Apo Adenylate Kinase Encodes Its Holo Form: A Principal Component and Varimax Analysis

Robert I. Cukier*

Department of Chemistry and the Quantitative Biology Modeling Initiative, Michigan State University, East Lansing, Michigan 48824

Received: June 18, 2008; Revised Manuscript Received: November 14, 2008

Adenylate kinase undergoes large-scale motions of its LID and AMP-binding (AMPbd) domains when its apo, open form closes over its substrates, AMP and Mg^{2+}-ATP. It may be an example of an enzyme that provides an ensemble of conformations in its apo state from which its substrates can select and bind to produce catalytically competent conformations. In this work, the fluctuations of the enzyme apo *Escherichia coli* adenylate kinase (AKE) are obtained with molecular dynamics. The resulting trajectory is analyzed with principal component analysis (PCA) that decomposes the atom motions into orthogonal modes ordered by their decreasing contributions to the total protein fluctuation. In apo AKE, a small set of the PCA modes describes the bulk of the fluctuations. Identification of the atom motions that are important contributors to these modes is improved with the use of a varimax rotation method that rotates the PCA modes to a new mode set that concentrates the atom contributions to a smaller set of atoms in these new modes. In this way, the nature of the important motions of the LID and AMPbd domains are clarified. The dominant PCA modes are used to investigate if apo AKE can fluctuate to conformations that are holo-like, even though the apo trajectory is mainly confined to a region around the initial apo structure. This is accomplished by expressing the difference between the protein coordinates, obtained from the holo and apo crystal structures, using as a basis the PCA modes from the apo AKE trajectory. The coherent motion described by a small set of the apo PCA modes is shown to be able to produce protein conformations that are quite similar to the holo conformation of the protein. In this sense, apo AKE does encode in its fluctuations information about holo-like conformations.

1. Introduction

Adenylate kinase (AK) enzymes catalyze the reversible transformation Mg^{2+}-ATP + AMP ⇔ Mg^{2+}-ADP + ADP.[1] AKs possess a core, a LID, and an AMP-binding (AMPbd) domain. The apo forms of AKs exist in an open form, while in the presence of ATP, AMP, and Mg^{2+} the LID and AMPbd domains undergo major conformational rearrangements, resulting in the enzyme closing to form the ternary complex and expelling waters to prevent ATP and AMP hydrolysis.[2] Kinetic studies established that the reaction occurs by a random bi bi mechanism, whereby the substrates can bind to one isoform while the products bind to another isoform.[3] These features suggest that apo-AKE can exist in at least two conformations, whereby one form of the enzyme can bind Mg^{2+}-ATP and AMP and the other Mg^{2+}-ADP and ADP. There also is evidence[4,5] suggesting that each of these forms of the enzyme produces its own subensemble of conformers. NMR studies of the apo enzyme provide evidence for a conformational ensemble.[6] Crystallographic studies[2] and energy-transfer experiments[7] indicate a two-phase process whereby binding AMP leads to one conformational change and then binding ATP, with this "ternary complex" usually modeled with AP_5A (ATP and AMP linked by a fifth phosphate group), results in further conformational changes resulting in the closed, catalytically competent form. The transition from apo (open) to closed AK requires large conformational changes in which the LID and AMPbd domains undergo extensive translational and rotatational displacements.[2] Large conformational changes in adenylate kinase have been investigated on the basis of the elastic network model (ENM), which provides a reduced description relative to atomistic approaches.[8−10] Temiz et al.[10] used a form of the ENM method and found that the lowest mode of the apo form describes a closing motion of the LID domain toward the AMPbd domain required for ligand binding. We[11] used DynDom[12] to find the locations of hinges in AK based on an analysis of the apo and ternary complex structures, consistent with the results of other related methods.[8,9]

In previous work, we carried out molecular dynamics (MD) simulations of apo-AKE at ambient and elevated temperatures to contrast their predictions regarding the LID and AMPbd domain motions.[13] The potential of mean force along a reaction coordinate that reflects the open to closed motion of AKE shows a distinct minimum around the distance corresponding to the holo structure distance for binding substrates.[11] This finding suggests that apo-AKE does have stable conformations that are suited to binding its substrates.

These large conformational rearrangements of AK in response to the presence of ligands were taken[2] as evidence of the induced fit mechanism, in contrast to the lock and key concept appropriate to other enzymes. The history of ligands binding to proteins forms a progression of models that in essence increase the role of protein plasticity in the binding of ligands.[14−17] The "lock and key" concept asserts that a protein has a cavity into which a ligand can be fit into with minor rearrangements of protein and ligand. In contrast, the induced fit model suggests that a ligand induces a conformational change at the binding site, shifting it toward an activated state. Of more recent vintage, the preexisting equilibrium hypothesis,[18] with

* To whom correspondence should be addressed. Phone: 517-355-9715ext. 263. Fax: 517-353-1793. E-mail: cukier@cem.msu.edu.

10.1021/jp8053795 CCC: $40.75
Published on Web 01/21/2009

origins in funnel energy landscape protein-folding concepts,[19,20] asserts that the native state of the protein exhibits an ensemble of conformations that can span apo to more ligand-bound-like conformations. A ligand can select from a large array of conformations and then may bias the equilibrium toward the catalytically competent conformation.

An important potential implication of conformational selection is that the protein by itself *encodes* the direction of fluctuations that are appropriate for ligand selection and binding. Proteins may be designed such that their apo fluctuations have embedded information about conformations that are favorable to trap ligands. This encoding hypothesis has been suggested a number of times and in various forms.[14−17,21−24,13,11,25−30] There are two approaches to this issue that are most closely allied to the scheme pursued in this work. Ikeguchi et al. used (equilibrium) linear response theory,[29] where the perturbation is a ligand, to show that apo protein fluctuations are appropriate for providing good conformers for ligand binding. They found that suitably applied forces, associated with sites in the protein close to where known ligands bind, when vectorially multiplied with the correlation matrix of the apo fluctuations, could predict the direction of protein motion for ligand binding. Lange et al.[30] obtained very long apo protein MD trajectories (∼200 ns) and showed that short slices of the long trajectory could be used to predict distinct conformations that were found in the long trajectory. They did so by showing that the dominant modes of motion of the short time slices, obtained by a principal components analysis (PCA), describe a major subspace of the long trajectory conformation space fluctuations, which includes conformations that could be appropriate for ligand binding.

The encoding hypothesis is a stronger statement than the preexisting equilibrium hypothesis in the sense that rather than the ligand selecting from a very broad array of conformations, the conformational space explored by the apo protein is directed toward specific conformations that are "set up" for ligand binding. Of course, a continuum of possibilities exist between the limits of a broad, diffuse ensemble, from which a ligand happens to select, and a specific set of encoded conformations that are ideal for ligand binding. Exploring the encoding notion on the basis of MD simulations is problematic because integrating Newton's equations of motion with a femtosecond time step will have difficulty reaching times that can capture the conformational transitions and fluctuations that may be occurring on the microseconds to even seconds time scales. Various strategies to surmount barriers in the complex configuration space of a protein, and thereby accelerate the ability of MD to sample configuration space, have been proposed and are under active development.[31]

In this work, we pursue another approach, relying on the encoding hypothesis, which focuses on the apo fluctuations of a protein and tries to identify the important motions of the protein that are "buried" in its fluctuations. If the apo protein has identifiable motions that in some sense point toward conformations that are holo-like, then one can conclude that the apo fluctuations encode ligand-bound conformations. To pursue this approach, we employ a PCA-based method that uses the covariance matrix of the atom coordinate fluctuations from their respective average values to decompose the protein motions into modes that capture significant amounts of the overall (mean square) protein fluctuations. We introduce a rotation of a set of these PCA modes known as a varimax rotation that has been used in other disciplines.[32,33] The objective of the varimax rotation is to maximize, in an average sense, the atom contributions to the modes. That is, many atoms often have significant amplitude in a given PCA mode. The varimax rotation seeks to minimize the number of atoms that significantly contribute to the modes. The reduction in the number of atoms with significant amplitude in a mode can simplify a modes interpretation and potentially provide a clearer picture of the important protein motions.

While, as we show, a varimax analysis does provide a clearer view of the protein atoms that contribute to the various modes, a separate analysis is required to see if holo-like conformations are encoded in the apo motion. We express the difference between known (from crystal structures) apo and holo configurations of AKE in the basis of the apo fluctuations obtained from an MD trajectory of the apo-based structure and investigate how well a small number of apo PCA modes can predict holo-like conformations. The encoding analysis does not rely on the varimax rotation—it uses a sum over a number of modes and the prediction is invariant to this rotated basis.

The remainder of this work is organized as follows. In Methods (section 2), after introducing the PCA in a form suited to our purposes, we present a version of the varimax procedure appropriate for analyzing a MD trajectory. In Results (section 3), the implications of the varimax-rotated modes to the motion of AKE are given and the degree to which the apo fluctuations of AKE describe holo-like conformations is explored. Concluding remarks are presented in section 4.

