

## Protein Motions at Zero-Total Angular Momentum: The Importance of Long-Range Correlations

Yaoqi Zhou,\* Michael Cook,<sup>†</sup> and Martin Karplus<sup>‡</sup>

\*Department of Chemistry and Chemical Biology, Harvard University, Cambridge, Massachusetts 02138 USA; <sup>†</sup>Neotrope Research, Pelham, Massachusetts 01002 USA; and <sup>‡</sup>Laboratoire de Chimie Biophysique, Institut le Bel, Université Louis Pasteur, 67000 Strasbourg, France

**ABSTRACT** A constant-energy molecular dynamics simulation is used to monitor protein motion at zero-total angular momentum. With a simple protein model, it is shown that overall rotation is possible at zero-total angular momentum as a result of flexibility. Since the rotational motion is negligible on a time scale of 1000 reduced time units, the essentially rotation-free portion of the trajectory provides an unbiased test of the common approximate methods for separating overall rotation from internal motions by optimal superposition. Removing rotation by minimizing the root-mean-square deviation (RMSD) for the entire system is found to be more appropriate than using the RMSD for only the more rigid part of the system. The results verify the existence of positive cross-correlation in the motions of atoms separated by large distances.

The biological functions of proteins require internal motions which can be characterized by the atomic fluctuations and their cross-correlations (Brooks et al., 1988). The determination of the atomic fluctuations from molecular dynamics simulations is complicated by global rotation, which is difficult to remove since the moments of inertia change with time for a non-rigid object like a protein. The most common method for removing rotational motion is to minimize the root-mean-squared deviation (RMSD) through finite rotations of all configurations with respect to the first configuration (Kabsch, 1976; Brooks et al., 1983). This method, although exact for a rigid body, is only approximate for a flexible system. Furthermore, a question has been raised as to whether the RMSD should be minimized for the most rigid part of the system (Hünengeberger et al., 1995), or for the entire system (Ichiye and Karplus, 1991; Karplus and Ichiye, 1996). Different implementations have been found to yield significantly different results for long-range cross-correlations of the internal motions (Hünengeberger et al., 1995; Ichiye and Karplus, 1991; Karplus and Ichiye, 1996). Thus, it is of interest to examine the internal motions of proteins under the constraint of a zero-total angular momentum. One way of achieving this is by doing normal mode dynamics to determine the internal motions (Brooks and Karplus, 1983; Ichiye and Karplus, 1991). Here, we perform molecular dynamics at zero-angular momentum for a protein model and show that the results can be used to obtain a well-defined separation of the internal motions and the overall rotation.

The total angular and linear momentum can be constrained to zero by using a constant-energy molecular dy-

namics simulation in which the total angular and linear momentum are conserved (Allen and Tildesley, 1987). Consequently, if total angular and linear momentum are set to zero at the beginning of a simulation, a trajectory with zero total angular and linear momentum will be obtained if numerical errors do not accumulate. In this paper, we obtain such a trajectory for a simple model three-helix bundle protein (Zhou and Karplus, 1997; Zhou and Karplus, 1999a,b).

The model protein (Fig. 1) consists of 46 freely jointed beads each of which represents an amino acid residue that can interact with other residues via a square-well potential. The square-well depth is  $-\epsilon$  ( $\epsilon > 0$ ) if the interaction pair are in contact in the global-minimum structure and is  $-0.7\epsilon$ , otherwise. Details of the model and the discrete molecular dynamics method used for the simulations can be found in Zhou and Karplus (1999b). Despite the simplicity of the model, it possesses many essential features of proteins, such as the existence of both disordered and ordered globule states, a two-state folding transition and a low-temperature frozen state similar in structure to the native state (Zhou and Karplus, 1997). The folding behavior of the model has been used to study general aspects of the folding of  $\alpha$ -helical proteins (Zhou and Karplus, 1999a,b).

The method for obtaining a constant-energy trajectory is as follows. First, the value for the total energy of the system is calculated from constant-temperature simulations. Based on results from the constant-temperature simulations (Zhou and Karplus, 1997), the model is known to be in its native state at a reduced temperature of  $T^* = 0.25$  ( $T^* = k_B T / \epsilon$ ,  $k_B$ , the Boltzmann constant and  $T$ , the temperature). The corresponding average total energy  $\langle TE \rangle$  from a simulation at that temperature is found to be  $-212.37\epsilon$  from the equation

$$\langle TE \rangle = \langle KE \rangle + \langle PE \rangle = \frac{(3N - 6)}{2} k_B T + \langle PE \rangle \quad (1)$$

where  $\langle KE \rangle$  and  $\langle PE \rangle$  are the average kinetic and potential energies, respectively;  $N = 46$  is the total number of residues. Note that six global rotational and translational degrees of freedom are explicitly removed from the kinetic

Received for publication 30 May 2000 and in final form 15 August 2000.

