in the Appendix A. The time-evolved coordinates are transformed into the frame xyz, and the corresponding atomic fluctuations are organized in the trajectory matrix $\Delta\mathbf{R}$ (Eq. (1)). SVD of $\Delta\mathbf{R}$ yields the singular modes and directions. The first three dominant modes will be analyzed inasmuch as the slowest modes operating over large length scales are presently explored. The contribution of the first three modes to the overall motion is found from the relation $\sum_{i=1}^{3} \lambda_i / \sum_{i=1}^{3m} \lambda_i$ to be 70% in vacuum and 40% in explicit solvent simulations.

4. Results and discussion

(i) *Effect of the first mode*. In Fig. 3(a)–(c), the displacements $\langle u_x(1,x)\rangle$, $\langle u_y(1,x)\rangle$ and $\langle u_z(1,x)\rangle$ imposed by the first mode on residues located at the position x of the longitudinal axis are presented as a function of x. The horizontal straight lines in the figures serve as references. The broken, thin lines go through the positions of all the C^{α} atoms along the x-direction. The solid curves, on the other hand, are obtained by fitting the thin lines with polynomials of order seven. One may make the plausible assumption that these curves are representative of the mean displacements defined by Eq. (5). The term $\langle u_x(1,x)\rangle$, for example, describes the motion of the slab illustrated in part (a) of Fig. 2 along the longitudinal axis x; whereas $\langle u_y(1,x)\rangle$ and $\langle u_z(1,x)\rangle$ refer to the motion of the same cross-sectional volume element in the directions of the y- and z-axes, respectively. These motions are extracted from one of the vacuum trajectories generated.

According to the solid curve in Fig. 3(a), the left half of the molecule along the longitudinal axis, or domain I, is almost stationary along the x-direction. The right half