2. Methods

2.1. Principal Component Analysis. A PCA[34−38] diagonalizes the covariance matrix **C**

$$\mathbf{C} = \frac{1}{T}\int_0^T \delta\mathbf{R}(t)\,\delta\mathbf{R}^T(t)\,\mathrm{d}t = \langle \delta\mathbf{R}(t)\,\delta\mathbf{R}^T(t)\rangle \quad (2.1)$$

of the atom fluctuations from their trajectory-averaged values, $\delta\mathbf{R}(t) = \mathbf{R}(t) - \langle\mathbf{R}(t)\rangle$, where $\mathbf{R}(t) = (x_1(t),y_1(t),..,z_N(t))^T$ denotes configurations along the trajectory and $\langle ...\rangle$ denotes a time average over trajectory snapshots. A PCA decomposes the trajectory as

$$\mathbf{R}(t) = \sum_{\nu=1}^{3N} [\mathbf{R}(t)\cdot\mathbf{m}_\nu^T]\mathbf{m}_\nu \equiv \sum_{\nu=1}^{3N} p_\nu(t)\mathbf{m}_\nu \quad (2.2)$$

where the $\mathbf{m}_\nu$ ($\nu = 1,2,..., 3N$) are the (orthonormal) eigenvectors (principal vectors) that will be referred to as PCA modes of the covariance matrix, and the corresponding eigenvalues of **C** are denoted as λ_ν^2. The principal component $p_\nu(t)$ is the projection of the trajectory onto the νth eigenvector. In the rotated Cartesian coordinate basis defined by the $\mathbf{m}_\nu$, the largest eigenvalue λ_1^2 captures the largest fraction of the mean square fluctuation (MSF), the second largest the next largest fraction of the MSF, etc. Ordering the eigenvalues from large to small leads, in favorable cases, to a small set of modes that capture most of the protein's fluctuation. To examine the time course of the displacement of atom j corresponding to some set of PCA modes, eq 2.1 is used in the form

$$\mathbf{R}^j(t) = \mathbf{R}^j(0) + \sum_\nu (p_\nu(t) - p_\nu(0))\mathbf{m}_\nu^j \quad (2.3)$$

where the sum runs over the included modes. The contribution of atom j to the νth PCA mode's fluctuation is obtained from

$$|\mathbf{m}_\nu^j| = \sqrt{(m_\nu^{jx})^2 + (m_\nu^{jy})^2 + (m_\nu^{jz})^2} \tag{2.4}$$

The $\mathbf{m}_\nu^j = \{m_\nu^{jx}, m_\nu^{jy}, m_\nu^{jz}\}$ will be referred to as *component vectors* of the jth atom for the νth mode. The total *MSF* can be decomposed as

$$\begin{aligned} \mathrm{MSF} &= \sum_\nu \lambda_\nu^2 = \sum_\nu \lambda_\nu^2 \mathbf{m}_\nu^T \cdot \mathbf{m}_\nu = \sum_\nu \sum_j \lambda_\nu^2 \mathbf{m}_\nu^{jT} \cdot \mathbf{m}_\nu^j \\ &= \sum_\nu \sum_j \lambda_\nu{}^2 (m_\nu^{jx} m_\nu^{jx} + m_\nu^{jy} m_\nu^{jy} + m_\nu^{jz} m_\nu^{jz}) \\ &\equiv \sum_\nu \sum_j (RC_\nu^j)^2 \end{aligned} \tag{2.5}$$

The relative component RC_ν^j is the contribution of atom j in mode ν to the total fluctuation of the protein, with inclusion of the eigenvalue size. This makes it possible to compare the importance of an atom across different modes, thus the name relative component.

2.2. Varimax Rotation. The idea of the varimax method, as applied to the analysis of atomistic trajectory data, is to rotate a set of the PCA vectors to maximize the contribution of some of the atoms (minimize the number of atoms that significantly contribute) to each vector in the rotated set.[32,33] The $\nu_{\max}$ new rotated vectors, which we denote by $\tilde{m}_\nu$, and refer to as VMX modes, are formally related to the PCA $\mathbf{m}_\nu$ by an orthogonal transformation matrix with elements $T_{\mu\nu}$

$$\tilde{\mathbf{m}}_\nu = \sum_{\mu=1}^{\nu_{\max}} T_{\mu\nu} \mathbf{m}_\mu \quad (\mu, \nu = 1, 2, ..., \nu_{\max}) \tag{2.6}$$

Since the transformation is orthogonal, the new vectors also form an orthonormal set and still represent independent modes of motion. In the varimax scheme, the rotation is defined as the one that maximizes the variance, VAR, of the rotated orthogonal vectors $\tilde{m}_\nu$

$$\begin{aligned} \mathrm{VAR} &= \sum_{\nu=1}^{\nu_{\max}} \sigma_\nu{}^2 \\ &= \sum_{\nu=1}^{\nu_{\max}} \mathrm{var}_\nu(\delta \tilde{\mathbf{m}}_\nu^T \cdot \delta \tilde{\mathbf{m}}_\nu) \\ &= \sum_{\nu=1}^{\nu_{\max}} \left\{ \frac{1}{3N} [\sum_{j=1}^{3N} (\tilde{m}_\nu^j)^4] - \left[\frac{1}{3N} \sum_{j=1}^{3N} (\tilde{m}_\nu^j)^2 \right]^2 \right\} \end{aligned} \tag{2.7}$$

The jth element $\tilde{m}_\nu^j$ of the νth vector consists of the values of the component vector (3-vector) of the jth atom's displacement in that mode. The range of VAR is readily obtained as $0 \leq \mathrm{VAR} \leq (\nu_{\max}/3N)(1 - 1/3N)$ with the lower limit (most dispersed) arising from all $\tilde{m}_\nu^j$ equal (all the atoms contributing equally), and the upper limit (most concentrated) arising from one component equal to $1/(3N)^{1/2}$ and all others zero. Thus, an increase in variance corresponds to a decrease in the number of atoms with significant amplitudes in the rotated modes.

The procedure is actually accomplished[39] by rotating all $\nu_{\max}$ $(\nu_{\max} - 1)/2$ pairs of vectors in turn in each cycle (one pass through all the pairs). $\sigma_{\nu_i}{}^2 + \sigma_{\nu_k}{}^2$ can be maximized analytically[39,32] for each mode pair indexed by ν_i and ν_k because a two-dimensional rotation can be expressed using one angle. VAR is then calculated for a cycle, and the procedure is iterated until VAR no longer increases. The varimax method maximizes the total VAR; therefore the individual $\sigma_\nu{}^2$ may not all be maximized, but for the data considered here all $\sigma_\nu{}^2$ are quite close to their largest values.

By construction, because the PCA principal components are uncorrelated, $\langle p_\nu(t)\, p_\mu(t) \rangle = 0$ $(\mu \neq \nu)$, as shown in Appendix A, any rotation of the coordinate system to new axes must correlate the new principal components. Thus, the resulting rotated vectors are still orthogonal, but the corresponding varimax components $\tilde{p}_\nu(t) = \tilde{\mathbf{m}}_\nu^T \cdot \Delta\mathbf{R}(t)$ are correlated:

$$\begin{aligned} \langle \tilde{p}_{\mu_1}(t)\, \tilde{p}_{\mu_2}(t) \rangle &= \frac{1}{T} \int (\tilde{\mathbf{m}}_{\mu_1}^T \cdot \Delta\mathbf{R}(t))(\Delta\mathbf{R}^T(t) \cdot \tilde{\mathbf{m}}_{\mu_1})\, \mathrm{d}t \\ &= \sum_{\nu=1}^{\nu_{\max}} \lambda_\nu{}^2 T_{\nu\mu_1} T_{\nu\mu_2} \end{aligned} \tag{2.8}$$

where we used eq 2.6 and the orthogonality of the $\mathbf{m}_\nu$. Giving up the property of uncorrelatedness does sacrifice the eigenvalue ordering of the PCA; however, if interest centers on a description provided by a set of modes, then, because the variance totaled over those modes is not changed by the rotation, little is lost. On the other hand, much may be gained, as discussed below, by concentrating the atom contributions in the VMX modes.

Using the varimax-rotated vectors, the MSF can be reexpressed in their basis as

$$\begin{aligned} \mathrm{MSF} &= \sum_\nu \lambda_\nu{}^2 = \sum_\nu \lambda_\nu{}^2 \tilde{\mathbf{m}}_\nu^T \cdot \tilde{\mathbf{m}}_\nu \\ &= \sum_\nu \sum_j \lambda_\nu{}^2 (\tilde{m}_\nu^{jx} \tilde{m}_\nu^{jx} + \tilde{m}_\nu^{jy} \tilde{m}_\nu^{jy} + \tilde{m}_\nu^{jz} \tilde{m}_\nu^{jz}) \\ &\equiv \sum_\nu \sum_j (RC_\nu^j)^2 \end{aligned} \tag{2.9}$$

where, for convenience, we do not indicate with a new symbol that the relative components are constructed from a set of varimax vectors. If the VMX vectors have concentrated particular atom coordinates at the expense of other atom coordinates (the vectors are unit normalized) relative to PCA vectors, then a simplification of the description of the motion of the protein characterized by that mode should occur. That may be monitored by examining the RC_ν^j versus j plots of PCA and VMX modes.

