

Principal Component Analysis and Long Time Protein Dynamics

M. A. Balsera, W. Wriggers, Y. Oono, and K. Schulten*

Department of Physics and Beckman Institute, University of Illinois at Urbana–Champaign,
405 North Mathews Avenue, Urbana, Illinois 61801

Received: December 12, 1995; In Final Form: December 21, 1995[⊗]

It has been suggested that principal component analysis can identify slow modes in proteins and, thereby, facilitate the study of long time dynamics. However, sampling errors due to finite simulation times preclude the identification of slow modes that can be used for this purpose. This is demonstrated numerically with the aid of simulations of the protein G-actin and analytically with the aid of a model which is exactly recoverable by principal component analysis. Although principal component analysis usually demonstrates that a set of a small number of modes captures the majority of the fluctuations, the set depends on the particular sampling time window and its width.

I. Introduction

Molecular dynamics has become an important tool to study proteins.^{1,2} One of the major obstacles limiting its usefulness is the shortness of achievable simulation times. These times, on the order of a few nanoseconds, are much shorter than the time scales of many important protein processes, e.g., folding.² Many attempts have been made (e.g., refs 3, 4) to extend the time scale of molecular dynamics simulations, but as yet, no satisfactory solution has been found.

One possible approach to reduce the number of degrees of freedom is principal component analysis (PCA), which is also known as the Karhunen–Loeve expansion⁵ in time series analysis. A detailed implementation of the method in the current context can be found in ref 6, and its use has been recently reviewed in ref 7. This method was introduced to the protein research community under the name “quasi-harmonic analysis” by Karplus and co-workers^{8–10} to facilitate the computation of configurational entropies. The method has also been employed with the hope of describing molecular dynamics trajectories in terms of a small number of variables, sometimes called essential degrees of freedom.^{11–15}

It appears natural to use the PCA method to reduce the phase space of proteins for long time molecular dynamics. For this purpose one may determine a small number of important modes by PCA and project the equations of motion on the resulting low-dimensional vector space. The idea has been successfully applied to reduce the dynamics of hydrodynamic systems.¹⁶ In this case a partial differential equation is reduced to a set of a small number of ordinary differential equations. This success strengthens one's expectation that PCA could serve as a useful method for protein dynamics reduction.

The purpose of this paper is to demonstrate that PCA is not suitable for long time protein dynamics reduction with the computational power available at present or in the foreseeable future. The PCA method works superbly in hydrodynamics, because in this case the computationally obtained correlation function is very accurate due to a sufficiently long sampling time. In contrast, PCA for proteins is problematic, since the longest relaxation time τ for many proteins is on the order of microseconds or longer.¹⁷ The error in the correlation matrix for a mode of time scale τ is on the order $\sqrt{\tau/t_{\text{tot}}}$, where t_{tot} is the sampling time. Due to these errors in slow modes, the fast

modes, for which $\sqrt{\tau/t_{\text{tot}}}$ is very small, are contaminated and also become unreliable. The difficulties due to the shortness of simulation have been clearly pointed out by Hodel et al.¹⁸ in the context of free energy calculations and by Clarage et al.¹⁹ and Faure et al.²⁰ in the context of incoherent light scattering experiments. In a comparison of light-scattering patterns derived from PCA and normal mode analysis, the latter method gives better agreement with experiment.²⁰ Our aim here is to shed light on the difficulties of PCA to extract essential modes from a MD simulation, which can be employed for long time descriptions of proteins.

How can we reconcile the difficulties of PCA observed in this paper and apparent successes in analyzing protein dynamics reported by other authors?^{11–15} If a particular large-scale motion arises in a short simulation, PCA surely captures the feature. However, such features may be detectable just as well without PCA, e.g., through atomic root mean square (rms) fluctuations. A nontrivial use of PCA has been successful in describing the fast motions relevant to NMR spectroscopy.²¹ As will be discussed later, however, this success does not justify the application of PCA to describe long time dynamics of proteins.

In section II we will outline PCA and its difficulties in conjunction with long time dynamics. In section III, the application of the PCA method is illustrated through simulations of G-actin. In section IV we explain analytically why PCA does not yield reliable slow modes. Section V provides concluding remarks.