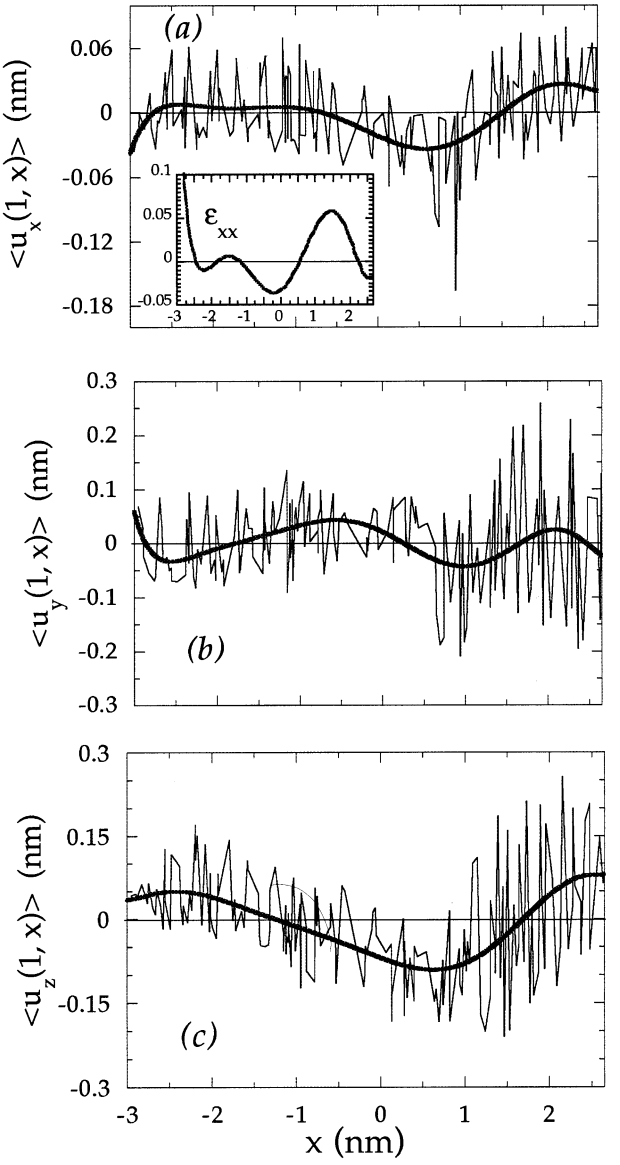

Fig. 3. Collective motions induced by the first ($k = 1$) dominant mode of motion identified by the SVD of a vacuum trajectory of 800 ps. Displacements of the slab shown in part (a) of Fig. 2 are shown as a function of the position of the slab along the x-axis. Parts (a), (b) and (c) display the x-, y- and z-components, respectively, of the mean displacements for the slab. The thin lines connect the results $u_{ix}(k)$, $u_{iy}(k)$, and $u_{iz}(k)$ obtained for individual residues i, as a function of their position along the x-axis. The solid curves represent the best fitting seventh order polynomials. They describe the collective motions along the longitudinal axis of the molecule (see Eq. (5)). The inset in part (a) displays the normal strain ϵ_{xx} along the x-axis. The collective motions observed in parts (a)–(c) may be characterized as earthworming, wave-like and wiggling types of motions, respectively.

(domain II), on the other hand, is rather mobile, part of it exhibiting a contractile motion while the remaining part undergoes an elongational motion along the longitudinal direction. This mode of motion resembles the locomotion of an earthworm, and may be referred to as 'earthworming'. The characteristic time of this mode is obtained from the correlation analysis (see below) of the RSVs as 58.6 ps. The inset displays the corresponding component ϵ_{xx} of the strain tensor ϵ, obtained by differentiating the solid curve representative of $\langle u_x(1,x)\rangle$. Domain I is verified to undergo relatively small normal strains along the longitudinal direction, except for the N-terminus which is highly stretched. In Domain II, strong variations from tension to compression are observed. The average magnitude of the strain is about 5% which is relatively large and may necessitate the consideration of the non-linear components of the displacement gradient in the finite strain tensor.

In Fig. 3, part (b), the shearing motions of the same slab are displayed as a function of x. The solid curve indicates 'wave-like' motions. In part (c), the transverse motions, $\langle u_z(1,x)\rangle$, are presented as a function of x. These motions are observed from the top view of the molecule, which is illustrated in part (b) of Fig. 1. A 'wiggling' motion is indicated by the solid curve.

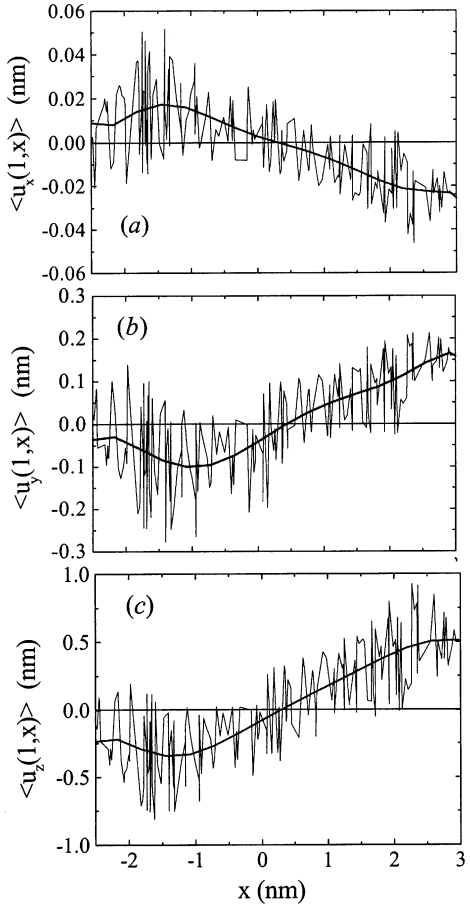

Fig. 4. Same deformations as in Fig. 3, replotted in the presence of solvent.

The effect of solvent on the first mode may be seen by comparing Fig. 3 with Fig. 4. The latter is obtained in the presence of solvent. The contractile and elongational motions along the major axis of the molecule are seen to be distributed over the whole molecule in the presence of solvent, as seen from Fig. 4a. About half of the chain exhibits contractile and the other half exhibits an elongational motion. In Fig. 4b and c, on the other hand, we observe lateral shrinkage accompanying the longitudinal tension in the first domain, while the second domain expands radially, consistent with its longitudinal contraction. It should be stressed that although the results for $\langle u_x(1,x)\rangle$, $\langle u_y(1,x)\rangle$, and $\langle u_z(1,x)\rangle$ differ for the vacuum and solvated molecule, calculations show that the other components of deformation in the first three modes are similar in the absence and presence of solvent. For this reason, only the results in vacuum will be presented in the remaining parts of the manuscript.