Y. Zhou's current address: Dept. of Physiology and Biophysics, State University of New York at Buffalo, 124 Sherman Hall, Buffalo, NY 14214.

Address reprint requests to Martin Karplus, Dept. of Chemistry and Chemical Biology, Harvard University, 12 Oxford St., Cambridge, MA 02138. Tel.: 617-495-4018; Fax: 617-496-3204; E-mail: marci@tammy.harvard.edu.

© 2000 by the Biophysical Society

0006-3495/00/12/2902/07 \$2.00

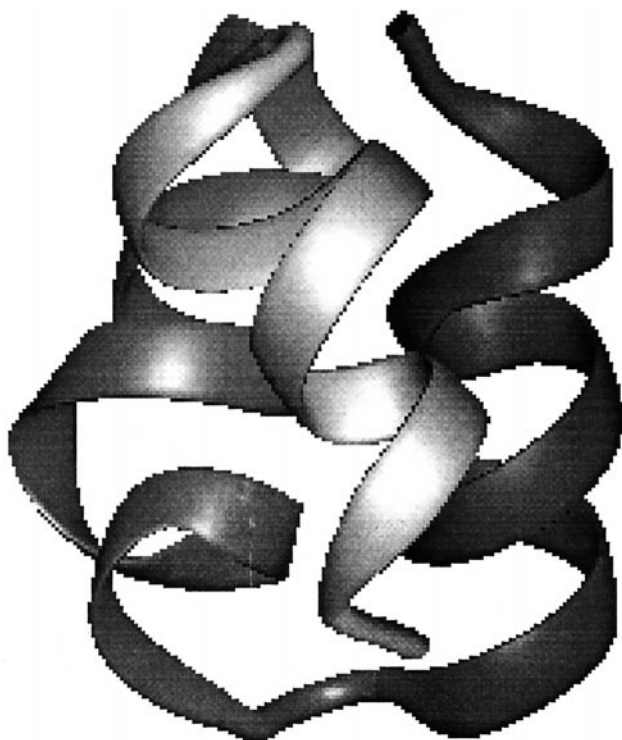


FIGURE 1 The global minimum structure of the model with helix I to III colored light to dark.

energy in Eq. 1 and they do not contribute to the potential energy in this isolated system. The instantaneous total angular momentum  $\mathbf{L}$  of the system about the center of mass is (Goldstein, 1980)

$$\mathbf{L} = M \sum_i \mathbf{r}_i \times \mathbf{v}_i \quad (2)$$

where  $\mathbf{r}_i$  and  $\mathbf{v}_i$  are the position and velocity of residue  $i$ , respectively and  $M$  is the mass of a residue; the same mass is assigned to all residues. The instantaneous angular velocity  $\boldsymbol{\omega}$  of the system satisfies the equation

$$\boldsymbol{\omega} = \mathbf{I}^{-1} \mathbf{L} \quad (3)$$

where  $\mathbf{I}$  is the moment of inertia tensor;  $I_{xx} = M \sum_i (r_i^2 - x_i^2)$  and  $I_{xy} = -M \sum_i x_i y_i$ ; etc.; principal moments of inertia correspond to the eigenvalues of the inertia tensor  $\mathbf{I}$  (Goldstein, 1980). In this paper, all quantities are reported in reduced units; for the total angular momentum and the inertia tensor, they are  $\mathbf{L}^* = \mathbf{L} / \sqrt{\epsilon M \sigma^2}$  and  $\mathbf{I}^* = \mathbf{I} / M \sigma^2$ , respectively. For time,  $t^* = t \sqrt{\epsilon / M \sigma^2}$ . Here,  $\sigma$  is the hard core diameter which is equal to 4.27 Å (Zhou and Karplus, 1997).

