Molecular Dynamics of Apo-Adenylate Kinase: A Principal Component Analysis

Hongfeng Lou and Robert I. Cukier*

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

Received: March 30, 2006; In Final Form: May 1, 2006

Adenylate kinase from *E. coli* (AKE) is studied with molecular dynamics. AKE undergoes large-scale motions of its Lid and AMP-binding domains when its open form closes over its substrates, AMP and $Mg^{2+}-ATP$. The third domain, the Core, is relatively stable during closing. The resulting trajectory is analyzed with a principal component analysis method that decomposes the atom motions into modes ordered by their decreasing contributions to the total protein fluctuation. Simulations at 303 K (normal *T*) and 500 K (high *T*) reveal that at both temperatures the first thee modes account for 70% of the total fluctuation. The residues that contribute the most to these three modes are concentrated in the Lid and AMP-binding domains. Analysis of the normal *T* modes indicates that the Lid and AMP-binding domains sample a broad distribution of conformations indicating that AKE is designed to provide its substrates with a large set of conformations. The high *T* results show that the Lid initially closes toward the Core. Subsequently, the Lid rotates to a new stable conformation that is different from what is observed in the substrate-bound AKE. These results are discussed in the context of experimental data that indicate that adenylate kinases do sample more than one conformational state in solution and that each of these conformational states undergoes substantial fluctuations. A pair of residues is suggested for labeling that would be useful for monitoring distance fluctuations by energy transfer experiments.

1. Introduction

Protein kinases catalyze phosphate transfer reactions.[1] Adenylate kinase (AK) catalyzes the reversible transformation

$$Mg^{2+}-ATP + AMP \Leftrightarrow Mg^{2+}-ADP + ADP \quad (1.1)$$

AKs are composed of three domains: Core, Lid, and AMP-binding (AMP-bd) domains (Figure 1). In the absence of substrates, AKs are in an open form. In the presence of ATP, AMP, and Mg^{2+}, the Lid and AMP-bd undergo major conformational rearrangements, resulting in the enzyme closing to form the ternary complex and expelling waters to prevent ATP and AMP hydrolysis.[2] Extensive kinetic studies of the reaction have shown that it occurs by a random bi−bi mechanism, whereby the substrates can bind to one isoform while the products bind to another isoform of the enzyme.[3] In this case, it leads to the suggestion that the apo form of *E. coli* adenylate kinase (apo-AKE) can exist in at least two conformations, such that one form of the enzyme can bind $Mg^{2+}-ATP$ and AMP and the other $Mg^{2+}-ADP$ and ADP.[3] In addition, other work[4,5] suggests that each of these forms of the enzyme produces its own subensemble of conformers. NMR studies also support the existence of a conformational ensemble for the apo-enzyme.[6] Evidence from crystallographic studies[2] and energy transfer experiments[7] shows that binding AMP is associated with an initial conformational change and binding the next substrate, usually modeled with a binary substrate mimic AP_5A (ATP and AMP linked by a fifth phosphate group), results in formation of the final closed, catalytically competent form. The transition from apo (open) to closed form requires large motions of AK; in particular, the Lid and Amp-bd domains must translate and rotate extensively in order to close the interdomain cleft.[2]

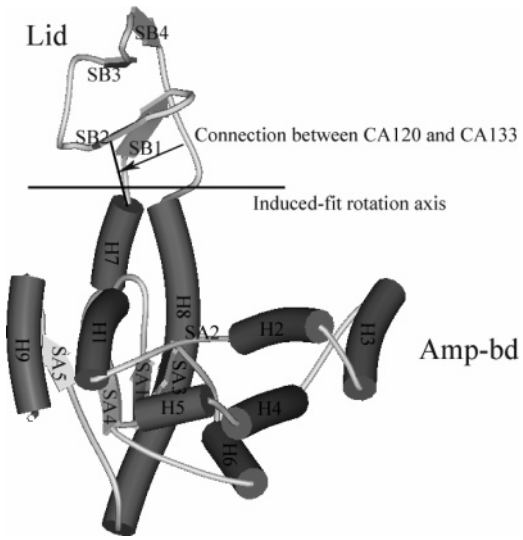

Figure 1. Schematic structure of 4AKE according to the Table 1 assignments. The Amp-bd domain includes H2 and H3 and the Lid domain includes SB1−SB4. The Core is the remainder of the protein.