II. Principal Component Analysis and Its Difficulty

Let $\vec{q}(t) = (q_1(t), q_2(t), \dots, q_{3N}(t))'$ be the coordinates of N atoms in a protein at time t , where $'$ denotes transposition. The basic idea of PCA is to seek the least-squares approximant for the position vector in the form

$$\vec{q}(t) \approx \sum_{\alpha=1}^M a_{\alpha}(t) \vec{\eta}_{\alpha} \quad (2.1)$$

where M is the number of relevant degrees of freedom and the vectors $\vec{\eta}_{\alpha}$ are orthonormal basis vectors, which are determined by the eigenvalue problem

$$\lambda_{\alpha} \vec{\eta}_{\alpha} = \mathbf{K}_T \vec{\eta}_{\alpha} \quad (2.2)$$

Here, \mathbf{K}_T is the correlation matrix determined by the molecular dynamics trajectory $\vec{q}(t)$:

[⊗] Abstract published in *Advance ACS Abstracts*, February 1, 1996.

$$\mathbf{K}_T = \langle (\bar{q}(t) - \langle \bar{q}(t) \rangle_{t_{\text{tot}}}) (\bar{q}(t) - \langle \bar{q}(t) \rangle_{t_{\text{tot}}})' \rangle_{t_{\text{tot}}} \quad (2.3)$$

where $\langle \rangle_{t_{\text{tot}}}$ denotes the time average over $0 \leq t \leq t_{\text{tot}}$. The eigenvalue λ_α is the time average of $|a_\alpha(t)|^2$.

The hope is to obtain a set of equations of motion for a small number of modes $a_\alpha(t)$ corresponding to the largest eigenvalues λ_α which provide a reduced representation of the long time dynamics of a protein. If one knew the correct basis vectors $\bar{\eta}_\alpha$, then one could obtain the equations with the aid of the (nonlinear) Galerkin²² method, as done in fluid dynamics.¹⁶ However, the shortness of the sampling time prevents the identification of suitable basis vectors $\bar{\eta}_\alpha$. The longest time scale τ of a protein scales as $\tau \sim n^\beta$, where n is the number of amino acid residues. The exponent is of order unity. It cannot be shorter than 2/3, i.e., the exponent which applies to diffusive relaxation of modes in a solid body. Probably the upper bound is given by a single-chain polymer in a good solvent with $\beta \approx 1.67$,²³ which may be interpreted as the fluctuation time scale of the completely unfolded protein conformation. According to NMR studies, the longest relaxation time cannot be smaller than 1 μs for a protein of the size $n \approx 300$.¹⁷ Thus, for many proteins the necessary sampling time must be extremely long. The longest simulation time realistically accessible today, even for a modest sized protein as G-actin, is a few nanoseconds. Thus, slow motions cannot be properly identified.

Even if we could manage to obtain the data for at least one relaxation time of the slowest mode, the existence of fluctuations prohibits an accurate determination of the correlations between the slow modes. Consequently, one cannot obtain reliable collective coordinates for them and, thus, one cannot use PCA to reduce protein dynamics without spoiling the most important slow motion.

There remains some hope that the PCA method, though not capable of extracting long time scale modes, may capture properly the modes with relaxation times well within the simulation time window. We will show below that this is not the case, either.

III. Principal Component Analysis of G-Actin

To illustrate the difficulties one encounters in the application of PCA to protein dynamics, we simulate the protein G-actin (globular actin). G-actin is the monomeric form of the polymeric protein system F-actin (filamentous actin), which is an important component of the cell cytoskeleton, where it is involved in cellular motility.^{24–26} F-actin also forms the major component of the thin filament of muscle tissue.^{27,28} The crystal structure of G-actin has been resolved by Kabsch and co-workers in 1990.²⁹ G-actin's 375 amino acid residues are organized in four subdomains, and in the present study the whole protein was simulated with explicit water molecules, a bound adenosine diphosphate nucleotide, and an associated calcium ion. It should be noted that a normal mode analysis of G-actin has been carried out by Tirion and ben-Avraham.³⁰ The resulting modes have been used to reduce the search span for building F-actin polymers from G-actin monomers.³¹

The molecular dynamics simulations were carried out with X-PLOR,³² using a 12 Å cutoff and an all-hydrogen force field, a 1 fs integration time step, and a dielectric constant $\epsilon = 1$. The protein was solvated in a 5.6 Å shell of explicit water molecules (1189 water molecules). The total system size was 9441 atoms. The system was first energy-minimized, then assigned initial atom velocities according to a Maxwell distribution, heated up to 300 K in steps of 30 K in a 5 ps time period,