The initial configuration and velocities are obtained from a phase point on the equilibrium constant-temperature simulation at  $T^* = 0.25$ . The total linear and angular momenta are removed by removing the linear ( $\sum \mathbf{v}_i / N$ ) and angular ( $\boldsymbol{\omega} \times \mathbf{r}_i$ ) components of the velocities ( $\mathbf{v}_i^{\text{new}} = \mathbf{v}_i^{\text{old}} -$

$\sum_j \mathbf{v}_j^{\text{old}} / N - \boldsymbol{\omega} \times \mathbf{r}_i$ ) (Brooks et al., 1983). The resulting velocities are then scaled by a constant multiplication factor so that the total energy is equal to  $-212.37\epsilon$ , the average total energy for the system (see above). The first 100 million steps (equilibration collisions) are discarded. The equilibrium trajectory is sampled every 10 reduced time units ( $\sim 10,000$  collisions). Each reduced time unit corresponds approximately to 1 ns (Zhou and Karplus, 1999a) based on the experimental time of collapse in protein folding ( $\sim 1 \mu\text{s}$ ) (Ballew et al., 1996; Muñoz et al., 1997).

The equilibrium simulation of the model protein lasts for  $10^5$  reduced time units, which corresponds to roughly 102 million collisions. The average reduced temperature is found to be 0.255 and the average potential energy,  $\langle PE \rangle$ , is  $-229.2\epsilon$ . This is in good agreement with  $\langle PE \rangle = -227.5\epsilon$  at the same average temperature obtained from a constant-temperature simulation. Independent constant-energy simulations (with different initial configurations and velocities but with the same total energy) yield essentially the same equilibrium results.

The total angular momentum is found to increase systematically from  $10^{-16}$  reduced units (the limit of numerical precisions) to  $10^{-12}$  reduced units. To avoid possible error accumulations, the angular and linear momentum are reset to zero after every collision in the production simulation. Since the resetting changes only global rigid-body motions, the equilibrium averages of the thermodynamic properties (i.e., energy, heat capacity) are essentially unchanged, as expected. Due to numerical precision, the instantaneous total angular momentum is not exactly zero even though it

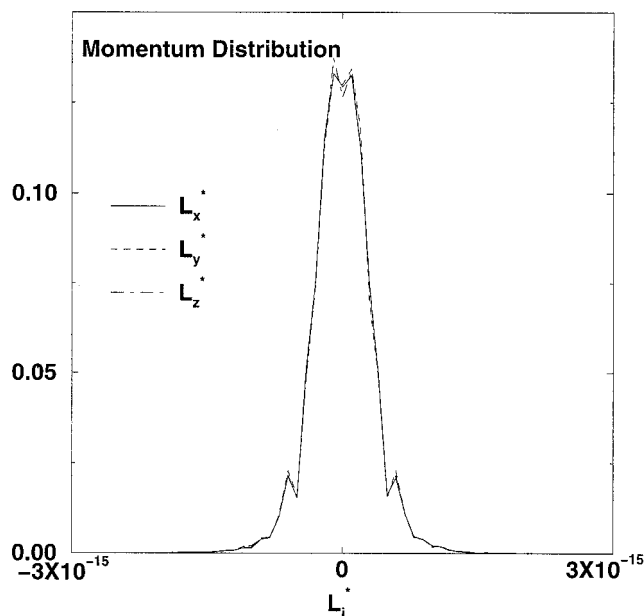


FIGURE 2 An example of the distribution of the reduced total angular momentum  $L_x^*$ ,  $L_y^*$ , and  $L_z^*$  found in a constant-energy simulation during which the linear and angular momentum are set to zero after every collision.

is reset to zero at every collision. Fig. 2 shows a nearly symmetric essentially Gaussian distribution around zero with a width at half-height equal to  $3.5 \times 10^{-16}$  reduced units for the instantaneous total angular momentum. This indicates that the numerical errors in the value of the angular momentum are random and of the order expected from the numerical precision. Thus, there should be no cumulative error that could lead to global rotations.

Fig. 3 shows six instantaneous structures obtained from the constant-energy simulation under the constraint of zero-total angular and linear momentum. As time increases, the structure of the model protein appears to gradually “rotate” in a counter-clockwise direction away from the starting structure; no translational motion is observed. The various structures differ from the  $t^* = 0$  structure by only a very small ( $0.15\sigma - 0.21\sigma$ ) root-mean-square deviation after