We now consider the collective motions as observed from the end-view of the molecule. Results are displayed in Fig. 5(a) and (b). In parallel with Fig. 3, the thin line in each figure reflects the behavior of the individual residues

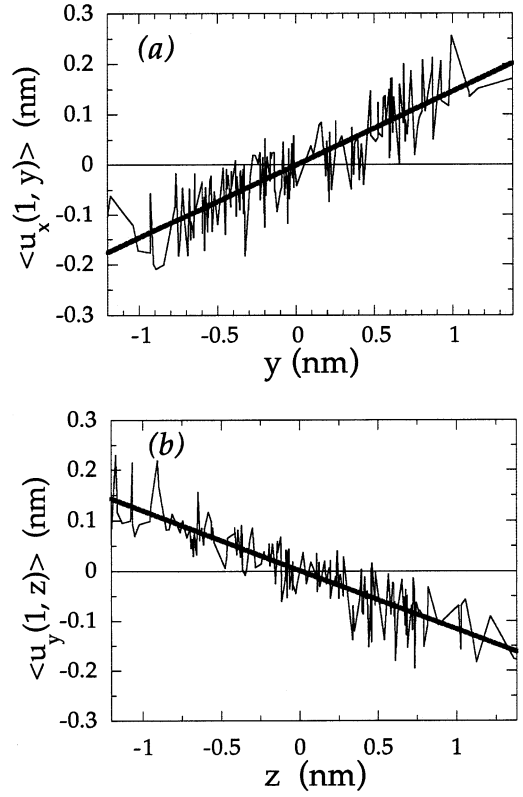

Fig. 5. Shear displacements of slabs observed from the end-view of the molecule from a vacuum trajectory. (a) The transverse component $\langle u_z(1,y)\rangle$ of the displacement tensor $\mathbf{D}(1)$ as a function of the lateral position y of the examined slab. (b) The component $\langle u_y(1,z)\rangle$ as a function of the position along the z-axis. The jagged lines refer to the displacements of individual residues, and the solid curves reflect the collective motions as observed from the end-view, mainly a right-handed rigid body rotation along the longitudinal (x) axis.

as a function of their position along the molecule. The best fitting lines through the data, on the other hand, reflect the shear components $\langle u_z(1,y)\rangle$ and $\langle u_y(1,z)\rangle$ of the collective displacement tensor $\mathbf{D}(1)$. In part (a), the transverse displacements $\langle u_z(1,y)\rangle$ of the slab are shown as a function of the lateral position y. The curve thus reflects the behavior of the successive slabs obtained by sliding along the y-axis. The fact that this is a smoothly increasing curve with increasing y, crossing the value $\langle u_z(1,0)\rangle = 0$, reveals the overall counterclockwise rotation of the molecule as viewed in part (b) of Fig. 2. Thus, the molecule undergoes a right-handed rotation about the longitudinal axis, by the effect of the first mode of motion. The lateral displacements $\langle u_y(1,z)\rangle$, displayed in part (b) of Fig. 5, confirm this rotatory behavior. Here the slabs illustrated in part (c) of Fig. 2 are considered, as a function of their position along the z-axis and their motion in the y-direction are analyzed. The occurrence of a rigid body rotation is also confirmed by substituting the slopes of the straight lines in Fig. 5(a) and (b) into the shear strain $\epsilon_{yz} = (u_{y,z} + u_{z,y})/2$ which equates approximately to zero.

(ii) *Effect of the second mode*. In Fig. 6(a), the longitudinal motions $\langle u_x(2,x)\rangle$ of the molecule along the x-axis induced by the second mode are presented for the vacuum simulation. The points of domain I move along the positive x-direction while those of domain II move in the opposite direction. This reveals a uniform contraction of the molecule along the x-direction. However, since the molecule executes a cyclic motion, it will be expanding at the next instant. This results in a 'breathing' type of motion, the characteristic time of which is found to be 38.0 ps (see below).