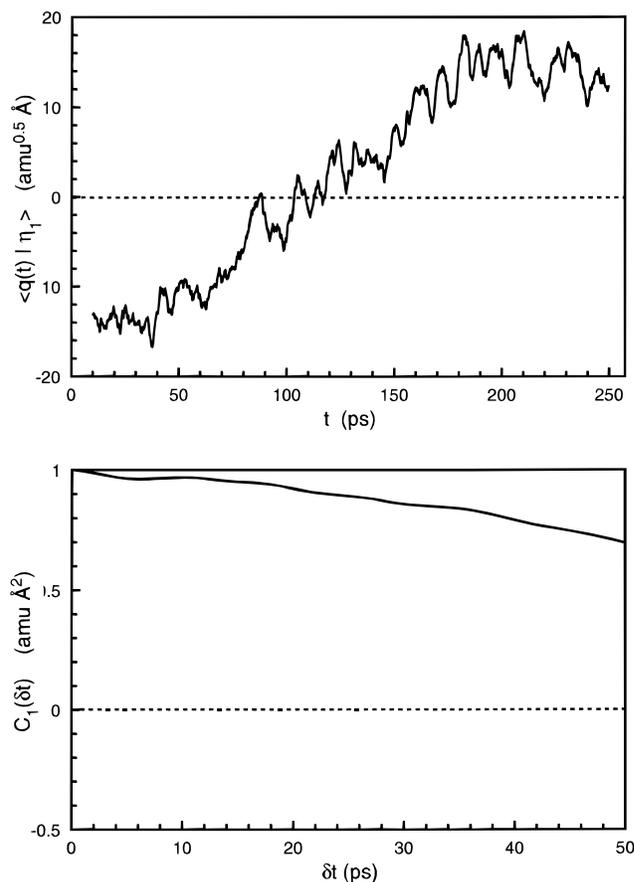


Figure 1. Projection of the trajectory along the largest principal component and its autocorrelation function.

and equilibrated at 300 K for 5 ps. Finally, free molecular dynamics was performed for 490 ps.

The trajectories have been analyzed after 20 ps of relaxation in the form of two consecutive periods of 235 ps. Coordinates were sampled at a 0.1 ps interval. The simulation results will be described elsewhere.³³

We sampled in our principal component analysis only the positions of the 375 C_α carbons resulting in $3N = 1275$ degrees of freedom. Amadei et al.^{13,34} demonstrated that the character of the fluctuation spectrum and even the identity of the larger amplitude modes do not change significantly by a reduction to C_α atoms.

We subtracted, as done customarily in PCA, the translational and rotational modes of the whole protein. These modes are conserved for the whole system and are expected to have very long relaxation times for the reduced set of variables we consider. The stated subtraction endows the covariance matrix with six vanishing eigenvalues associated with the conserved modes. The eigenvalues are actually not zero, since the subtraction of the rigid-body modes only involves the C_α values, but are much smaller than the next smallest eigenvalues. The former eigenvalues are on the order of $\lambda \approx 10^{-8} \text{Å}^2$, while the latter eigenvalues measure $\lambda \approx 10^{-4} \text{Å}^2$.

Figures 1–3 present the projections $a_\alpha(t)$ of the trajectory onto the eigenvectors corresponding to the three largest eigenvalues λ_α of the correlation matrix. The results demonstrate that motion along these modes exhibits a relaxation time which is longer than or comparable to the sampling window of 235 ps. Notice that if the relaxation time of a mode is significantly longer than the window width, its relaxation cannot be observed, and the mode does not show up in the PCA result. Consequently, significant modes reflecting the long term dynamics could be missed.

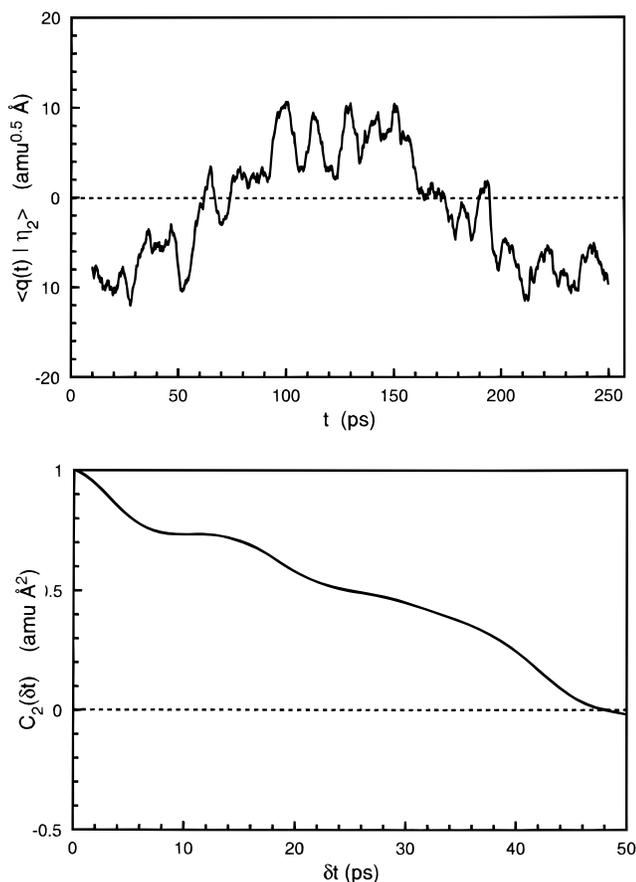


Figure 2. Projection of the trajectory along the second largest principal component and its autocorrelation function.