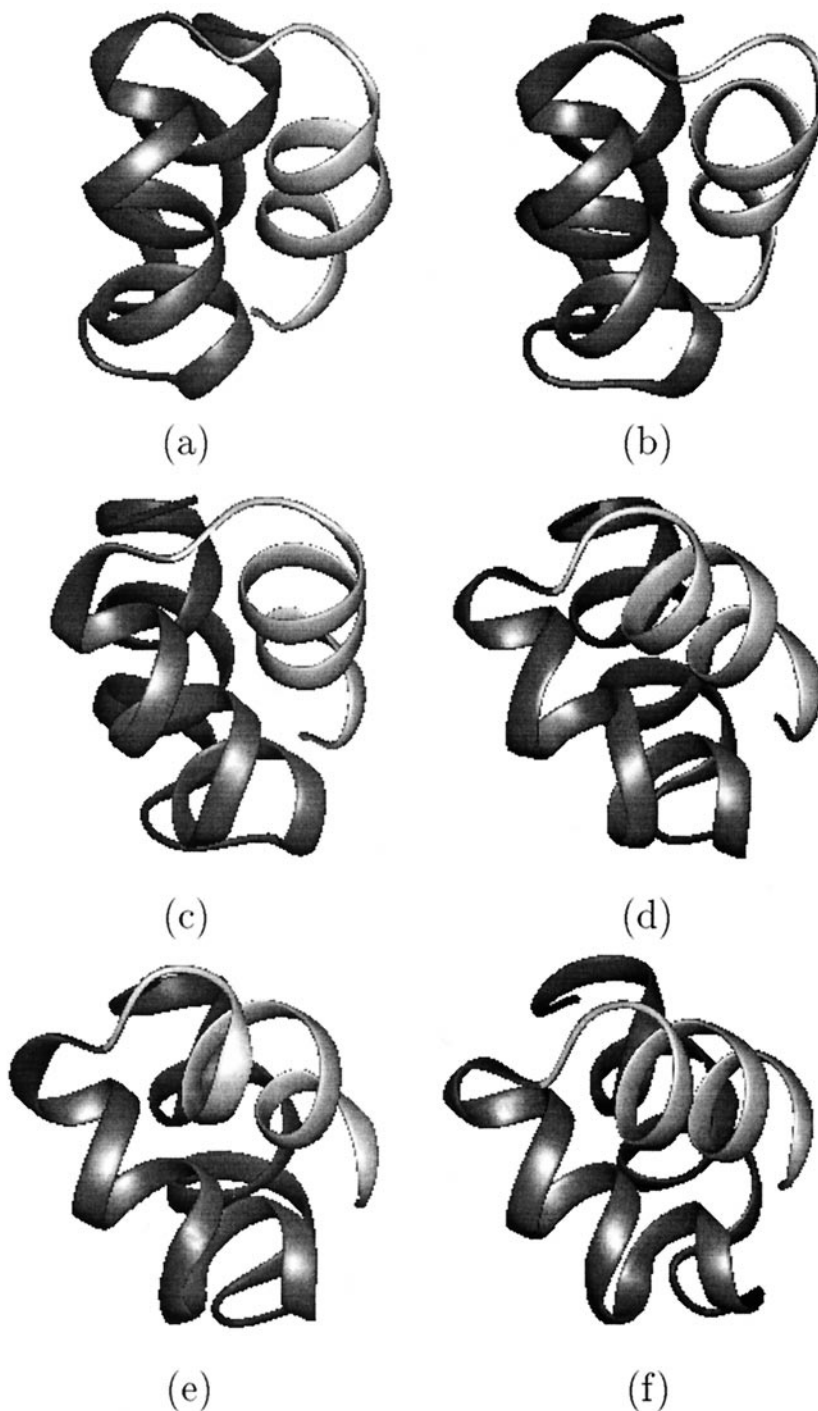


FIGURE 3 The instantaneous structures of the constant-energy simulation at (a)  $t^* = 0$ , (b)  $2 \times 10^4$ , (c)  $4 \times 10^4$ , (d)  $6 \times 10^4$ , (e)  $8 \times 10^4$ , and (f)  $10^5$ , respectively. The three helices (I–III) are colored light to dark.

optimal superposition. Thus, the observed rotation is approximately “rigid-body” rotation, even though the system has essentially zero total angular momentum (see below). The rotational matrix elements used to minimize the RMSD with respect to the starting structure are shown as a function of time in Fig. 4. The overall rotational matrix elements are small for the first 1000 reduced time units and increase in a nonuniform manner.

Where does the long-time-scale rotation come from? The principal moments of inertia ( $I_{xx}^*$ ,  $I_{yy}^*$ , and  $I_{zz}^*$ ) for the starting configuration are 91.77, 96.12, and 60.38, respectively. To estimate the maximum amount of rotation caused by error accumulation, we consider a rigid system that rotates with the reduced angular momentum  $L_x^* = 0$ ,  $L_y^* = 0$ , and  $L_z^* = 10^{-13}$  for  $t^* = 10^5$ . The value  $10^{-13}$ , which is two orders of magnitude larger than the maximum value of angular momentum shown in Fig. 2, is used to estimate an upper bound for the accumulated error. Such a rigid system would rotate over  $L_z^*/I_{zz}^* \sim 10^{-8}$  degrees during the entire simulation. This is to be compared with the actual rotation of  $42^\circ$  (the Euler angle  $\theta$ ) at  $t^* = 10^5$ . Thus, small numerical errors in angular momentum cannot explain the large rotation found in the simulation, even if they were to accumulate.