Furthermore, the more concentrated the vectors into particular atom coordinates, the better they are at picking out nonlinear motions, e.g., rotations. That is simply understood as follows. First, note that any one mode can only describe linear motions of atoms, because all the atoms must move in phase. Rotations can then only be characterized by simultaneous consideration of two or more modes. For a pair of modes, significant rotational motions will be revealed by finding modes ν and μ where RC_ν^j and RC_μ^j are large *and* the angle between these vectors, $\cos\Theta_{\nu\mu}^j$, is close to zero (so approaching orthogonality), where

$$\cos\Theta_{\nu\mu}^j = (\mathbf{m}_\nu^j / RC_\nu^j) \cdot (\mathbf{m}_\mu^j / RC_\mu^j) \tag{2.10}$$

Recall that $\mathbf{m}_\nu = (\mathbf{m}_\nu^1, ..., \mathbf{m}_\nu^j, ..., \mathbf{m}_\nu^N)^T$ with $\mathbf{m}_\nu^j$ the component vector (3-vector) for the jth atom in mode ν. If, for example, two mode vectors are both dominated by the same atom, j, then

TABLE 1: PCA Mode Eigenvalues

mode	λ_ν (Å)	λ_ν^2 (Å^2)	cumulative fraction
1	0.675159	0.455839	0.455839
2	0.404124	0.163316	0.619155
3	0.267685	0.071655	0.690811
6	0.160294	0.025694	0.805651
12	0.086067	0.007408	0.875595
24	0.049385	0.002439	0.924789
48	0.029105	0.000847	0.958338

the requirement of orthogonality ($\mathbf{m}_\nu^T \cdot \mathbf{m}_\mu \approx \mathbf{m}_\nu^{jT} \cdot \mathbf{m}_\mu^j = 0$) shows that this atom's motion (as described by these two modes) can be expressed in an orthonormal basis and could reveal in an extreme case a circular trajectory. Thus, giving up the property of uncorrelatedness can be a useful tradeoff in exchange for concentrating the atom contributions to the vectors. The more atoms with significant weights in a given mode, the more the atom contributions will be dispersed, and rotations, for example, will be correspondingly harder to detect.

2.3. Molecular Dynamics. The MD simulations were carried out with the Sander module of Amber 8[40] using the Amber 94 force field.[41] For the electrostatic energies and forces, the PME method[42] was applied with a direct-space cutoff of 9.0 Å, a Ewald coefficient of 0.34864, and a 80 × 80 × 80 reciprocal space grid. The starting configuration was taken from the apo crystal structure of *Escherchia coli* (*E. coli*) adenylate kinase[43] (PDB code 4AKE). The crystallographic dimer's chain A was used. Four Na^+ cations were added to neutralize the system (with residue ionization states set appropriate to a pH of 7), far away from the protein and from each other. A 20 ps fixed number, pressure, and temperature (*NPT*) simulation at $T = 303$ was run with a 1 fs time step to solvate the protein and adjust to the normal density. Then, 5 ns of fixed *NVT* was run in a box of volume 70.7 × 73.3 × 73.2 Å^3 with 11105 water molecules at $T = 303$. The simulations are done with only hydrogen-containing bonds constrained by SHAKE,[44] a 2 fs time step, and a temperature coupling time of 0.5 ps.[45] The first 1 ns of data was discarded as an equilibration period.

The various PCA methods were carried out with ANALYZER,[46] a program written for the purpose of analyzing trajectory data by a wide variety of methods. Before diagonalization of the covariance matrix, the overall translational and rotational motion of the protein was removed by superimposing all the trajectory snapshots onto the core CA atoms (see Results for the definition of the core) of the X-ray structure. The varimax rotation appropriate to atomistic trajectory data was coded on the basis of the source code of a generic statistical package.[47]

3. Results

3.1. Principal Component Analysis of AKE. The trajectory snapshots of the apo-AKE fluctuations were best fit on the protein core atoms. The core is defined as residues 1−29, 60−121, and 160−214 leaving the AMPbd domain as residues 30−59 and the LID domain as residues 122−159.[43] When PCA is carried out using the CA atoms, there are 642 modes (2568 when using backbone atoms) with 6 zero eigenvalue modes arising from the elimination of the overall translation and rotation of the protein. The PCA mode contributions to the mean square fluctuation (MSF) are listed in Table 1 for the CA atoms and, clearly, there is strong dominance by the first few modes. Essentially the same data are obtained by PCA on the backbone atoms, and this shows that collective motion is being extracted by the PCA analysis. Indeed, using half of the CAs or even a quarter of the CAs (evenly spaced) did not change the mode

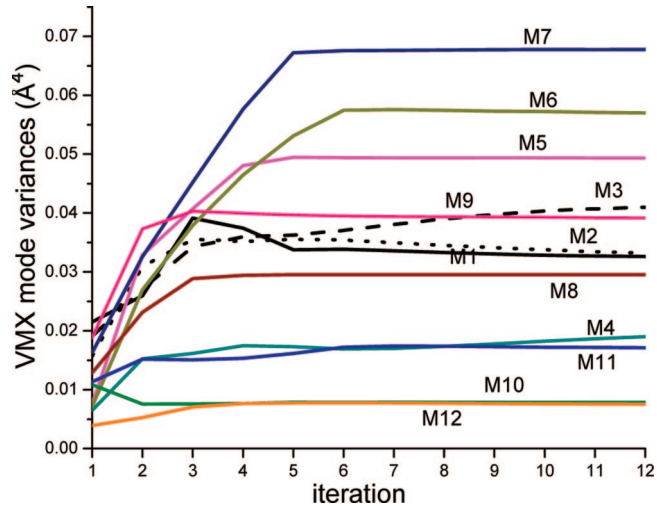

Figure 1. Individual variances σ_ν^2 for modes 1−12 as a function of iteration number in the varimax rotation. The σ_ν^2 do not necessarily have to maximize as the varimax algorithm maximizes the sum variance VAR (see eq 2.7) over all the modes used.

percentages of the dominant modes. The PCA of the trajectory summarized in Table 1 shows that the first 12 modes account for ∼88% of the total MSF of 4AKE. These modes are good candidates for the varimax rotation to see if the rotated modes do concentrate the atom contributions.

3.2. Varimax Rotation. The goal of a varimax rotation is to concentrate the atom contributions to a given mode into a smaller set of atoms than in the original PCA mode and in doing so provide a more readily interpretable set of modes. The variance σ_ν^2 defined in eq 2.7 monitors this concentration. It would be zero for all atoms contributing equally and increases in value (related to the number of degrees of freedom, 3*N*) when a smaller number of atoms have significant amplitude in the mode. The convergence characteristics of the first 12 modes are displayed in Figure 1, which plots the σ_ν^2 of the mode vectors. The increase of variance depends quite strongly on the mode, with some hardly changing and others increasing significantly. As noted in Methods, the total variance VAR is maximized; thus all terms need not be maximized. Other numbers of modes were tried for the varimax rotation, and little difference with the 12-mode result was found. The varimax rotation by this measure does increase the concentration of atom contributions to the new mode vectors. This feature becomes explicit by comparing the RC_ν^j values that measure the contributions of atom j to mode ν between the PCA and VMX modes, as shown in Figure 2 for modes 1 and 5. The RC_ν^j definition in eq 2.5 incorporates the effect of the eigenvalue and permits comparison not only of the PCA versus VMX results of a particular mode but also permits comparisons among different modes. (In this regard, it should be noted that it might be better to consider the squares of the RC_ν^js when making comparisons because the PCA partitions the MSF into additive contributions.) Both modes show an increase in concentration, with the increase for mode 5 being quite dramatic.

The PCA-based RC_ν^j values in Figure 2 do show the primary importance of the LID (residues 122−159) motion, dominating mode 1, and secondarily the AMPbd domain (residues 30−59), dominating mode 5; however, the RC_ν^j values are still quite spread out over residues. The varimax-rotated vectors concentrate the residues that contribute to RC_ν^j values and are able to identify a small set as mainly contributing to the AKE fluctuations. Figure 3 displays the 12 VMX RC_ν^j, where about 10 CA atoms (indicated in Table 2) can be readily identified

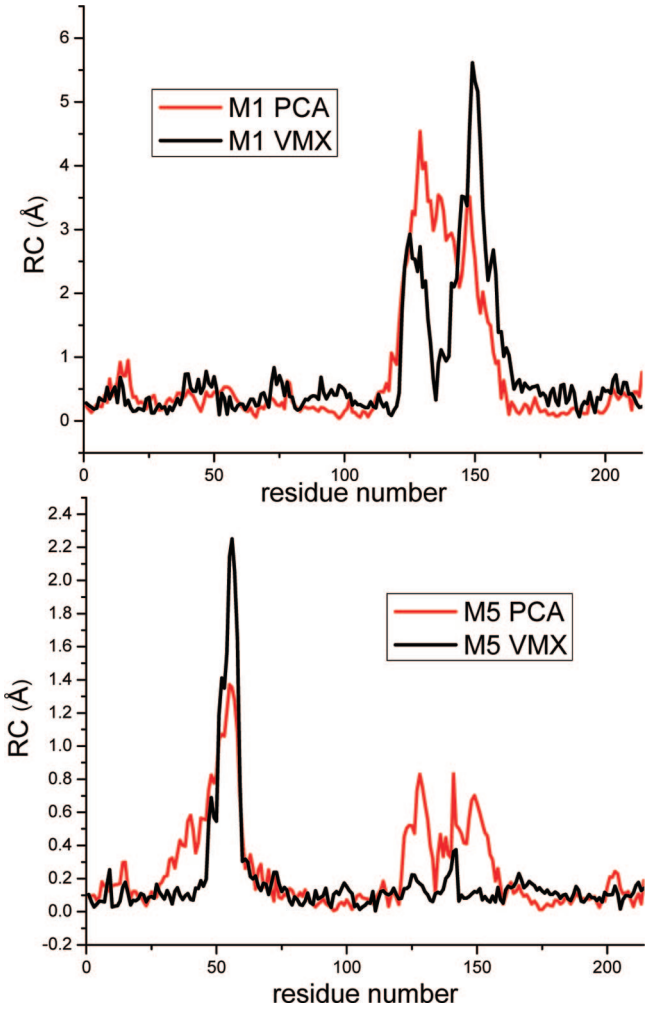

Figure 2. Comparison of the *R*-component (RC_ν^j) values for the PCA and VMX modes 1 and 5. A VMX (varimax) mode tends to concentrate the atom contributions to a given mode better than does the corresponding PCA mode.

from the plots as being key contributors. Of course, the choice of these atoms is somewhat subjective, but they have been picked as peak positions and, again, since it is the squares of the RC_ν^js that are most relevant, the peaks become even more prominent with use of this measure. These residues can be grouped according to their contributions to the various modes, as listed in Table 2. Most of the AKE fluctuations are associated with the LID region. The varimax rotation effectively partitions the modes into AMPbd modes (3, 5, and 8), and the other modes describe LID motion. Figure 4 shows that the LID residues featured by the VMX modes are distributed throughout the LID, while Gly56 is picked out in the AMPbd domain. It is true that mode 8 (see Figure 3) is quite broad, but the weight of this mode is really quite small.