The eigenvalues λ_α of the covariance matrix \mathbf{K}_T and the correlation times τ_α of the associated motion along the eigenmodes are shown for the two 235 ps simulations in Figures 4 and 5. The times τ_α were obtained through

$$\tau = \int_0^{\tau_1} dt \frac{\langle [a_\alpha(t) - \langle a_\alpha(t) \rangle][a_\alpha(0) - \langle a_\alpha(t) \rangle] \rangle}{\langle a_\alpha^2(t) \rangle - \langle a_\alpha(t) \rangle^2} \quad (3.1)$$

where τ_1 is the shortest time at which the integrand, i.e., the autocorrelation function of $a_\alpha(t)$, vanishes. The τ_α vs λ_α curve obeys approximately the scaling laws $\tau \sim \lambda^{0.63}$ and $\lambda_\alpha \sim \alpha^{-1.5}$ for $\alpha \leq 200$. Although for a given length of the sampling period the eigenvalues λ_α and relaxation times τ_α do not depend on the actual time window, these quantities increase with the sampling time. This increase arises because longer times allow further explorations of the phase space and is consistent with the observation that the **rms** fluctuations of the atoms increase with the length of simulation.³⁴ This behavior is a clear indication of insufficient simulation time.

To demonstrate further how unreliable the obtained eigenmodes are, we compare the modes obtained for the consecutive time windows of the simulations. For this purpose we define the matrix

$$\mathbf{P}_{\alpha\beta} = \bar{\eta}_\alpha^{(1)} \cdot \eta_\beta^{(2)} \quad (3.2)$$

where $\eta_\alpha^{(i)}$ denotes the α th eigenvector obtained from the i th window ($i = 1, 2$) and where both α and β run from 1 to $M \ll 3N$.

If the motion was really constrained to an "essential" subspace, the projections of the principal components should be nearly diagonal or block diagonal, or at least one should

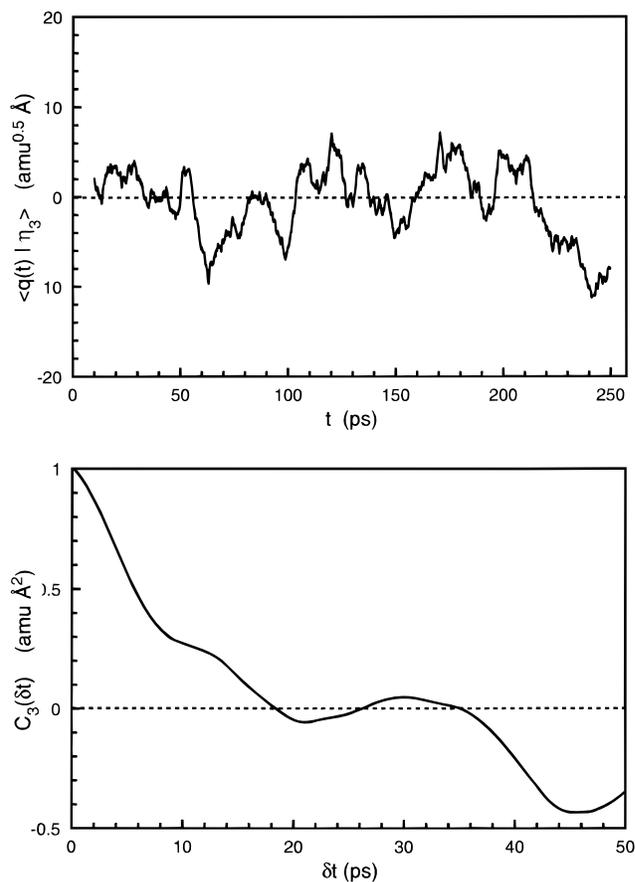


Figure 3. Projection of the trajectory along the third largest principal component and its autocorrelation function.

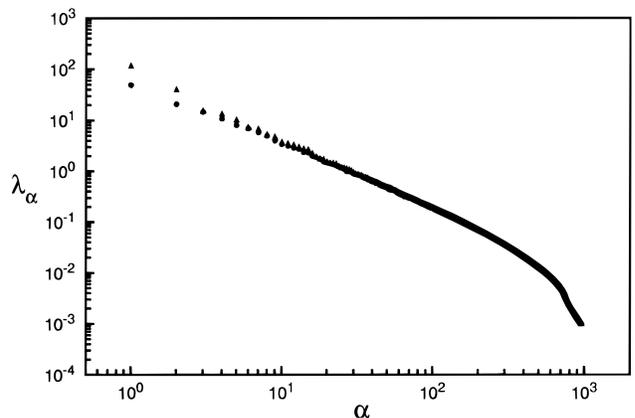


Figure 4. Mean square fluctuations of two consecutive 235 ps trajectories of G-actin. The dependence of λ on α can be fitted to $\lambda_\alpha \propto \alpha^{-1.5}$ for several decades.

recognize certain nonrandom patterns. However, this is not the case, as Figure 6 demonstrates; in fact there is very little structure in \mathbf{P} .