A few molecular dynamics (MD) simulations of AK have appeared previously. A 300 ps study of AK complexed with AP_5A was carried out in the gas and solvated states.[8] A number of residues that would be candidates for strong interaction with a Mg^{2+} ion were proposed in this study. We simulated the ternary complex by creating a model for $Mg^{2+}-ATP$ and AMP and found that the complex was stable during the MD trajectory in a conformation that would permit phosphoryl transfer by an associative mechanism.[9] A weighted masses molecular dynamics method was used to explore the nature of the open conforma-

* Corresponding author: Professor R. I. Cukier, Department of Chemistry, Michigan State University, East Lansing, MI 48224-1322. E-mail: cukier@cem.msu.edu. Phone: 517-355-9715 (x 263). Fax: 517-353-1793.

10.1021/jp061976m CCC: $33.50
Published on Web 06/07/2006

tions of apo-AK that could be reached starting from a number of closed form (with different bound substrates) crystallographic structures.[10] AKE also has been analyzed from the point of view of the Gaussian network model (GNM).[11] The GNM represents a protein as an elastic network with the CA atoms taken as nodes. The nodes within a cutoff radius are connected by springs of a common force constant. A GNM analysis of apo-AKE points to the Lid and Amp-bd domains as the ones with the largest fluctuations.

When substrates bind to proteins, a recurring theme is that of induced fit whereby the substrates are largely responsible for the protein conformational changes. On the other hand, it may be that protein fluctuations are responsible for setting up conformations that are predisposed to capture substrates to yield the final substrate-bound structure.[12−15] In this work, we use MD to investigate the apo form of *E. coli* adenylate kinase (AKE) with the goal of investigating large-scale domain motions. MD simulations rely on thermal fluctuations to explore configuration space and therefore, if there are substantial barriers to certain regions of configuration space, they will not be sampled on a typical MD time scale. Consequently, in addition to simulations at 303 K we also run MD at 500 K to accelerate exploration of configuration space.

Extracting information about large-scale motions, such as those corresponding to domain movements, from a MD trajectory is not straightforward. Garcia[16,17] and Berendsen and coworkers[18] developed methods, which are applications of principal component analysis[19] (PCA), that attempt to decompose the $3N$ protein degrees of freedom into a small set of modes, p, that capture much of the overall atom fluctuation of the protein and a large remaining set, $3N - p$, that describes small more-or-less Gaussian fluctuations. There are a number of examples[20−26] where this separation is obtained, and it leads to a reduced space description of the important protein motions.

To the extent that the open form of AK is "susceptible" to motions that describe closing, the PCA method may be able to reveal those closing motions. While one cannot expect open AK to close on the MD time scale, nor would it, most likely, close in the absence of substrates, it may be that the PCA will reveal directions of motion that are indicative of the closing path required to entrain its substrates.

Domain closure in AKE has been studied by time-resolved dynamic nonradiative excitation energy transfer.[7] When pairs of residues selected to probe the dynamics of AKE are labeled, distance distributions that reflect the conformational space exploration can be monitored. The apo-enzyme was found to have a broad distance distribution, with a mean and width around 30 Å, between Ala55, a residue in the AMP-bd domain and V169, in helix 8, which is below the Lid (see Figure 1). The width of this distribution indicates that apo-AKE samples a very large conformational ensemble in solution. From their data, they also infer that the rate of conformational transitions is small on a nanosecond time scale, in part motivating our high T simulations. We attempt to correlate the results of our simulations with these studies and, where possible, suggest other residue pairs that could provide additional information about AKE dynamics.