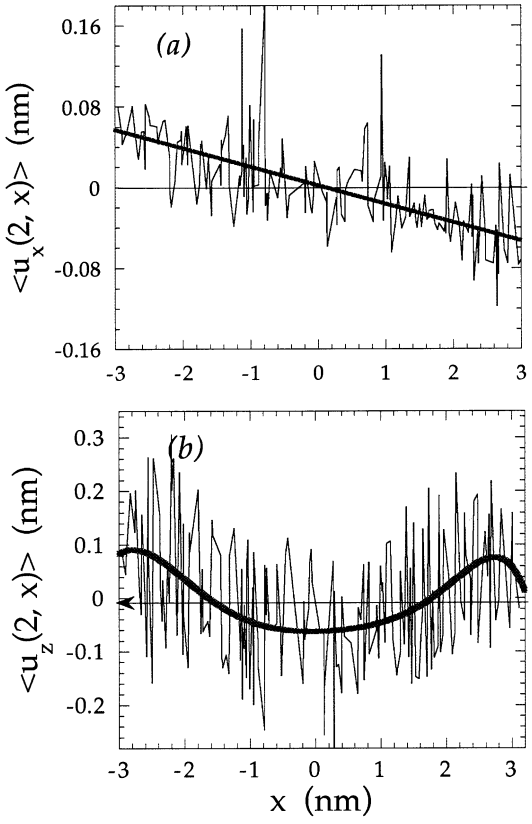

Fig. 6. Motions driven by the second slowest mode of motion from a vacuum trajectory. Part (a) displays the normal displacements $\langle u_x(2,x)\rangle$ of the slabs along the x-axis, in the same format as in Fig. 3. A breathing type of motion is revealed. Part (b) shows the shear displacements $\langle u_z(2,x)\rangle$ of the same cross-sections, i.e. slabs normal to the x-axis, as a function of their position along the x-axis. A sharp bending motion of the molecular principal axis is indicated.

Fig. 6(b) shows the shearing motions of the same cross-sections (shown in part (a) of Fig. 2) as a function of their position along the x-axis. The solid curve for $\langle u_z(2,x)\rangle$ through the data indicates a sharp 'bending' motion of the molecular principal axis when viewed from the front xy-plane. According to this curve, there are two stationary points along the molecule, symmetrically placed at approximately one fourth of the length from each end. The transverse fluctuations, i.e. $\langle u_y(2,x)\rangle$, do not exhibit a pronounced collective pattern, and therefore are not shown here. Results from the second vacuum trajectory as well as that in solution showed, however, pronounced collective features in $\langle u_y(2,x)\rangle$ strikingly similar to those of Fig. 6(b), while no collective pattern was detected in $\langle u_z(2,x)\rangle$. This interchange of the two axes in the two runs, in relation to the effect of the second mode, suggests that the mechanism of motion induced by the second mode is the bending of the longitudinal axis of the molecule without discrimination of the lateral (y) or transverse (z) directions.

The shear components $\langle u_y(2,z)\rangle$ and $\langle u_z(2,y)\rangle$ of $\mathbf{D}(2)$ exhibit a rigid body rotation of the molecule, similar to that of the first mode illustrated in Fig. 5. However, while the rotations in the first mode are counterclockwise when viewed along the negative x-axis, those of the second mode are clockwise. Thus, the two modes lead to rotations of opposite sense. This is expected because overall rotations of the molecule are prevented in the simulations. The shear strain component ϵ_{xy} obtained from $(u_{y,z} + u_{z,y})/2$ is again approximately zero, verifying that a rigid body rotation is driven by this mode.