A most significant measure of the merit of the modes $\bar{\eta}_\alpha$ resulting from the PCA method is furnished by the dimension of the common subspace U of the subspaces spanned by $\{\bar{\eta}_\alpha^{(1)}\}$ and $\{\bar{\eta}_\alpha^{(2)}\}$. This dimension, given by

$$\text{tr}(\mathbf{P} \cdot \mathbf{P}') = \sum_{\alpha, \beta} \mathbf{P}_{\alpha\beta} \mathbf{P}'_{\beta\alpha} \quad (3.3)$$

is shown in Figure 7 as a function of M . For a small number of eigenvectors, $M = 10$, less than 20% of the 300 largest amplitude modes remain comparable in the sense of spanning

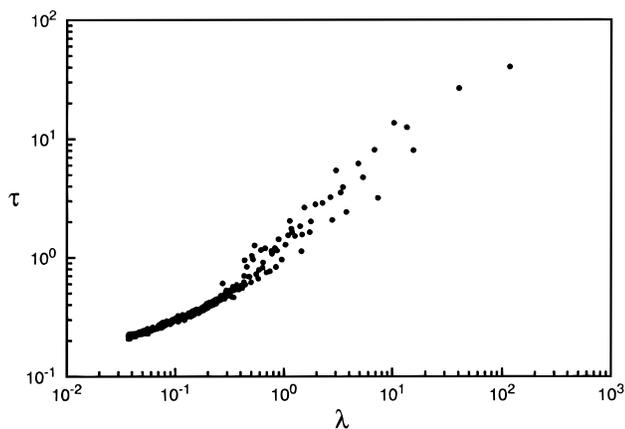


Figure 5. Correlation time versus mean square fluctuations of two consecutive 235 ps trajectories of G-actin. Note that the decay times exhibit a very good correlation with the amplitude fluctuation; the data can be fitted to $\tau \propto \lambda^{0.63}$.

the same linear vector space on going from the first to the second time window. For $M = 50$ apparently about 50% of the subspaces spanned by the modes overlap in successive windows, and for $M = 300$ almost 70% overlap. However, the reader should note, as shown in Figure 7, that purely randomly chosen subspaces have significant overlap as well. Thus, the quoted high overlap values do not attest to the reliability of the PCA modes in sampling the relevant subspace. In any case it is clear that reduction of dynamics cannot be achieved with a small number of the largest amplitude modes. Even if one weights the modes with the amplitudes in order to measure the amount

of fluctuations in the common subspace U , the situation does not improve significantly.

We conclude from the PCA analysis of molecular dynamics trajectories of G-actin that the PCA method fails to capture the essential subspace. We can explain this failure through an analysis of a model system which is exactly recoverable by PCA if sufficient sampling time is provided.

IV. Why PCA Does Not Work for System Reduction

The actual motion of the modes in a protein are (non)linearly coupled, so the $a_\alpha(t)$'s (2.1) are dynamically correlated. Hence, the nonrelaxing slow modes slave faster modes (that is, the parameters of the faster modes are dependent on the slow modes), making the short time estimation of the faster modes unreliable. However, as we will see below, even without any dynamical couplings, an insufficient length of sampling time yields mode vectors $\vec{\eta}_\alpha$ which suffer from strong statistical errors.

We ask the following question. Can we recover the equation

$$\frac{da_\alpha}{dt} = \frac{1}{\tau_\alpha} a_\alpha + N_\alpha \quad (4.1)$$

by PCA with limited sampling time? Here we assume N_α represents Gaussian white noise with zero mean and correlation function

$$\langle N_\alpha(t) N_\beta(0) \rangle = 2k_B T \frac{\lambda_\alpha}{\tau_\alpha} \delta(t) \delta_{\alpha\beta} \quad (4.2)$$