3.3. LID and AMPbd Domain Motions. Underlying motions that would be obscured by the trajectory with all its complicated oscillations can be revealed by using the atom displacements corresponding to sums over selected modes. It is useful to use the varimax modes since they isolate the atom motions more purely. If one mode were used in eq 2.2, each atom motion $\mathbf{R}^j(t)$ is confined to a line (in 3-space). If two modes are used, $\mathbf{R}^j(t)$ is confined to a plane in 3-space for any atom's motion. Thus, as shown in Methods, by using two modes, important atom rotations can be ferreted out by finding pairs of residues that have large RC_ν^j and RC_μ^j values as well as small

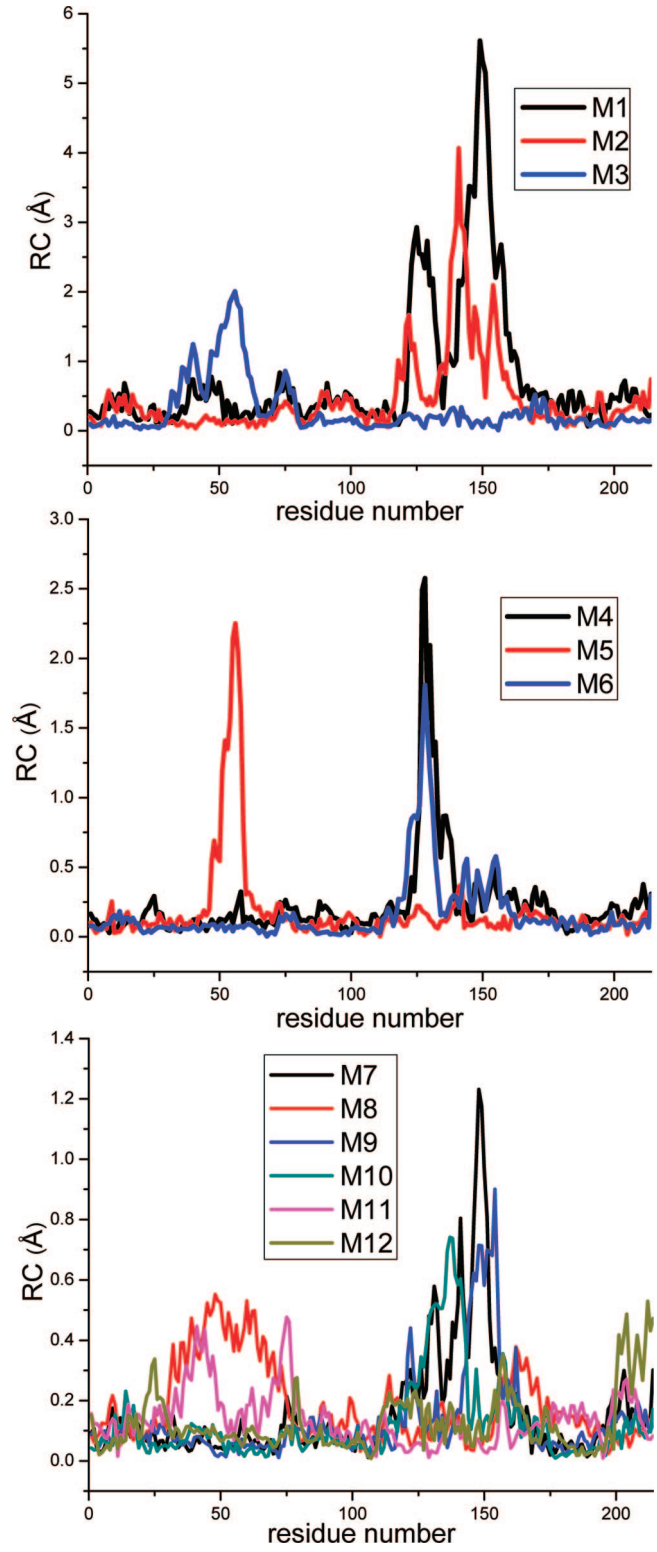

Figure 3. The *R*-components (RC_ν^j) for the 12 VMX modes. The atoms with prominent RC_ν^j values are collected in Table 2.

$\cos(\Theta_{\nu\mu}^j)$ values. Simultaneously satisfying these conditions will point to regions of the protein where there are large coherent fluctuations for two modes that have mode atom vectors that are close to orthogonal. Satisfaction of these properties could reveal in the extreme case a circular trajectory.

These ideas are illustrated by VMX modes 1 and 4 where atom 127 in the LID domain is prominent (see Figure 3 and Table 2). A plot of the magnitude of $\cos(\Theta_{\nu\mu}^j)$ is given in Figure 5. There are a number of small values of $\cos(\Theta_{\nu\mu}^j)$, but, to be

TABLE 2: Atoms That Contribute to the Main VMX Modes and Their Segregation into Groups

mode	residues	group	residues/modes
1	125, 129, 149	1	128,129/1, 4, 6
2	141	2	56/3, 5, 8
3	56	3	148, 149/1, 7, 9
4	128	4	141/2, 7
5	56	5	125/1
6	128	6	154/ 9
7	141, 148	7	137/10
8	56 (broad)		
9	148, 154		
10	137		

of interest, a small value must be associated with correspondingly large RC_ν^j and RC_μ^j values, as is the case for residue 127.

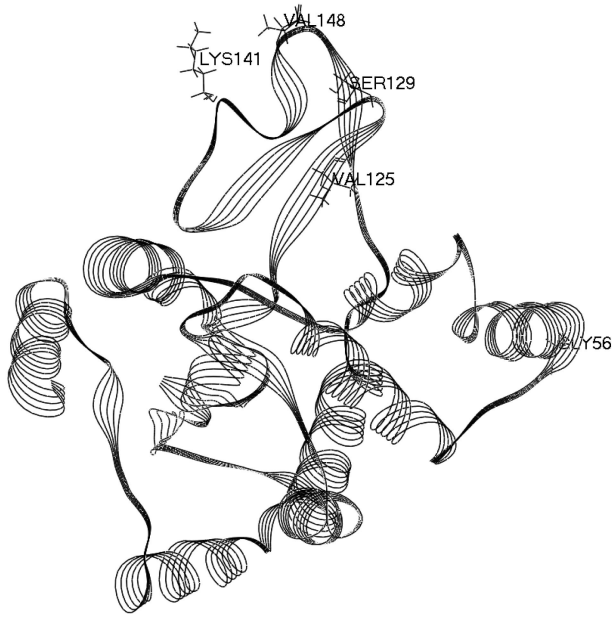

Figure 4. Apo form of AKE from the crystal structure with some residues identified from the VMX rotation of the first 12 PCA modes. The VMX results concentrate on a few residues in the LID (residues 125, 129, 141, and 148) and one (residue 56) in the AMPbd domain.

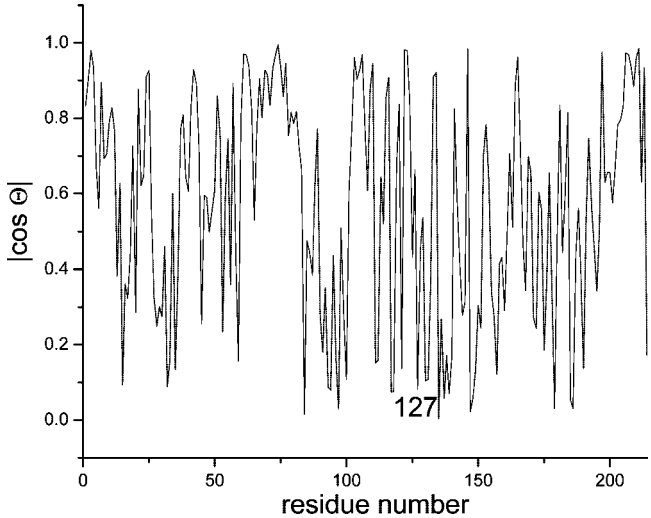

Figure 5. Magnitude of cos $\Theta_{\nu\mu}^j$ of modes 1 and 4 for the residues (CA atoms). A small value indicates that the component vectors (3-vectors) $\mathbf{m}_\nu^j$ and $\mathbf{m}_\mu^j$ are almost orthogonal.