(iii) *Effect of the third mode*. In the third mode, as well as in the higher modes, the motions of the molecule become less featured. The collective motions are not as clear and well defined as those in the first two modes. The correlation time τ_3 for the third mode is found as 23.0 ps, as will be shown below. Two typical motions worth noting are shown in parts (a) and (b) of Fig. 7. Part (a) displays the lateral fluctuations of C^{α}-atoms along the x-axis. The solid curve through the data shows a pronounced 'wagging' of domain I while domain II exhibits negligible cooperativity in this direction. This behavior extracted from one of the vacuum simulations was observed to be inverted in the other two runs (vacuum and explicit solvent runs): mainly a wagging motion in the y-direction was undergone by the domain II, while domain I exhibited large amplitude non-collective fluctuations. Thus, in the third mode, either the first or the

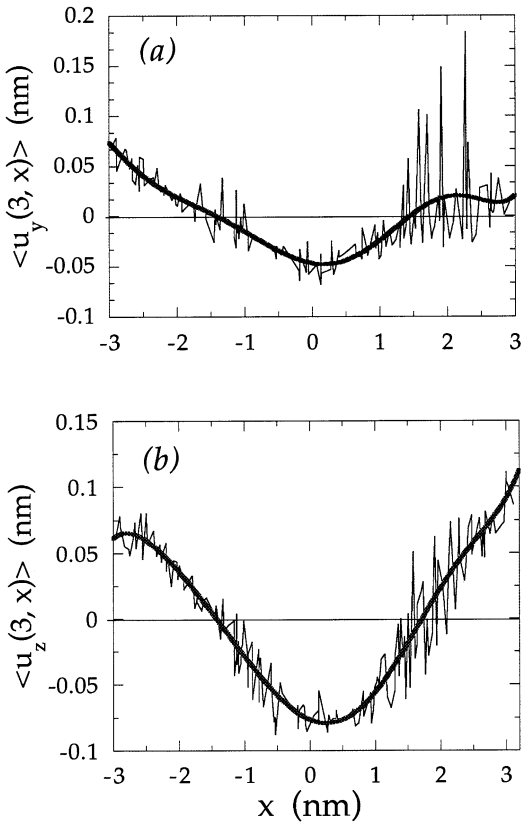

Fig. 7. Shear motions induced by the third slowest mode of a vacuum trajectory on the slabs perpendicular to the x-axis, as a function of the position of the slabs (a) lateral displacements, $\langle u_y(3,x)\rangle$. (b) Transverse displacements $\langle u_z(3,x)\rangle$. Jagged and solid curves refer to the results obtained for individual residues, and the best fitting seventh order polynomials, respectively. Wagging and bending motions are distinguished in the respective parts (a) and (b).

second domain of the molecule undergoes highly directed (minimal noise) wagging motions, while the other domain displays no collective preference. Part (b), on the other hand, reveals a bending motion, comparable to the one driven by the second mode (Fig. 6b). The stationary points in this case are located at about one-third of the overall length of the molecule.

(iv) *Characteristic times of the slowest modes*. Each mode i is characterized by a given correlation time τ_i; and consequently all motions induced by a given mode exhibit the same time dependence, irrespective of their types or amplitudes. Thus, the correlation functions $C_{ij}(k;t)$ associated with the time-delayed fluctuations of atoms i and j, as induced by the kth mode, defined as

$$C_{ij}(t,k) = \langle \Delta\mathbf{R}_i(t_0,k).\Delta\mathbf{R}_j(t_0+t,k)\rangle \tag{8}$$

exhibit the same time decay, $C(t,k)$, provided that they are normalized with respect to their equilibrium ($t = 0$) values. The brackets in Eq. (8) refer to the averaging over various t_0. Fig. 8 shows, for example, the time decay of the normalized