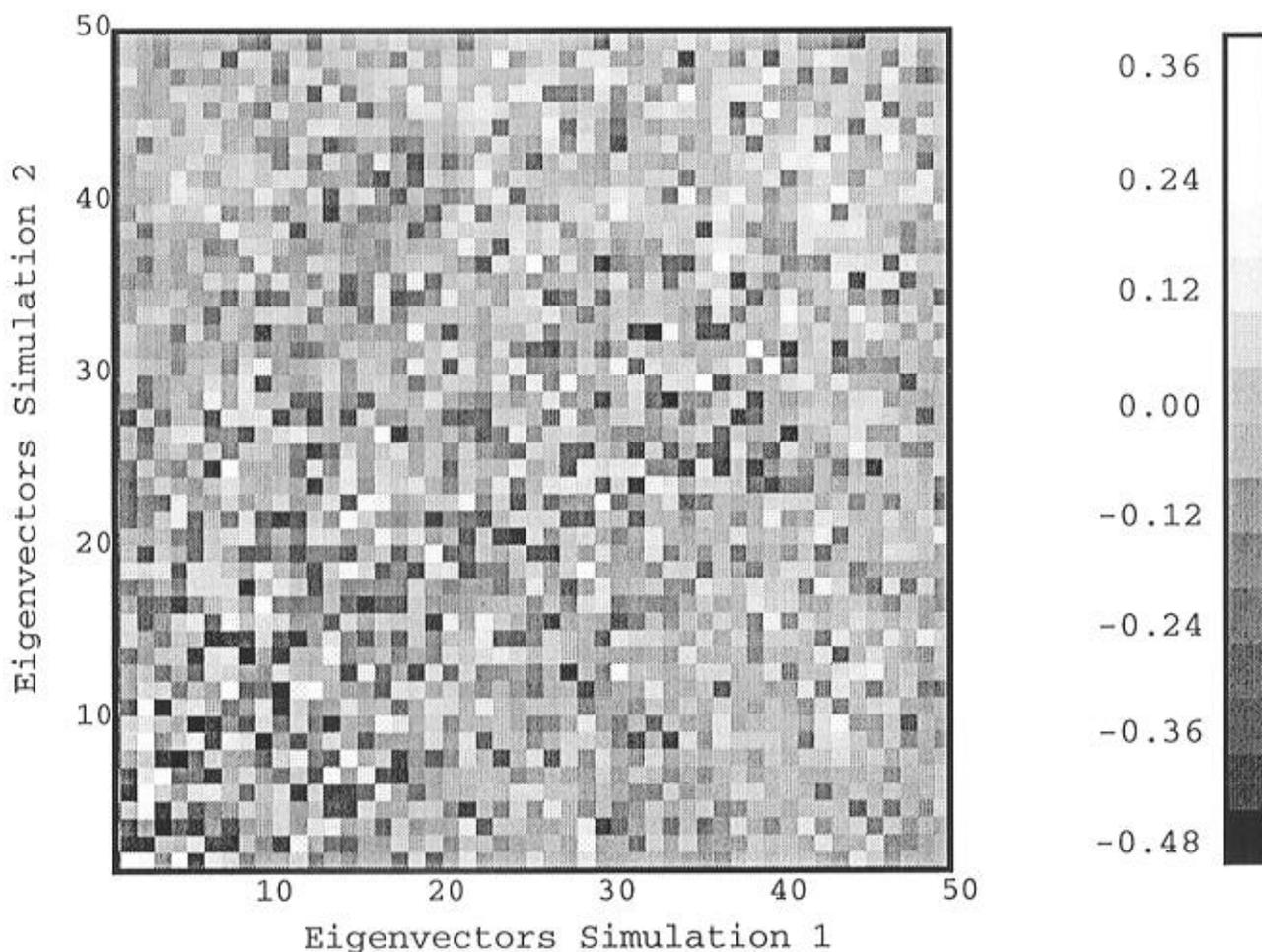


Figure 6. Comparison of the principal components for two consecutive 235 ps trajectories of G-actin provided by the overlap matrix (3.2).

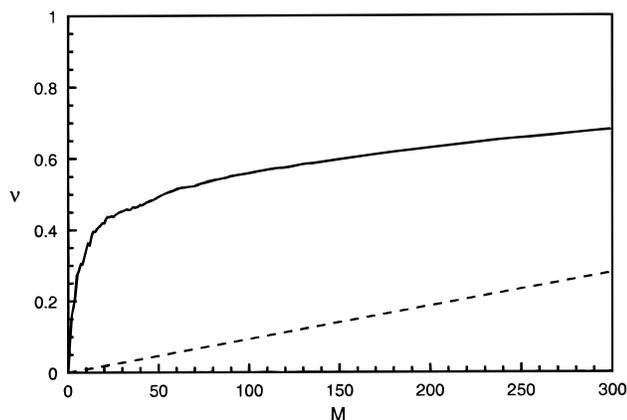


Figure 7. $\nu = \text{tr}(\mathbf{P} \cdot \mathbf{P})/M$ as a function of M (see (3.3)). The dashed line represents the expected value for ν if both subspaces were formed by sets of random, uncorrelated vectors.

where $\langle \rangle$ denotes the average over noise. Henceforth, we choose the unit of energy so that $k_B T \equiv 1$. Equations 4.1 and 4.2 characterize $a_\alpha(t)$ as a Gaussian process with zero mean whose correlation function is given by

$$\langle a_\alpha(t_2) a_\beta(t_1) \rangle = \lambda_\alpha e^{-|t_2 - t_1|/\tau_\alpha} \delta_{\alpha\beta} \quad (4.3)$$

If the true correlation matrix \mathbf{K} , which is diagonal, is given, then by the Galerkin method²² we can trivially recover the original linear equation of motion. Thus, the question of whether we can obtain the correct equation of motion boils down to the question of an accurate estimate of the equal time correlation matrix. We study the variance of $(\mathbf{K}_T)_{\alpha\beta} \equiv \langle a_\alpha(t) a_\beta(t) \rangle_{t_{\text{tot}}}$, where $\langle \rangle_{t_{\text{tot}}}$ is the time average over the sampling time t_{tot} , which can be written as

$$\frac{1}{t_{\text{tot}}} \int_0^{t_{\text{tot}}} dt_1 \frac{1}{t_{\text{tot}}} \int_0^{t_{\text{tot}}} dt_2 [\langle a_\alpha(t_1) a_\beta(t_1) a_\alpha(t_2) a_\beta(t_2) \rangle - \langle a_\alpha(t_1) a_\beta(t_1) \rangle \langle a_\alpha(t_2) a_\beta(t_2) \rangle] \quad (4.4)$$