Another possibility is that the rotation arises from the presence of systematic errors in the program. To ensure that

this is not the case, we simulated a rigid three-helix model using the same program. The rigid three-helix model corresponds to  $t^* = 0$  structure used in the equilibrium constant-energy simulation. The interaction between the pair of residues  $i$  and  $j$  is given by an infinitely deep square-well potential,

$$u_{i,j}(r) = \begin{cases} \infty, & r < (1 - \delta)R_{ij}, \\ 0, & (1 - \delta)R_{ij} < r < (1 + \delta)R_{ij}, \\ \infty, & r > (1 + \delta)R_{ij}. \end{cases} \quad (4)$$

where  $R_{ij}$  is the distance between residues  $i$  and  $j$  in the  $t^* = 0$  structure and  $\delta$  is the bond flexibility parameter. All residues are connected via a nearly rigid bond. For  $\delta = 0$ , we have a rigid three-helix bundle with fixed distances between all pairs of residues. Over the time period used for the flexible protein model ( $10^5$  reduced time units), the rigid model with  $\delta = 0.01$  rotates very little (Fig. 5 *a*); the rotation is only  $2^\circ$  in terms of the Euler angle  $\theta$ . This is reflected by the fact that the elements of the rotational matrix for optimal superposition is either very close to 1 ( $u_{ii}$ ) or 0 ( $u_{ij}$ ) during the entire  $10^5$  reduced time unit period. As the flexibility of the model increases ( $\delta$  increases from 0.01 to 0.1), relative rotation between different instantaneous structures starts to appear (Fig. 5 *b*). Since  $\delta = 0.1$  is used for bond constraint in the actual model for the protein,

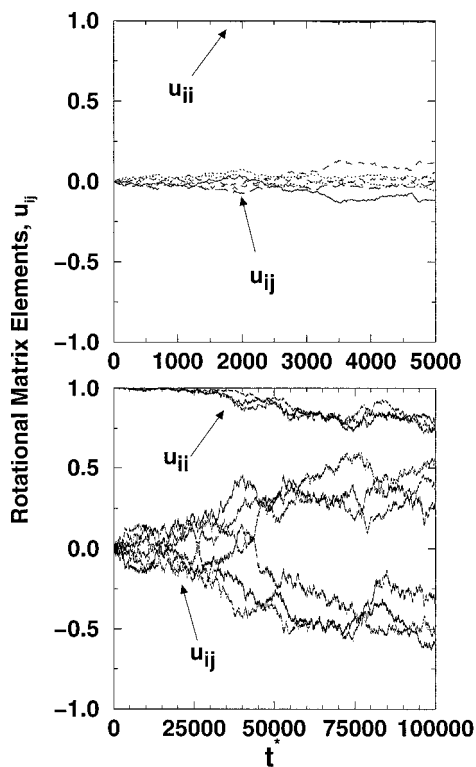


FIGURE 4 The nine rotational matrix elements,  $u_{ij}$ , that are obtained when the structure is rotated to minimize the RMSD with respect to the starting structure. The results of  $t^*$  from 0 to 5000 reduced units are shown at the top. The results for the entire simulation are shown on the bottom.

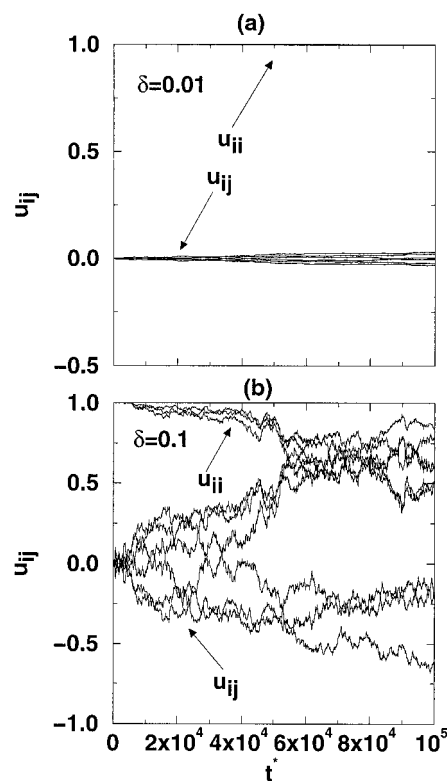


FIGURE 5 As in Fig. 4 but for the rigid model with the flexibility parameter (a)  $\delta = 0.01$  and (b)  $\delta = 0.1$ , respectively.

this points to protein flexibility as the source of the time-dependent rotation shown in Figs. 3 and 4.