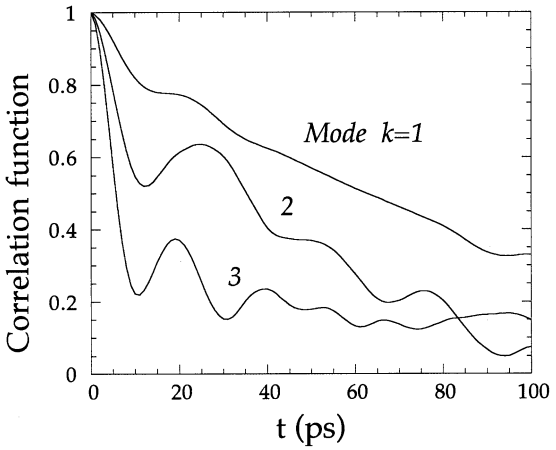

Fig. 8. Time decay of correlation functions associated with the fluctuations of C^{α}-atoms, contributed by the first three dominant modes of motion of a vacuum trajectory. The ordinate values are normalized in the interval [0,1]. They are representative of the time evolution of auto- and cross-correlations for the fluctuations of all pairs of C^{α}-atoms.

cross-correlation functions among any two C^{α} atoms for the first three modes, separately. The correlation times associated with the three dominant modes are estimated from [18]

$$\tau_k = \int_0^{\infty} \frac{C(t,k) - C(\infty,k)}{C(0,k) - C(\infty,k)}\mathrm{d}t \tag{9}$$

to be $\tau_1 = 58.6$ ps, $\tau_2 = 38.0\,ps$ and $\tau_3 = 23.0$ ps, as mentioned above, for the vacuum simulation.

(v) *Local Dynamics*. In the interest of characterizing the local dynamics of the protein molecule, we introduce the *mobility amplitude* of the pth C^{α} atom in the kth mode by the following expression:

$$M_p(k) = u_{px}^2(k) + u_{py}^2(k) + u_{pz}^2(k) \tag{10}$$

Thus, $M_p(k)$ is proportional to the amplitude of the fluctuations undergone by the pth atom under the action of the kth mode. The collective contribution of the first three modes to the fluctuations of the pth C^{α} atom is described by:

$$M_p(1 \le k \le 3) = \sum_{k=1}^{3} M_p(k)\lambda_k^2 \tag{11}$$

where the square singular values λ_k^2 are used for weighting the contributions of the individual modes. Results of explicit solvent simulations based on Eq. (11) are shown by the thick solid curves in Fig. 9 as a function of the residue index along the chain coordinate, starting from the N-terminus. The thin dashed curve represents the average over the two vacuum runs. The temperature factors obtained from crystallographic measurements [7] are displayed by the thick dotted curve. The amplitude fluctuations from simulations are normalized so that the areas enclosed by the curves from simulations are equal to that of the temperature factors.

The distribution curves obtained with the three dominant modes of motion are more structured than the experimental

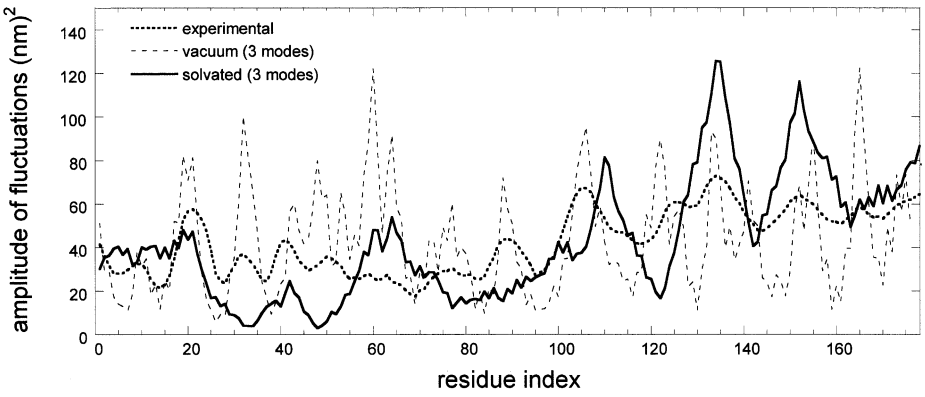

Fig. 9. Distribution of the amplitudes of fluctuations of C^{α}-atoms as a function of residue index along the chain coordinate. The vacuum and explicit solvent simulation results evaluated using Eq. (11) are plotted together with the experimental temperature factors [7]. The simulation results include only the contribution of the three dominant modes of motion identified by SVD.

one, particularly in the case of vacuum simulations. This difference may be attributed to the fact that the theoretical curves are obtained on the basis of the slowest three modes, only, after filtering out all the noise induced by the higher frequency modes, which would otherwise randomize the fluctuation behavior. On the other hand, the elimination of uninteresting modes provides a clear picture of the collective motions, and emphasizes the correlations between segmental motions [19–22].