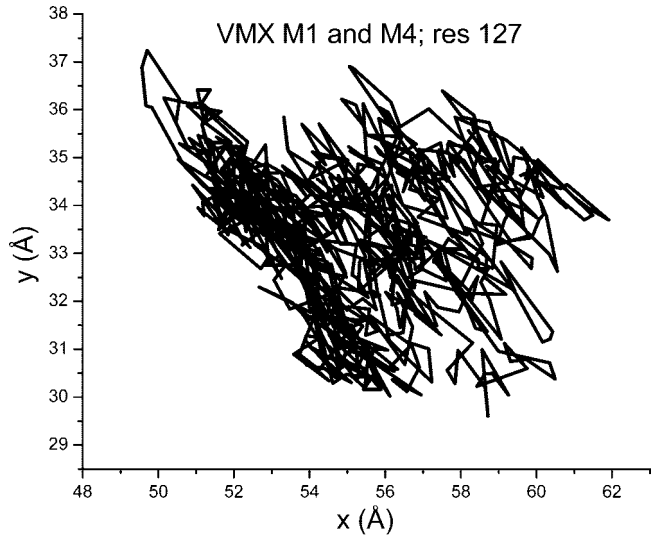

Figure 6. Trajectory for residue 127 projected onto VMX modes 1 and 4 exhibiting rotational motion with origin in the large RC_ν^k and RC_μ^k and correspondingly small cos ($\Theta_{\nu\mu}^k$) values.

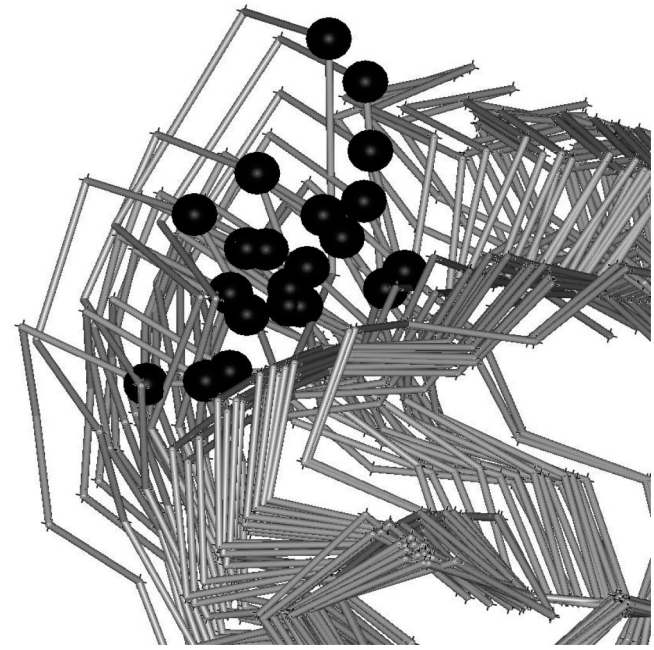

Figure 7. Part of the LID domain. The large black spheres are res 127 CA positions for the VMX modes 1 and 4 based trajectory snapshots. The curvature is quite clear here.

The resulting two-mode trajectory for residue 127, displayed in Figure 6 as a projection into a 2D plane for clarity, shows that the motion is indeed rather circular. To visualize this motion in the context of the total protein, Figure 7 displays some snapshots from the protein trajectory based on these two modes, for part of the LID domain. The black spheres are the residue 127 CA atom locations. The motion in a plane is rather clear from this view and corresponds to a twisting of the LID domain.

The other prominent feature of the VMX analysis is the AMPbd domain where residue 56 is picked out in modes 3 and 5, as indicated in Table 2. The motion of residue 56 is illustrated in Figure 8 in a two-dimensional projection for clarity, and it again provides a picture of more-or-less circular motion. It is interesting to contrast the PCA and VMX values of cos($\Theta_{\nu\mu}^j$) for these modes, as shown in Figure 9. The graph is truncated to feature the small (absolute) values of cos($\Theta_{\nu\mu}^j$) and shows

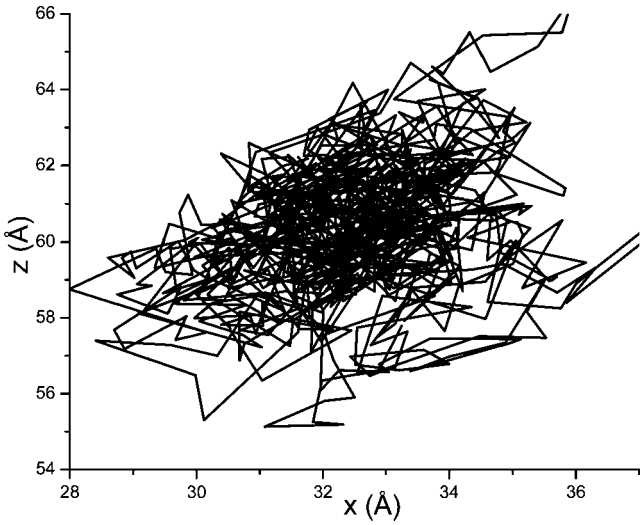

Figure 8. Trajectory for the CA atom of residue 56 projected onto VMX modes 3 and 5 exhibiting rotational motion with origin in the large RC_ν^k and RC_μ^k and correspondingly small cos $(\Theta_{\nu\mu}^k)$ values.

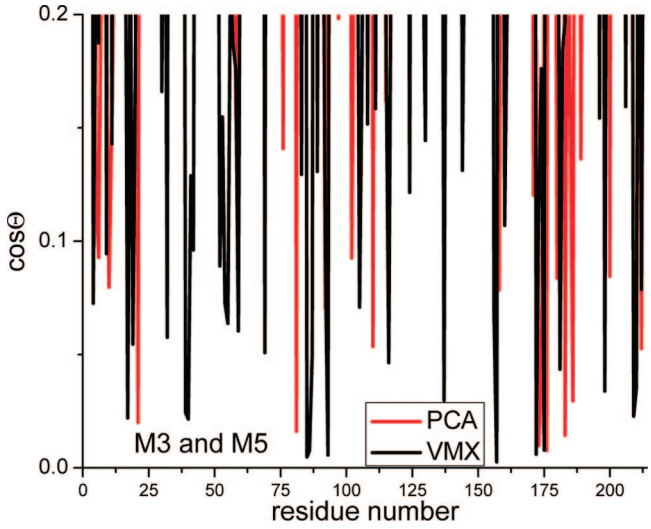

Figure 9. Plot of cos $(\Theta_{\nu\mu}^j)$ (the absolute values) for the indicated modes contrasting the PCA and VMX results, indicating that the VMX M3 and M5 modes are on average closer to orthogonal than are the corresponding PCA modes.

that here there are substantially more small VMX than PCA cos$(\Theta_{\nu\mu}^j)$ values. However, it should be stressed that these small values must be associated with correspondingly large RC_ν^j values to be of interest. The protein motion corresponding to modes 3 and 5 is displayed for some trajectory snapshots in Figure 10. Because the picture is constructed from two modes, some care must be exercised in concluding that the actual motion is essentially two-dimensional. To do so, note that modes 3 and 5 dominate the AMPbd motion, as shown in Figure 3 (mode 8 is there, too, with its diffuse character, but the λ_ν^2 for mode 8 is down by a factor of almost 3 relative to mode 5). Thus, modes 3 and 5 dominate the AMPbd motion, and, while the motion must be in a plane because of the two-mode projection, this will be the actual dominant motion. The orientation of Figure 10 was chosen as a side-on view of residue 56 to illustrate that the AMPbd domain oscillates in the direction of the LID domain.

3.4. Apo Fluctuations Encode Holo-like Conformations. PCA on the MD trajectories of numerous proteins does show that their fluctuations consist of numerous small, one-basin oscillatory modes and a small number of large, often nonharmonic modes. If this separation holds for an apo protein and the amplitudes of all these modes were artificially increased, then it would be a remote possibility that holo conformations would result. On the other hand, if it is true that holo conformations are encoded in the apo fluctuations, but they are not seen in the MD trajectory because the trajectory is not run for a sufficiently long time, then it may be that some set of the large "productive" modes from the apo PCA will point toward holo conformations. If these properties are manifest in the trajectory, then (1) a small subspace contained in the covariance matrix **C** characterizes the important apo motions and (2) that subspace is the relevant one for holo-directed motion.

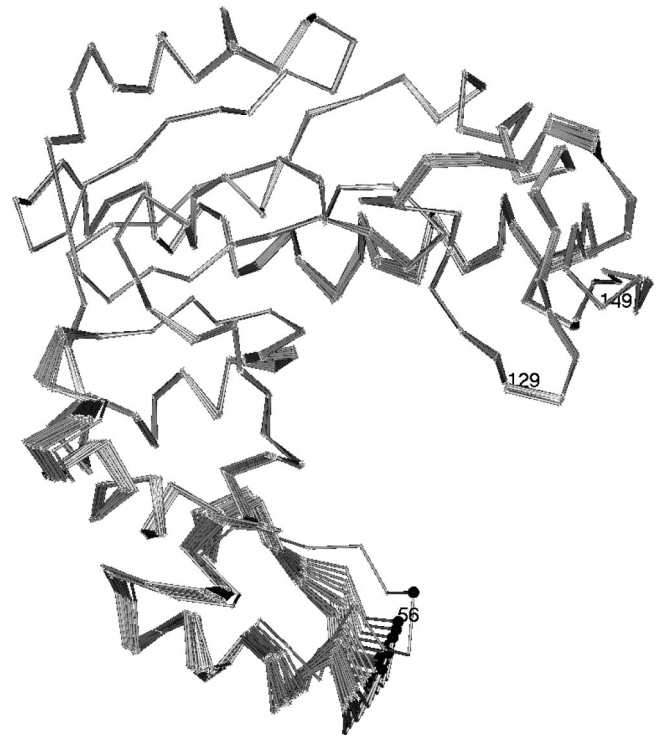

Figure 10. Trajectory of residue 56 from modes 3 and 5, denoted with black spheres, in an orientation that shows that it is mainly confined to a plane and oscillates toward the LID (indicated by the labeled residues, 129 and 149). Modes 3 and 5 dominate residue 56, as evident in Figure 3. When the figure is reoriented, the motion is approximately circular, as shown in Figure 8.