That a flexible system can rotate with zero total angular momentum is well established and has been used to explain such everyday observations as that divers, gymnasts, and cats can rotate under overall zero-angular momentum conditions (Frohlich, 1980). For proteins, the movement of this flexible system at each time step can be decomposed conceptually into a rigid rotational motion in which the distance between all pairs of residues are unchanged and the individual translational motion of each residue that moves the residues from the rigidly rotated positions to their final positions. A rigid rotation under zero-total angular momentum is possible when the angular momentum associated with the rigid rotation is cancelled by the angular momentum associated with the individual atomic motions.

The zero-angular momentum trajectory from the protein model permits us to analyze how overall (rigid) rotations affect the cross-correlation,  $c_{ij}$ , of the internal motions. The cross-correlation of the internal motions for residues (atoms)  $i$  and  $j$  (Ichiye and Karplus, 1991) is defined by the equation

$$c_{ij} = \frac{\langle (\mathbf{r}_i - \langle \mathbf{r}_i \rangle) \cdot (\mathbf{r}_j - \langle \mathbf{r}_j \rangle) \rangle}{\sqrt{\langle (\mathbf{r}_i - \langle \mathbf{r}_i \rangle)^2 \rangle} \sqrt{\langle (\mathbf{r}_j - \langle \mathbf{r}_j \rangle)^2 \rangle}} \quad (5)$$

where  $\langle \rangle$  denotes configurational averages. Fig. 6 shows  $c_{ij}$  as a function of the average distance between residues  $i, j$ ,  $\langle r_{ij} \rangle$ , obtained via two different methods for removing rotational motions from the trajectory. When the rotational motion is removed by rotating each coordinate to minimize the RMSD for the entire molecule from the initial coordinate set (Fig. 6 *a*), the  $c_{ij}$  values show positive correlations at small and large separations and negative correlations at intermediate separations. This is in agreement with all-atom molecular dynamics simulations of bovine pancreatic trypsin inhibitor (BPTI) as well as normal mode results (Ichiye and Karplus, 1991). The use of rigidly packed residues (core residues here) as the reference frame for minimizing the RMSD changes the long range correlation to nearly zero (Fig. 6 *b*), a result similar to that obtained in all-atom simulations of BPTI and lysozyme by Hünengeberger et al. (1995).

In Fig. 7, the cross-correlation matrix  $c_{ij}$  is calculated directly from the zero-total angular momentum trajectory. The trajectory is analyzed for eight different time intervals: 0–1000, 0–2000, 0–3000, 0–4000, 0–5000, 0–10000, 0–50000 and 0–100000. The matrix elements for each time interval are calculated from the configurations obtained in that length of time. On the short time scale (up to 4000 reduced time units,  $\sim 4,000,000$  collisions), the overall shape of the  $c_{ij}$  versus distance curve is essentially the same as that obtained by minimizing the RMSD for all 46 residues. As time increases, the correlation of  $c_{ij}$  at long distance changes from positive to negative and the distribution

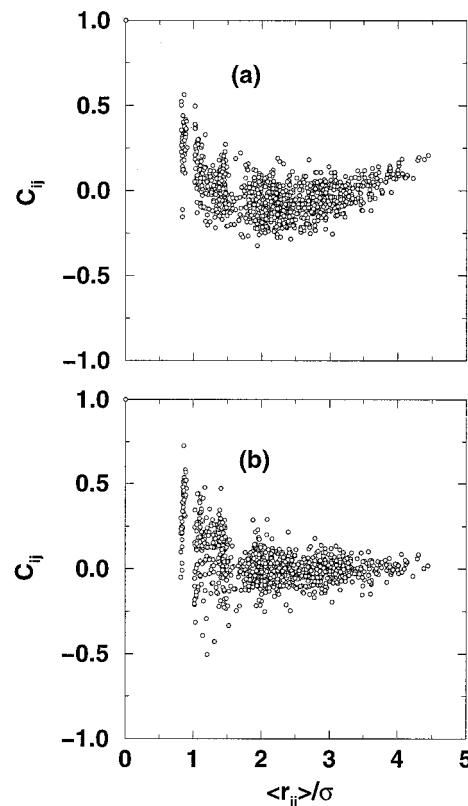


FIGURE 6 The cross-correlation of residue motions as a function of the average distance between a pair of residues average over entire equilibrium simulation ( $t^* = 0$  to  $10^5$ ). Overall rotations are removed (*a*) via minimizing RMSD for all 46 residues and (*b*) via removing RMSD for the 9 residues that are closest to the center of mass on average.