The fluctuation behavior obtained from the simulation with explicit solvent are found to be in closer agreement with experimental than that obtained from the vacuum simulations. Previous simulations have shown that the presence of solvent smoothes out the collective motions, which is also confirmed by the present results [22].

5. Conclusions

A new method is proposed for extracting information on the global conformational deformations of macromolecules based on a modal decomposition of MD trajectories. Calculations are performed for a protein molecule, CD4, around its native state, and the effect of the dominant non-linear modes on the deformation of the molecule along the molecular principal axes is determined. Several types of normal and shear deformations are identified, which can be described as wave-like, wagging, wiggling, bending, and breathing motions, and rigid body rotations. These motions give rise to a complex distribution of intramolecular strains. Comparison of the results from vacuum simulation with those performed in solution shows that the net effect of solvent is to smoothen the distribution of residue fluctuations improving the overall agreement with the experimentally measured thermal fluctuations.

Acknowledgements

Partial support by Bogazici University Research Funds Projects 99HA501 and 00HA503 are gratefully acknowledged by P.D. and I.B., respectively.

Appendix A. Principal coordinate representation

In this Appendix, a straightforward computational scheme is developed to identify the principal axes of the moment of inertia tensor of the molecule. Let $\mathbf{R}_{0i}$ be the time-averaged position vector of the ith backbone carbon, the α carbon, of the protein in a Cartesian coordinate system, $x_0y_0z_0$, for $1 \le i \le n$. A centroidal coordinate system, $\bar{x}, \bar{y}, \bar{z}$, is defined with the respective axes parallel to $x_0y_0z_0$, in which the mean position vectors $\bar{\boldsymbol{R}}_{\boldsymbol{i}}$ are obtained from the expression:

$$\bar{\boldsymbol{R}}_{\boldsymbol{i}} = \mathbf{R}_{0i} - \frac{\sum_j \mathbf{R}_{0j}}{n} \tag{A1}$$

The centroidal coordinates of the molecule may be organized in the $3 \times n$ configuration matrix

$$\bar{\boldsymbol{R}} = \begin{bmatrix} \bar{x}_1 & \bar{x}_2 & \cdots & \bar{x}_n \\ \bar{y}_1 & \bar{y}_2 & \cdots & \bar{y}_n \\ \bar{z}_1 & \bar{z}_2 & \cdots & \bar{z}_n \end{bmatrix} \tag{A2}$$

The configuration matrix $\bar{\boldsymbol{R}}$ may be decomposed into the product of three matrices, using the SVD technique [10]

$$\bar{\boldsymbol{R}}_{[3\times n]} = \bar{\boldsymbol{U}}_{[3\times 3]} \bar{\boldsymbol{\Lambda}}_{[3\times 3]} \bar{\boldsymbol{V}}^T_{[3\times n]} \tag{A3}$$

where, the subscripts in brackets denote the dimension of each matrix, $\bar{\boldsymbol{U}}$ and $\bar{\boldsymbol{V}}$ are the matrices of the LSVs and RSVs of $\bar{\boldsymbol{R}}$, respectively, and $\bar{\boldsymbol{\Lambda}}$ is the diagonal matrix of the three singular values $\bar{\lambda}_i$ of $\bar{\boldsymbol{R}}$. Multiplying both sides of

Eq. (A3) by its transpose leads to

$$\bar{\boldsymbol{I}} = \bar{\boldsymbol{R}}\bar{\boldsymbol{R}}^{\mathrm{T}} = \begin{bmatrix} \sum_i \bar{x}_i^2 & \sum_i \bar{x}_i\bar{y}_i & \sum_i \bar{x}_i\bar{z}_i \\ \sum_i \bar{x}_i\bar{y}_i & \sum_i \bar{y}_i^2 & \sum_i \bar{y}_i\bar{z}_i \\ \sum_i \bar{x}_i\bar{z}_i & \sum_i \bar{y}_i\bar{z}_i & \sum_i \bar{z}_i^2 \end{bmatrix} = \bar{\boldsymbol{U}}\bar{\boldsymbol{\Lambda}}^2\bar{\boldsymbol{U}}^{\mathrm{T}} \quad \text{(A4)}$$