To explore this possibility, we express the difference $\Delta\mathbf{X}$ in holo and apo X-ray conformations in the apo protein PCA basis:

$$\Delta\mathbf{X} = \mathbf{R}^{\text{holo}} - \mathbf{R}^{\text{apo}} = \sum_{\nu}^{3N} (\Delta\mathbf{X}\cdot\mathbf{m}_\nu^T)\mathbf{m}_\nu \equiv \sum_{\nu}^{3N} a_\nu \mathbf{m}_\nu \tag{3.1}$$

In this fashion, the holo conformation is expressed in the apo PCA basis. For the purposes of this analysis, there is no need to use the varimax vectors because the following results are independent of whether PCA or VMX vectors are used, as shown in Appendix A. If the apo fluctuation directions that are expressed by the $\mathbf{m}_\nu$ basis do encode the holo direction, then the a_ν expansion coefficients in eq 3.1 should track the λ_ν coefficients of the apo PCA. That is, the a_ν should fall off roughly in proportion to the fall off in the λ_ν. Equivalently, if this encoding hypothesis is true, then

$$\Delta\mathbf{X} - \Delta\mathbf{X}^{M_\nu} = \sum_{\nu>M_\nu}^{3N} a_\nu \mathbf{m}_\nu \tag{3.2}$$

should also decay rapidly with ν (again, on the same scale as the λ_ν). A convenient (normalized) measure to use is

$$\begin{aligned}\mathrm{DEV} &= (\Delta\mathbf{X} - \Delta\mathbf{X}^{M_\nu})^T\cdot(\Delta\mathbf{X} - \Delta\mathbf{X}^{M_\nu})/(\Delta\mathbf{X})^T\cdot(\Delta\mathbf{X}) \\ &= (\Delta\mathbf{X} - \Delta\mathbf{X}^{M_\nu})^T\cdot(\Delta\mathbf{X} - \Delta\mathbf{X}^{M_\nu})/\mathrm{MSD}\end{aligned} \tag{3.3}$$

In DEV, the denominator, MSD, is the mean square deviation between the holo and apo forms (the rmsd = $(\mathrm{MSD})^{1/2}$ between then is approximately 100 Å). The scaling of DEV with λ_ν^2 is reasonably obeyed, as shown in Figure 11. The PCA squared eigenvalues are also plotted in normalized fashion ($\lambda_\nu^2/\lambda_0^2$) to put them on the same scale as DEV. One should not expect that DEV could be zero for $M_\nu < 3N$ modes. If a snapshot at time t_s is taken from the apo trajectory, the instantaneous value of $(\Delta\mathbf{X}(t_s) - \Delta\mathbf{X}^{M_\nu}(t_s))^T(\Delta\mathbf{X}(t_s) - \Delta\mathbf{X}^{M_\nu}(t_s)) = (\delta\mathbf{R}(t_s) - \delta\mathbf{R}^{M_\nu}(t_s))^T(\delta\mathbf{R}(t_s) - \delta\mathbf{R}^{M_\nu}(t_s))$, where $\Delta\mathbf{X}(t_s) = \mathbf{R}^{\mathrm{holo}} - \delta\mathbf{R}(t_s)$, will be a fluctuating quantity. Its time-averaged value is given by

$$\langle(\Delta\mathbf{X}(t) - \Delta\mathbf{X}^{M_\nu}(t))^T\cdot(\Delta\mathbf{X}(t) - \Delta\mathbf{X}^{M_\nu}(t))\rangle = \sum_{\nu>M_\nu}^{3N} \lambda_\nu^{\,2} \tag{3.4}$$

as shown in Appendix A. It provides an estimate of how close to zero a quantity such as $\Delta\mathbf{X} = \mathbf{R}^{\mathrm{holo}} - \mathbf{R}^{\mathrm{apo}}$ can actually be as a function of the value of M_ν. To the extent that M_ν modes do characterize the apo fluctuations, the complement in eq 3.4 provides the limitation on how well this number of modes can be expected to represent the holo configuration. For example, if the first 12 modes occupy 90% of the total MSF, then the complement in eq 3.4 will correspond to 10% and the DEV in Figure 11 could not be smaller than 0.1.

Also shown in Figure 11 is the deviation defined in eq 3.3 that would result if Cartesian basis vectors $\mathbf{e}_\nu$ with elements $e_\nu^j = \delta_{\nu j}$ were used, instead of the PCA $\mathbf{m}_\nu$ basis vectors. In that case, the deviation is simply given by $\mathrm{DEV} = \sum_{\nu>M_\nu}^{3N} (\Delta X_\nu)^2$. Clearly, in this basis, DEV is not tracking the λ_ν^2 decay behavior. (The "breaks" in the behavior of the Cartesian DEV correspond to the peaks in the residue MSF values.) Thus, the PCA modes of the apo trajectory are quite effective in "pointing" in the direction of the holo-derived protein crystal structure.

3.5. Visualization of the Holo-Directed Apo Fluctuations. The $\Delta\mathbf{X}^{M_\nu}$ vector can be used to obtain a picture of the configuration of AKE implied by this number of vectors according to $\mathbf{R}^{M_\nu} = \mathbf{R}^{\mathrm{apo}} + \Delta\mathbf{X}^{M_\nu}$. Figure 12 plots the PCA predicted configurations for $M_\nu = 12$ and 24 (blue), along with the apo (red) and holo (black) crystal structure configurations, for the backbone atoms in a ribbon representation. These are what the apo-derived fluctuations, when analyzed by PCA (on the CA atoms) extrapolate to, for a given number of PCA modes, when the target is the holo crystal structure. It is evident that even for 12 modes the extrapolation already is quite holo-like. The "break" in the eigenvalue curve (see Figure 11) occurs between 12 and 24 modes, and therefore the predicted configurations for $M_\nu = 12$ and 24 should be quite similar, with the latter somewhat closer to the holo configuration. The displacement of the LID that consists of bending down and twisting toward the AMPbd domain is quite dramatic. The movement of the AMPbd site is also substantial though it is clear that it has not moved as far toward the holo-based conformation as has the LID. The remaining ~600 modes do little to improve on this result.

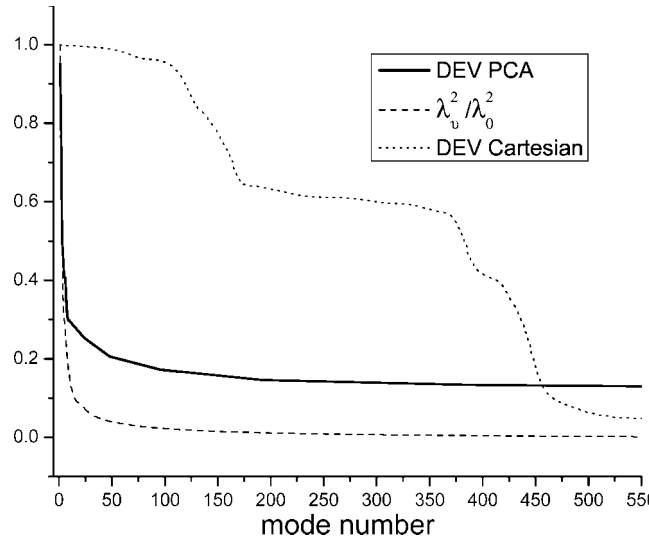

Figure 11. DEV (see eq 3.3) that monitors the approach toward the holo structure as a function of the number of modes used in the expansion of the difference of holo and apo crystal structures in the apo PCA basis, along with the decay of the apo fluctuation eigenvalues, λ_ν^2. That their decays with the number of modes are similar shows that the apo fluctuations do encode hololike protein conformations. Also shown for comparison is the DEV that would result if a Cartesian basis were used for the expansion, instead of the PCA basis.

Table 3 presents some key distances between residue CA atoms that serve as monitors of the conformational change. The LID to AMPbd distances (56−129 and 56−149) indicate the similarity with the holo configuration; though distances of course are not designed for picking out orientational changes. Also included in Table 3 are the CA distances between residues 56 and 169. Residue 169 is below the LID, and it was used in our previous[11] distance replica exchange method simulations to obtain the potential of mean force for the 56−169 distance reaction coordinate. This reaction coordinate mainly reports on the AMPbd motion because residue 169 is in the core of the protein and therefore quite stable. Figure 12 shows also that the PCA extrapolation does not move the AMPbd domain as much as the LID domain. The proposed[2] mechanism of substrate binding to AKE is for AMP to bind first to the AMPbd domain, which induces some motion of the AMPbd domain, followed by ATP binding that produces the final, holo structure.