of  $c_{ij}$  at a specific distance becomes broader. A “pure” rigid rotational motion has large negative correlation between the parts that move in opposite directions. This leads to significant increases of negative correlations as the protein rotates away from its original position (Fig. 3). This overall rotation is “random” and, thus, it is expected to average to zero, if the system is simulated long enough to explore fully its rotational degrees of freedom. The similarity between  $c_{ij}$  in Fig. 6 *a* and the “short” time behavior of  $c_{ij}$  in Fig. 7 suggests that removing rotational motions by minimizing the RMSD for the entire system is more appropriate than doing it by minimizing the RMSD only for the more rigid part of the system, in accord with the conclusions of Karplus and Ichiye (1996) and Abseher and Nilges (1998).

We have shown that it is impossible to eliminate overall global rotation exactly in molecular dynamics simulations of a flexible protein-like model and by inference for a protein. Such behavior has been observed in a zero angular momentum simulation of BPTI (B. Brooks, unpublished). This is a consequence of the fact that even under the constraint of zero total angular momentum, global rotation

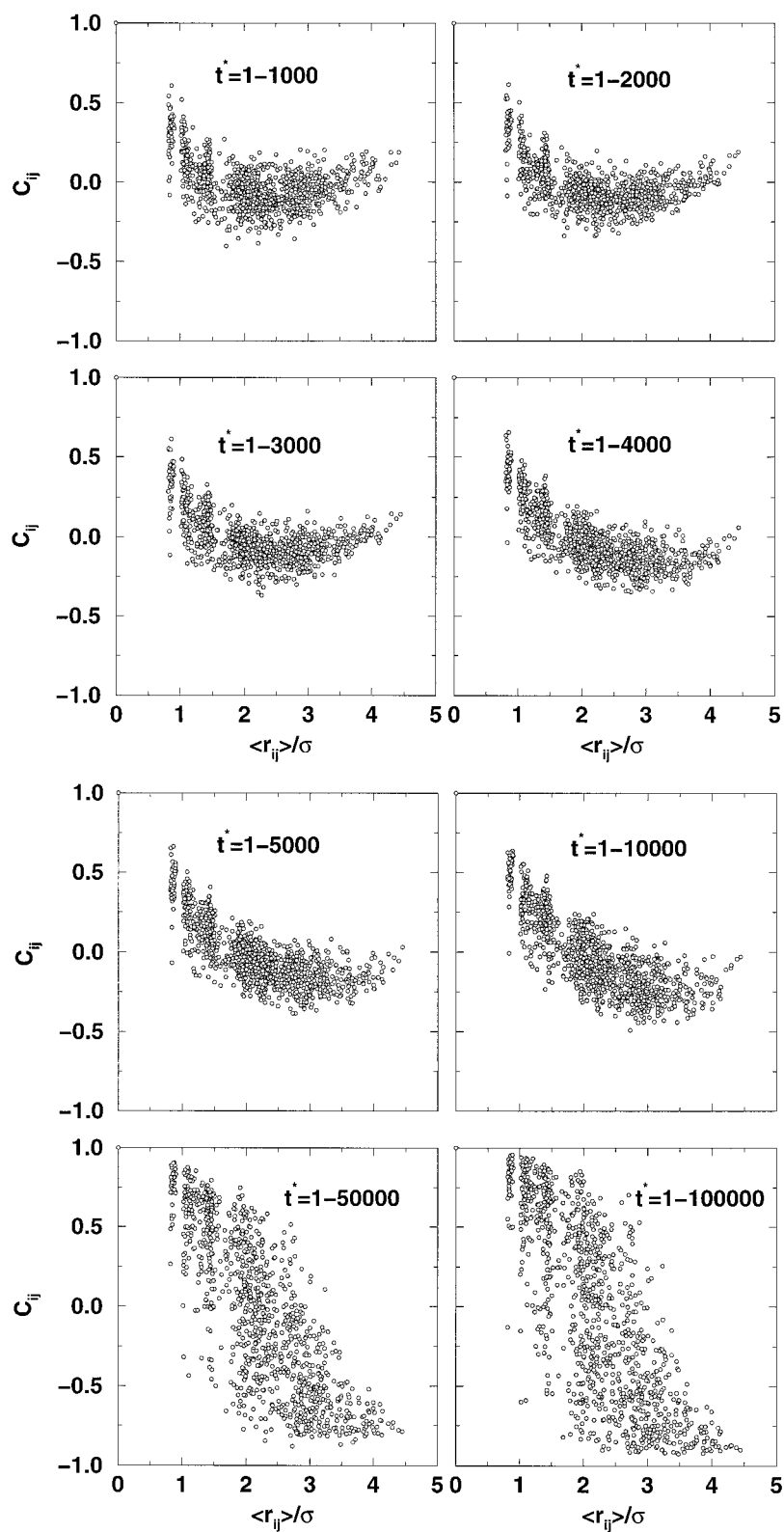


FIGURE 7 The cross-correlation of residue motions as a function of the average distance between a pair of residues average over entire equilibrium simulation. The zero-angular momentum trajectory was used directly. The trajectory was analyzed for eight different lengths of time intervals: 0–1000, 0–2000, 0–3000, 0–4000, 0–5000, 0–10000, 0–50000 and 0–100000, as labeled.

still occurs over a long time period due to the flexibility of the protein. The short-time behavior of the zero total momentum trajectory, which is essentially rotation-free, sug-

gests that removing rotations by minimizing the RMSD for the entire system is more appropriate than minimizing the RMSD for the more rigid part of the system.

This work was supported in part by a grant from the National Science Foundation and a grant from Pittsburgh Supercomputing Center. Y.Z. was a National Institutes of Health Postdoctoral Fellow (1996–1998). Figs. 1 and 3 are made with QUANTA (Molecular Simulations Inc.).