Since the a 's are mutually statistically independent Gaussian variables, using (4.3) we obtain³⁵

$$(\delta \mathbf{K}_{T\alpha\beta})^2 = (1 + \delta_{\alpha\beta}) \left(\frac{t_{\text{tot}}}{\tau_\alpha} + \frac{t_{\text{tot}}}{\tau_\beta} \right)^{-1} (\lambda_\alpha \lambda_\beta) \quad (4.5)$$

The variance is small only if the decay times of the modes τ_α and τ_β are much smaller than the simulation time t_{tot} . Only in that case will \mathbf{K}_T be a good approximation to the true correlation matrix which can be used to recover the dynamics.

Looking at (4.5), the reader might think that if the time scale of mode α is sufficiently small (i.e., $\tau_\alpha/t_{\text{tot}} \ll 1$), then the variance of $(\mathbf{K}_T)_{\alpha\beta}$ is small even if $\tau_\beta/t_{\text{tot}}$ is not small. However, in this case $(\mathbf{K}_T)_{\alpha\alpha}$ is also small, so we lose the diagonal nature of the true correlation matrix. This can be explicitly seen from

$$\frac{\langle \delta \mathbf{K}_{\alpha\beta}^2 \rangle}{\lambda_\alpha^2} = \frac{\tau_\alpha \lambda_\beta}{t_{\text{tot}} \lambda_\alpha} \quad (4.6)$$

As seen in Figures 4 and 5, $\lambda_\alpha \propto \alpha^{-1.5}$ and $\tau \propto \lambda^{0.63}$. Hence, $\tau_\alpha/\lambda_\alpha \propto \alpha^{0.56}$, so for large α (for small amplitude modes) this ratio diverges; this asymptotic relation holds for α sufficiently smaller than $3N$ but much larger than unity. Therefore, even fast modes whose relaxation time is well within the sampling window cannot be recovered by PCA.

V. Concluding Remarks

The PCA method has been applied previously to various proteins.^{11–15} From these studies and the study provided here

we conjecture the following universal features: there are scaling relations $\tau_\alpha \sim \alpha^{-\theta}$ and $\lambda_\alpha \sim \alpha^{-\zeta}$ for a wide range of α , where α is the mode index. In the molecular dynamics investigation of G-actin, we estimate $\theta = 1.5$ and $\zeta = 0.89$. These values do not depend on the particular choice of the sampling window so long as its width is the same. Even if the sampling window width is increased, such power law behaviors are observed. However, the relaxation times and amplitudes increase for each α ; that is, $\tau_\alpha^{(1)} > \tau_\alpha^{(2)}$ and $\lambda_\alpha^{(1)} > \lambda_\alpha^{(2)}$ for each α , if the window (1) is wider than window (2). Needless to say, the α th mode for window (1) and that for (2) can be completely different.

It is claimed that a very small number of PCA modes dominate protein motion; in fact, usually around 1% of the modes capture 90% of the total rms fluctuations of a molecular dynamics trajectory.¹³ However, these dominant modes change from one sampling window to another. Consequently, they cannot serve to describe the motion in the later time windows. Therefore, such modes do not have predictive power for the long time behavior of proteins.

If one increases the width of the analysis window, more and more slower modes will acquire larger amplitudes and show up as dominant modes. Consequently, the top 10 modes with largest λ_α values in our 235 ps sampling would become minor modes, even if preserved, if a longer sampling time is chosen.

In summary, PCA does not reveal any reliable information on time scales which are not actually sampled. A short molecular dynamics trajectory does not yield an accurate covariance matrix (\mathbf{K}_T) of protein fluctuations. Consequently, any PCA modes cannot be determined reliably.

Still, there are possible uses of PCA. As has been demonstrated,^{11–15} one can extract salient motions observed in the sampling window. There is no mystery here; PCA is designed to extract salient features from time series. The problems in this use of PCA are, however, (i) that the modes apply only for the analyzed window and (ii) that PCA results can be obtained directly probably without the cumbersome diagonalization of the correlation matrix (2.3).