4. Concluding Remarks

In this work, we introduced the varimax rotation as an aid in simplifying the results of PCA on an atomistic trajectory. Using the method on the AKE trajectory does succeed in increasing the concentration of the atom contributions to most of the modes in comparison with the original PCA modes, as evident from Figures 1 and 2. The structure for the first 12 VMX modes displayed in Figure 3 is indeed quite simple and points to a small set of residues as key for the important LID and AMPbd domain motions. Two mode representations pointed to by the varimax method show (see Figure 7) that the LID domain closes toward the AMPbd domain with a twisting motion comprised of rotation and translation. The AMPbd domain motion (Figure 10) is dominated by just two modes, and this motion represents mainly an oscillation in the direction of the LID domain.

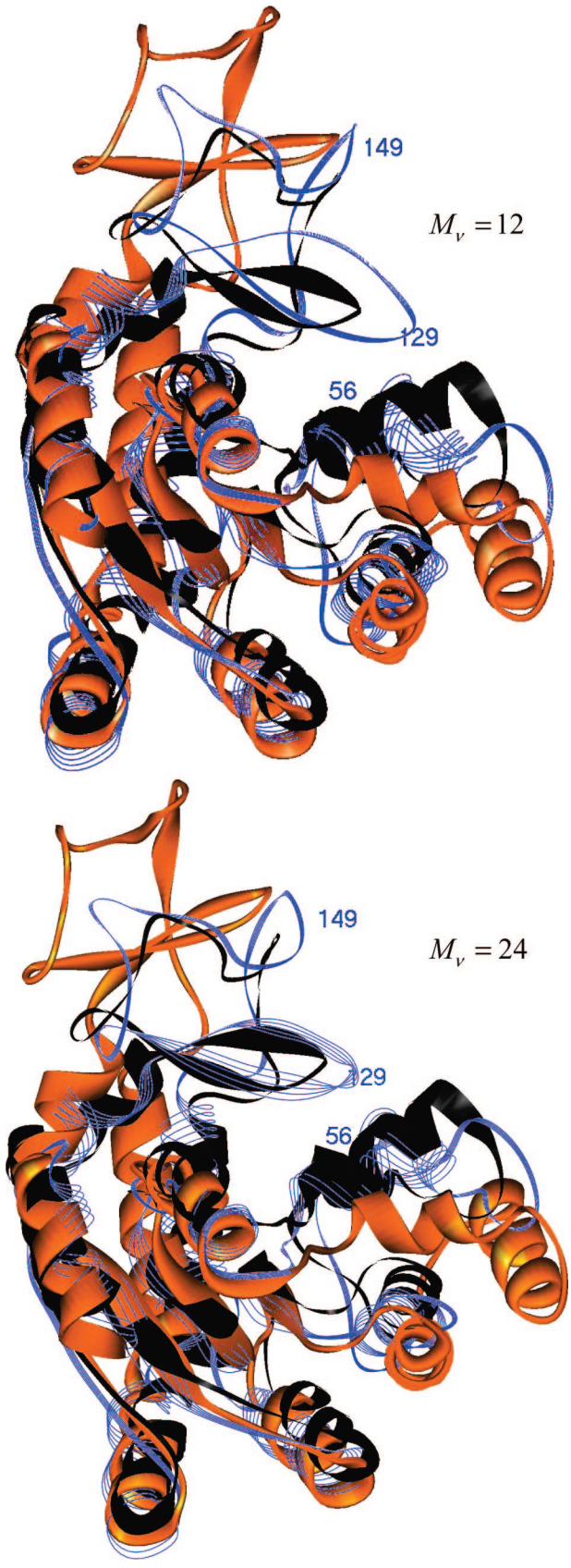

Figure 12. Apo crystal backbone (red ribbon); holo crystal backbone (black ribbon); 12-mode (top panel) and 24-mode (bottom panel) PCA extrapolated structure (blue line ribbon). The downward movement and rotation of the LID domain and the corresponding upward motion of the AMPbd domain toward the holo crystal structure dominates the PCA-determined conformational change.

TABLE 3: distances between Selected Residues for the Apo, PCA-Extrapolated, and Holo Configurations

	$d(56-129)$,[a] Å	$d(56-149)$,[a] Å	$d(56-169)$,[a] Å
apo	30.4	39.2	26.9
PCA-extrapolated[b]	22.4	30.3	17.0/15.0[c]
holo	22.0	29.5	10.0

[a] Distance between CA atoms of indicated residues. [b] From the $M_\nu = 12$ data in Figure 12. [c] From the $M_\nu = 24$ data in Figure 12.

It should be noted that LID and AMPbd domain motions have been predicted by various ENM-based coarse-grained methods, some of which use solely the apo structure[10] and others that interpolate between apo and ternary complex forms.[8,9,11] The varimax analysis of the MD apo trajectory refines the description of the global motion found in all these methods by concentrating them to as few atoms as possible under the varimax criterion.

Two properties of the PCA method are orthogonality of the $\mathbf{m}_\nu$, the PCA vectors, and the uncorrelatedness of the PCA principal components, the $p_\nu(t)$ (eq A.1). Any (orthogonal) rotation of the coordinate system to new axes must leave the resulting rotated $\tilde{m}_\nu$ vectors orthogonal, but the corresponding varimax components $\tilde{p}_\nu(t) = \tilde{m}_\nu^T \cdot \Delta\mathbf{R}(t)$ become correlated. The eigenvalue ordering property no longer holds for the VMX modes. That is, if some fraction $\sum_{\nu=1}^{M_\nu} \lambda_\nu^2 / \sum_{\nu=1}^{3N} \lambda_\nu^2$ of the total RMSF is represented by the first M_ν PCA modes, the same fraction is captured by the VMX modes, but the individual modes no longer are ordered. But, if the varimax rotation succeeds in concentrating atom contributions in the various modes, then a simplified physical description can be achieved. For example, physical atom rotations will be harder to identify when there are more atoms with significant weights in a given mode. Thus, giving up the property of uncorrelatedness can be a useful tradeoff in exchange for concentrating the atom contributions to the vectors. While PCA methods typically focus on collective motions, ultimately the atom contributions are of greatest interest. Other rotations are possible. Indeed, rotations that give up the orthogonality of the $\mathbf{m}_\nu$ are possible while preserving uncorrelatedness of the $p_\nu(t)$, or both can be given up.[32,33] But, the varimax rotation seems best suited to the dissection and simplification of protein motions.

In this regard, it is important to distinguish between correlation and statistical independence. Correlation is defined with respect to the centered second moment matrix of the atom fluctuations, $\mathbf{C}$, as is clear from eqs 2.1 and A.1, and this property is different from statistical independence that would require the joint probability distribution of the modes to factorize. Only if the joint probability distribution of the modes were Gaussian would correlation and statistical independence be equivalent. Since the principal components of the large eigenvalue PCA modes of a protein are typically characterized by non-Gaussian distributions (almost by definition interesting modes will not be Gaussian—if they were most likely they would be describing more-or-less random fluctuations as would be implied by central limit theorem arguments) they will be uncorrelated, by PCA construction, but not statistically independent. In this sense, the uncorrelatedness of PCA modes is not such a useful property to maintain. The varimax properties of orthogonal vectors and maximizing the atom contributions to a set of important modes may be more useful.

Proteins that undergo large-scale rearrangements between their apo and holo forms because of, for example, a requirement that their substrate(s) need to be sequestered from the solvent are good candidates for obeying the preexisting equilibrium (conformational selection) paradigm. From the perspective of

simulation where, typically, the trajectory is confined to some limited range around the initial state obtained from a crystal or NMR determined structure, it is difficult to explore the encoding hypothesis. The approach pursued here expresses the difference between apo and holo configurations of the protein on the basis of the PCA modes of the apo trajectory. The analysis can be done with the PCA or the VMX modes, for a specified number of included modes, since they capture the same fraction of the overall protein RMSF. The DEV variable defined in eq 3.3 with behavior displayed in Figure 11 does show that the apo fluctuations of AKE encode holo-like conformations. The comparisons of apo, holo, and projected on a number of modes configurations shown in Figure 12 illustrate that the projected conformations are much more holo than apo-like.

The PCA decomposition of trajectory fluctuations is useful in the context of the encoding hypothesis in the following sense. An MD trajectory starting from an apo crystal structure typically will not be able to sample productive, holo-like conformations because the simulation cannot be run for a sufficiently long time. Suppose that a PCA analysis of this apo trajectory does lead to a separation into a few large and many small eigenvalues. To obtain possible protein motions based on these PCA modes (which are directions of motion) amplitudes for the modes must be assigned. If finite amplitudes for all the modes were assigned, then it would be a remote possibility that holo-like conformations would result. The small eigenvalue modes can be viewed as "noise" and should not be involved in large-scale conformational changes. If, as found in this work, holo conformations are encoded in the apo fluctuations, then some set of the large "productive" modes that are obtained from the apo PCA will point toward holo conformations.

Appendix A

The uncorrelated nature of the PCA principal components $p_\nu(t)$ is readily demonstrated:

$$\langle \delta p_\nu(t)\, \delta p_\mu(t)\rangle = \frac{1}{T}\int (\mathbf{m}_\nu^T \cdot \delta\mathbf{R}(t))(\delta\mathbf{R}^T(t)\cdot\mathbf{m}_\mu)\, \mathrm{d}t = \mathbf{m}_\nu^T\cdot\mathbf{C}\cdot\mathbf{m}_\mu = \lambda_\nu{}^2\mathbf{m}_\nu^T\cdot\mathbf{m}_\mu = \lambda_\nu{}^2\delta_{\mu\nu} \quad \text{(A.1)}$$

with $\delta p_\nu(t) = p_\nu(t) - \langle p_\nu(t)\rangle$ and $\delta_{\mu\nu}$ the Kronecker delta.