2. Methodology

Molecular Dynamics. The MD simulations were carried out with the Sander module of Amber 7.0[27] at two temperatures, 303 K, which will be referred to as normal T, and 500 K, referred to as high T, both for 6 ns. The Amber 94 force field was used.[28] For the electrostatic energies and forces, the PME method[29] was applied with a direct-space cutoff of 9.0 Å, a Ewald coefficient of 0.30768, and a $80 \times 80 \times 80$ reciprocal space grid. The starting configuration was the apo form of the *E. coli* adenylate kinase[30] (PDB code 4AKE). The dimer's chain A was used. The 4AKE crystal structure coordinates were put into a cubic box of side 79.659 Å with 15 012 water molecules added. Four Na^+ cations were added to neutralize the system, far away from the protein and also from each other. All simulations are done with only hydrogen-containing bonds constrained by SHAKE,[31] a temperature coupling time of 0.5 ps and pressure coupling time of 2.0 ps.[32] Because of crystal strains, a small step size (0.2 fs) was first used to relax the system for 2 ps at 200 K, followed by normal 2 fs steps for 50 ps of equilibration under NVT to bring the temperature from 200 to 303 K. Then, it was equilibrated under NPT for 150 ps to bring it back to the normal density. After equilibration under NPT, 6 ns of simulation under NVT at 303 K were done and all data were collected. After the first nanosecond under NVT at 303 K, the system was heated at a 50 K per 100 ps rate under NVT until 500 K, and then another 100 ps at 500 K was carried out, followed by the 6 ns production run at 500 K under NVT.

Principal Component Analysis. A principal component analysis[19] diagonalizes the covariance matrix $\sigma_{ij} = \langle \delta\alpha_i \delta\alpha_j \rangle$ of the atom fluctuations $\delta\alpha_i$ from their trajectory-averaged $\langle ... \rangle$ values, where $\delta\alpha_i = \alpha_i - \langle \alpha_i \rangle$ and $\alpha_i = \{x_i, y_i, z_i\}$ denotes the Cartesian components of the ith atom. It decomposes the configuration point $\mathbf{r}^{3N}(t) = (x_1(t), y_1(t), ..., z_N(t))^T$ as

$$\mathbf{r}^{3N}(t) = \sum_{i=1}^{3N} [\mathbf{r}^{3N}(t)\cdot\mathbf{m}_i]\mathbf{m}_i = \sum_{i=1}^{3N} p_i(t)\mathbf{m}_i \qquad (2.1)$$

where the $\mathbf{m}_i$ are the (orthonormal) eigenvectors of the covariance matrix and the corresponding eigenvalues are denoted as λ_i^2. In the rotated Cartesian coordinate basis defined by the $\mathbf{m}_i$ ($i = 1, 2, ..., 3N$), the largest eigenvalue captures the largest fraction of the root-mean-square fluctuation (RMSF), the second largest the next largest fraction of the RMSF, 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. PCA is not restricted to harmonic motions; it can describe collective transitions between structures that differ greatly. Note that the $p_i(t)$ defined by eq 2.1 are projections on the configurations $\mathbf{r}^{3N}$ versus projections on the deviations $\delta\mathbf{r}^{3N}(t) = \mathbf{r}^{3N}(t) - \langle \mathbf{r}^{3N}(t) \rangle$ from the trajectory average.

The contribution of atom j to the ith mode's fluctuation is obtained from

$$|\mathbf{m}_i^j| = \sqrt{(m_i^{jx})^2 + (m_i^{jy})^2 + (m_i^{jz})^2} \equiv \text{Component}_i^j$$

The $\mathbf{m}_i^j = \{m_i^{jx}, m_i^{jy}, m_i^{jz}\}$ will be referred to as *component vectors* of the jth atom for the ith mode. The total RMSF^2 can be decomposed as

$$\text{RMSF}^2 = \sum_i \lambda_i^2 = \sum_i \lambda_i^2 \mathbf{m}_i\cdot\mathbf{m}_i = \sum_i \sum_j \lambda_i^2 (m_i^{jx} m_i^{jx} + m_i^{jy} m_i^{jy} + m_i^{jz} m_i^{jz}) = \sum_i \sum_j (\lambda_i \text{Component}_i^j)^2 \equiv \sum_i \sum_j (R\text{-Component}_i^j)^2 \qquad (2.2)$$

The product $\lambda_i \text{Component}_i^j \equiv R\text{-Component}_i^j$ is the contribution of atom j in mode i to the total fluctuation of the protein. This makes it possible to compare the importance of an atom across different modes; thus, we have termed it the relative contribution or R-Component. The time evolution of the ith mode is given