## REFERENCES

- Abseher, R., and M. Nilges. 1998. Are there non-trivial dynamic cross-correlations in proteins? *J. Mol. Biol.* 279:911–920.
- Allen, M. P., and D. J. Tildesley. 1987. *Computer Simulation of Liquids*. Oxford, Oxford University Press.
- Ballew, R. M., J. Sabelko, and M. Gruebele. 1996. Direct observation of fast protein folding: the initial collapse of apomyoglobin. *Proc. Natl. Acad. Sci. U.S.A.* 93:5759–5764.
- Brooks, B. R., R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus. 1983. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* 4:187–217.
- Brooks, B. R., and M. Karplus. 1983. Harmonic dynamics of proteins: normal modes and fluctuations in bovine pancreatic trypsin inhibitor. *Proc. Natl. Acad. Sci. U.S.A.* 80:6571–6575.
- Brooks, C. L., M. Karplus, and B. M. Pettitt. 1988. *Proteins: A Theoretical Perspective of Dynamics, Structure, and Thermodynamics*. New York, John Wiley & Sons.
- Frohlich, C. 1980. The physics of somersaulting and twisting. *Sci. Am.* 242:155.
- Goldstein, H. 1980. *Classical Mechanics*. Reading, MA, Addison-Wesley.
- Hünengeberger, P. H., A. E. Mark, and W. F. van Gunsteren. 1995. Fluctuation and cross-correlation analysis of protein motions observed in nanosecond molecular dynamics simulations. *J. Mol. Biol.* 252:492–503.
- Ichiye, T., and M. Karplus. 1991. Collective motions in proteins: a covariance analysis of atomic fluctuations in molecular dynamics and normal mode simulations. *Proteins*. 11:205–217.
- Kabsch, W. 1976. A solution for the best rotation to relate two sets of vectors. *Acta. Cryst. A.* 32:922–923.
- Karplus, M., and T. Ichiye. 1996. Comment on a fluctuation and cross correlation analysis of protein motions observed in nanosecond molecular dynamics simulations. *J. Mol. Biol.* 263:120–122.
- Muñoz, V., P. A. Thompson, J. Hofrichter, and W. A. Eaton. 1997. Folding dynamics and mechanism of beta-hairpin formation. *Nature*. 390:196–199.
- Zhou, Y., and M. Karplus. 1997. Folding thermodynamics of a model three-helix bundle protein. *Proc. Natl. Acad. Sci. U.S.A.* 94:14429–14432.
- Zhou, Y., and M. Karplus. 1999a. Interpreting the folding kinetics of helical proteins. *Nature*. 401:400–403.
- Zhou, Y., and M. Karplus. 1999b. Folding of a model three-helix bundle protein: a thermodynamic and kinetic analysis. *J. Mol. Biol.* 293:917–951.