The effect of $M_\nu < 3N$ modes on the DEV introduced in eq 3.3 can be analyzed as follows. Define

$$\Delta\mathbf{X}(t) = \mathbf{R}^{\mathrm{holo}} - \delta\mathbf{R}(t) = \mathbf{R}^{\mathrm{holo}} - \sum_{\nu=1}^{3N}\delta p_\nu(t)\mathbf{m}_\nu$$

$$\Delta\mathbf{X}^{M_\nu}(t) = \mathbf{R}^{\mathrm{holo}} - \delta\mathbf{R}^{M_\nu}(t) = \mathbf{R}^{\mathrm{holo}} - \sum_{\nu=1}^{M_\nu}\delta p_\nu(t)\mathbf{m}_\nu \quad \text{(A.2)}$$

Then, using eq A.1 and the relation $\delta\mathbf{R}(t) - \delta\mathbf{R}^{M_\nu}(t) = -(\Delta\mathbf{X}(t) - \Delta\mathbf{X}^{M_\nu}(t))$

$$\langle(\Delta\mathbf{X}(t) - \Delta\mathbf{X}^{M_\nu}(t))^T\cdot(\Delta\mathbf{X}(t) - \Delta\mathbf{X}^{M_\nu}(t))\rangle = \sum_{\nu>M_\nu}^{3N}\sum_{\nu'>M_\nu}^{3N}\frac{1}{T}\int_0^T \delta p_\nu(t)\, \delta p_{\nu'}(t)\, \mathrm{d}t\mathbf{m}_\nu^T\cdot\mathbf{m}_{\nu'} = \sum_{\nu>M_\nu}^{3N}\frac{1}{T}\int_0^T \delta p_\nu{}^2(t)\, \mathrm{d}t = \sum_{\nu>M_\nu}^{3N}\lambda_\nu{}^2 \quad \text{(A.3)}$$

$\Delta\mathbf{X}^{M_\nu}(t)$ is invariant to the use of PCA or VMX vectors for the same number of vectors as is readily shown:

$$\sum_{\nu}^{M_\nu}(\Delta\mathbf{X}(t)\cdot\tilde{\mathbf{m}}_\nu^T)\tilde{\mathbf{m}}_\nu = \sum_{\nu}^{M_\nu}\sum_{\mu}^{M_\nu}\sum_{\mu'}^{M_\nu}(\Delta\mathbf{X}(t)\cdot T_{\mu\nu}\mathbf{m}_\nu^T)T_{\mu',\nu}\mathbf{m}_\nu^T = \sum_{\nu}^{M_\nu}(\Delta\mathbf{X}(t)\cdot\mathbf{m}_\nu^T)\mathbf{m}_\nu = \Delta\mathbf{X}^{M_\nu}(t) \quad \text{(A.4)}$$

where we used the orthonormality of the $T_{\mu\nu}$ transformation between PCA and varimax vectors defined in eq 2.6.

References and Notes

(1) Walsh, C., *Enzymatic reaction mechanisms*; Freeman, W. H.: San Francisco, 1979.
(2) Schulz, G. E.; Muller, C. W.; Diederichs, K. *J. Mol. Biol.* **1990**, *213*, 627.
(3) Sheng, X. R.; Li, X.; Pan, X. M. *J. Biol. Chem.* **1999**, *274*, 22238.
(4) Han, Y.; Li, X.; Pan, X. M. *FEBS Lett.* **2002**, *528*, 161.
(5) Zhang, H. J.; Sheng, X. R.; Niu, W. D.; Pan, X. M.; Zhou, J. M. *J. Biol. Chem.* **1998**, *273*, 7448.
(6) Burlacu-Miron, S.; Gilles, A. M.; Popescu, A.; Bârzu, O.; Craescu, C. T. *Eur. J. Biochem.* **1999**, *264*, 765.
(7) Sinev, M. A.; Sineva, E. V.; Ittah, V.; Haas, E. *Biochemistry* **1996**, *35*, 6425.
(8) Maragakis, P.; Karplus, M. *J. Mol. Biol.* **2005**, *352*, 807.
(9) Miyashita, O.; Onuchic, J. N.; Wolynes, P. G. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 12570.
(10) Temiz, N. A.; Meirovitch, E.; Bahar, I. *Proteins: Struct., Funct., Bioinf.* **2004**, *57*, 468.
(11) Lou, H. F.; Cukier, R. I. *J. Phys. Chem. B* **2006**, *110*, 24121.
(12) Hayward, S.; Lee, R. A. *J. Mol. Graphics Modell.* **2002**, *21*, 181.
(13) Lou, H. F.; Cukier, R. I. *J. Phys. Chem. B* **2006**, *110*, 12796.
(14) Bahar, I.; Chennubhotla, C.; Tobi, D. *Curr. Opin. Struct. Biol.* **2007**, *17*, 633.
(15) Goh, C. S.; Milburn, D.; Gerstein, M. *Curr. Opin. Struct. Biol.* **2004**, *14*, 104.
(16) James, L. C.; Tawfik, D. S. *Trends Biochem. Sci.* **2003**, *28*, 361.
(17) Kumar, S.; Ma, B. Y.; Tsai, C. J.; Sinha, N.; Nussinov, R. *Protein Sci.* **2000**, *9*, 10.
(18) Ma, B. Y.; Kumar, S.; Tsai, C. J.; Nussinov, R. *Protein Eng.* **1999**, *12*, 713.
(19) Bryngelson, J. D.; Onuchic, J. N.; Socci, N. D.; Wolynes, P. G. *Proteins: Struct., Funct., Genet.* **1995**, *21*, 167.
(20) Dill, K. A.; Chan, H. S. *Nat. Struct. Biol.* **1997**, *4*, 10.
(21) Kamikubo, H.; Shimizu, N.; Harigai, M.; Yamazaki, Y.; Imamoto, Y.; Kataoka, M. *Biophys. J.* **2007**, *92*, 3633.
(22) Liu, T.; Whitten, S. T.; Hilser, V. J. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 4347.
(23) Cui, Q.; Li, G. H.; Ma, J. P.; Karplus, M. *J. Mol. Biol.* **2004**, *340*, 345.
(24) Krebs, W. G.; Alexandrov, V.; Wilson, C. A.; Echols, N.; Yu, H. Y.; Gerstein, M. *Proteins: Struct., Funct., Genet.* **2002**, *48*, 682.
(25) Nicolay, S.; Sanejouand, Y. H. *Phys. Rev. Lett.* **2006**, 96.
(26) Tama, F.; Valle, M.; Frank, J.; Brooks, C. L. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 9319.
(27) Cavasotto, C. N.; Kovacs, J. A.; Abagyan, R. A. *J. Am. Chem. Soc.* **2005**, *127*, 9632.
(28) Cavasotto, C. N.; Orry, A. J. W. *Curr. Top. Med. Chem.* **2007**, *7*, 1006.
(29) Ikeguchi, M.; Ueno, J.; Sato, M.; Kidera, A. *Phys. Rev. Lett.* **2005**, 94.
(30) Lange, O. F.; Grubmuller, H. *J. Phys. Chem. B* **2006**, *110*, 22842.
(31) Lei, H.; Duan, Y. *Curr. Opin. Struct. Biol.* **2007**, *17*, 187.
(32) Comrey, A. L.; Lee, H. B. *A First Course in Factor Analysis*; Academic Press: New York, 1973.
(33) Wilks, D. S. *Statistical Methods in the Atmospheric Sciences*, 2nd ed.; Elsevier: Amsterdam, 2006.
(34) Cox, T. F.; Cox, M. A. A. *Multidimensional scaling*, 2nd ed.; Chapman & Hall: Boca Raton, FL, 2001.
(35) Amadei, A.; Linssen, A. B. M.; Berendsen, H. J. C. *Proteins: Struct., Funct., Genet.* **1993**, *17*, 412.
(36) García, A. E. *Phys. Rev. Lett.* **1992**, *68*, 2696.
(37) García, A. E.; Blumenfeld, R.; Hummer, G.; Krumhansl, J. A. *Physica D* **1997**, *107*, 225.

(38) Romo, T. D.; Clarage, J. B.; Sorensen, D. C.; Phillips, G. N. *Proteins: Struct., Funct., Genet.* **1995**, *22*, 311.
(39) Kaiser, H. F. *Psychometrika* **1958**, *23*, 187.
(40) Case, D. A.; Darden, T. A., Cheatham, T. E., III; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Merz, K. M., Wang, B.; Pearlman, D. A.; Crowley, M., Brozell, S.; Tsui, V.; Gohlke, H.; Mongan, J.; Hornak, V.; Cui, G.; Beroza, P.; Schafmeister, C.; Caldwell, J. W.; Ross, W. S.; Kollman, P. A. *AMBER Simulation Software Package*, version 8; University of California: San Francisco, CA, 2004.
(41) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1996**, *117*, 5179.
(42) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *J. Chem. Phys.* **1995**, *103*, 8577.
(43) Müller, C. W.; Schlauderer, G. J.; Reinstein, J.; Schulz, G. E. *Structure* **1996**, *4*, 147.
(44) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327.
(45) Berendsen, H. H. C.; Postma, J. P. M.; Gunsteren, W. F.; DiNola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684.
(46) Krishnamurthy, H.; Lou, H. F.; Kimple, A.; Vieille, C.; Cukier, R. I. *Proteins: Struct., Funct., Bioinf.* **2005**, *58*, 88.
(47) Atkinson, J. R. PQMethod Software, 1993.

JP8053795