Defined in this manner, $\bar{\boldsymbol{I}}$ is the moment of inertia tensor per unit mass in the centroidal coordinate frame. The tensor $\bar{\boldsymbol{U}}^{\mathrm{T}}$ transforms from the centroidal coordinate frame $\bar{x}, \bar{y}, \bar{z}$ to the frame *xyz* spanned by the singular directions or LSVs of $\bar{\boldsymbol{R}}$. The coordinate system *xyz* is the *principal frame* associated with the investigated molecule. Premultiplication of $\bar{\boldsymbol{I}}$ by $\bar{\boldsymbol{U}}^{\mathrm{T}}$ and postmultiplication by $\bar{\boldsymbol{U}}$ yields the diagonal tensor $\bar{\boldsymbol{\Lambda}}^2$, the elements of which are the principal components of $\bar{\boldsymbol{I}}$. Thus, the principal axes of $\bar{\boldsymbol{I}}$ coincide with the singular directions of $\bar{\boldsymbol{R}}$.

References

[1] Amadei A, Linssen ABM, Berendsen HJC. Proteins 1993;17:412.
[2] Garcia AE. Phys Rev Lett 1992;68:2696.
[3] Romo TD, Clarage JB, Sorensen DC, Phillips GN. Proteins 1995;22:311.
[4] Garcia AE, Harman JG. Prot Sci 1996;5:62.
[5] Clarage JB, Romo T, Andrews BK, Pettitt BM, Phillips GN. Proc Natl Acad Sci USA 1995;92:3288.
[6] Doruker P, Atilgan AR, Bahar I. Proteins 2000;40:512.
[7] Wang J, Yan Y, Garrett TPJ, Liu J, Rodgers DW, Garlick RL, Tarr GE, Husain Y, Reinherz EL, Harrison SC. Nature 1990;348:411.
[8] Ryu S, Kwong PD, Truneh A, Porter TG, Arthos J, Rosenberg M, Dai X, Xuong N, Axel R, Sweet RW, Hendrickson WA. Nature 1990;348:419.
[9] Ptaszek LM, Vijayakumar S, Ravishanker G, Beveridge DL. Biopolymers 1994;34:1145.
[10] Golub GH, Van Loan CF. Matrix Computations. Baltimore: Johns Hopkins University Press, 1989.
[11] Sokolnikoff IS. Mathematical theory of elasticity. New York: McGraw-Hill, 1956.
[12] van Gunsteren WF, Berendsen HJC. Groningen Molecular Simulation (GROMOS) Library Manual. Groningen: Biomos, 1987.
[13] Dauber-Osguthorpe P, Roberts VA, Osguthorpe DJ, Wolff J, Genest M, Hagler AT. Proteins 1988;4:31.
[14] Bernstein FC, Koetzle TF, Williams GJB, Meyer EF, Brice MD, Rodgers JR, Kennard O, Shimanovichi T, Tasumi M. J Mol Biol 1977;112:535.
[15] Abola EE, Bernstein FC, Bryant SH, Koetzle TF, Weng J. Data Commission of the International Union of Crystallography 1987:107.
[16] Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, Haak JR. Chem Phys 1984;81:3684.
[17] Ryckaert J, Ciccotti G, Berendsen HJC. J Comput Phys 1977;23:327.
[18] Bahar I, Erman B, Monnerie L. Adv Polym Sci 1994;116:145.
[19] Dauber-Osguthorpe P, Osguthorpe DJ. J Comput Chem 1993;14:1259.
[20] Baysal C, Atilgan AR, Erman B, Bahar I. J Chem Soc Faraday Trans 1995;91:2483.
[21] Baysal C, Atilgan AR, Erman B, Bahar I. Macromolecules 1996;29:2510.
[22] Kitao A, Go N. Curr Opin Struct Biol 1999;9:164.