A more nontrivial use of the PCA modes is to study local short time scale motion of proteins as relevant to NMR spectroscopy.²¹ In this case what is needed is the summation of the contribution from many fast modes for a given atom or groups of atoms, so that the law of large numbers allows us to overcome the sampling error. That is, even though each mode is highly unreliable, its sum may become reliable. This is, however, true only for the quantities to which no long time contribution is significant. Thus, Debye–Waller factors cannot be computed reliably with the aid of the PCA modes. Diffusive light scattering is another problematic example, as clearly demonstrated in recent papers by Clarage et al.¹⁹ and Faure et al.²⁰

Acknowledgment. The authors thank Joost Beckers for drawing our attention to ref 13 and Tamar Schlick as well as Attila Szabo for reading the manuscript. The present work has been supported by the National Science Foundation Grant NSF DMR 93-1938, the National Institutes of Health Grant PHS 5 P41 RR05969-04, the National Science Foundation Grant NSF BIR 93-18159, the MCA93S028P computer time grant at Pittsburgh Supercomputing Center, the Roy J. Carver Charitable Trust (W.W.), and the La Caixa Fellowship Program (M.A.B.).

References and Notes

- (1) Karplus, M. *Molecular Dynamics: Applications to Proteins*. In Rivail, J.-L., Ed.; *Modelling of Molecular Structures and Properties*, volume 71 of *Studies in Physical and Theoretical Chemistry*; Elsevier Science Publishers: Amsterdam, 1990; pp 427–461.

- (2) Brooks, C. L.; et al. *Proteins*; Wiley-Interscience: New York 1988; Chapter 2.B, pp 14–21.
- (3) yan Chang, X.; Freed, K. F. *J. Chem. Phys.* **1993**, *99* (10), 8016–8030.
- (4) Nyberg, A.; Schlick, T. *Chem. Phys. Lett.* **1992**, *198*, 538.
- (5) Karhunen, K. *Ann. Acad. Sci. Fenn.* **1947**, 37.
- (6) Brooks, B. B.; et al. *J. Comput. Chem.* **1995**, *16* (12), 1522–1542.
- (7) Case, D. A. *Curr. Opin. Struct. Biol.* **1994**, *4*, 285–290.
- (8) Karplus, M.; Jushick, J. N. *Macromolecules* **1981**, *14*, 325–332.
- (9) Perahia, D.; et al. *Biopolymers* **1990**, *29*, 645–677.
- (10) Ichiye, T.; Karplus, M. *Proteins* **1991**, *11*, 205–217.
- (11) Horiuchi, T.; Go, N. *Proteins* **1990**, *10*, 106–116.
- (12) Kitao, A.; et al. *J. Chem. Phys.* **1991**, *158*, 447–472.
- (13) Amadei, A.; et al. *Proteins* **1993**, *17*, 412–425.
- (14) van Aalten, D. M. F.; et al. *Proteins* **1995**, *22*, 45–54.
- (15) Garcia, A. E. *Phys. Rev. Lett.* **1992**, *68*, 2696–2699.
- (16) Sirovich, I.; Rodriguez, J. D. *Phys. Lett. A* **1987**, *120*, 211–214.
- (17) Nicholson, L. K.; et al. *Struct. Biol.* **1995**, *2*, 274–280.
- (18) Hodel, A.; et al. *Protein Sci.* **1995**, *4*, 636–654.
- (19) Clarage, J. B.; et al. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 3288–3292.
- (20) Faure, P.; et al. *Struct. Biol.* **1994**, *1* (2), 124–128.
- (21) Bruschiweiler, R.; Case, D. A. *Phys. Rev. Lett.* **1994**, *72*, 940–943.
- (22) Peyret, R.; Taylor, T. D. *Computational Methods for Fluid Flow*; Springer Verlag: Berlin, 1983.
- (23) Doi, M.; Edwards, S. F. *The Theory of Polymer Dynamics*; Oxford Science Publications: Oxford, 1986.
- (24) Devreotes, P. N.; Zigmond, S. H. *Ann. Rev. Cell Biol.* **1988**, *4*, 649–686.
- (25) Small, J. V. *Curr. Opin. Cell Biol.* **1989**, *1*, 75–79.
- (26) Theriot, J. A.; et al. *Nature* **1992**, *357*, 257–260.
- (27) Holmes, K. C.; et al. *Nature* **1990**, *347*, 44–49.
- (28) Lorenz, M.; et al. *J. Mol. Biol.* **1993**, *234*, 826–836.
- (29) Kabsch, W.; et al. *Nature* **1990**, *347*, 37–44.
- (30) Tirion, M. M.; ben-Avraham, D. *J. Mol. Biol.* **1993**, *230*, 186–195.
- (31) Tirion, M. M.; et al. *Biophys. J.* **1995**, *68*, 5–12.
- (32) Brünger, A. T. *X-PLOR, Version 3.1, A System for X-ray Crystallography and NMR*; The Howard Hughes Medical Institute and Department of Molecular Biophysics and Biochemistry, Yale University, 1992.
- (33) Wriggers, W.; Schulten, K. Manuscript in preparation.
- (34) Brooks, B. B.; et al. *J. Comput. Chem.* **1995**, *16* (12), 1554–1566.
- (35) Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*; Oxford University Press: Oxford, 1987; Chapter 6.4.1.

JP9536920