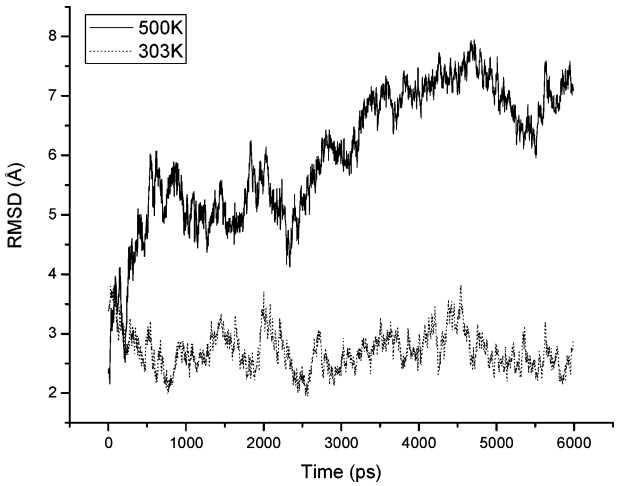

Figure 2. RMSD of CA atoms fit on the X-ray structure CA atoms for normal (303 K) and high *T* (500 K).

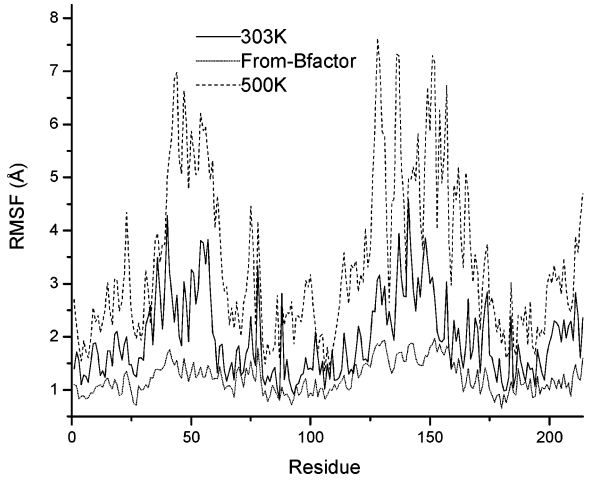

Figure 3. Residue RMSF of residues obtained from superimposing the MD snapshots on the 4AKE Core and from the *B* factors, related by *B* factor = $(8\pi^2/3)\mathrm{RMSF}^2$.[37]

TABLE 1: Secondary Structure Assignment for apo-AKE (4AKE)[a]

helix	initial residue	terminal residue	strand	initial residue	terminal residue	strand	initial residue	terminal residue
H1	13	24	SA1	2	7	SB1	123	126
H2	31	40	SA2	27	29	SB2	131	134
H3	44	54	SA3	81	84	SB3[b]	144	146
H4	61	73	SA4	105	110	SB4[b]	151	154
H6	90	98	SA5	192	197			
H6	90	98						
H7	113	121						
H8	161	184						
H9	202	213						

[a] Assignment of secondary structure from pdb bank 4AKE for chain A. [b] These two beta strands are not present in the pdb bank assignment of 4AKE, while it is in 1AKE and also in Müller et al.'s assignment.[30]

by $p_i(t)$ as introduced in eq 2.1. By examining the subspace spanned by displacements of the first few modes, one gets an idea of whether the system is constrained in one conformational state or not. To examine the time course of the displacement of atom *j* in the *i*th mode, eq 2.1 is used in the form

$$\mathbf{r}_i^j(t) = \mathbf{r}_i^j(0) + (p_i(t) - p_i(0))\mathbf{m}_i^j \qquad (2.3)$$

PCA is a linear transformation conventionally applied to the fluctuations in the Cartesian coordinates.[33] The atom dynamics

TABLE 2: Pearson Correlation Coefficients between the Two Simulation RMSFs and the RMSF Calculated from the *B* Factors[a]

	whole protein	AMP-bd	Lid
high *T*/*B* factor[b]	0.81	0.62	0.54
normal *T*/*B* factor[b]	0.72	0.40	0.54
high *T*/normal *T*	0.77	0.62	0.40

[a] The RMSF are based on superposition on the Core CA atom. [b] *B* factor denotes an RMSF calculated from the X-ray structure *B* factors, where *B* factor = $(8\pi^2/3)\mathrm{RMSF}^2$.[37]

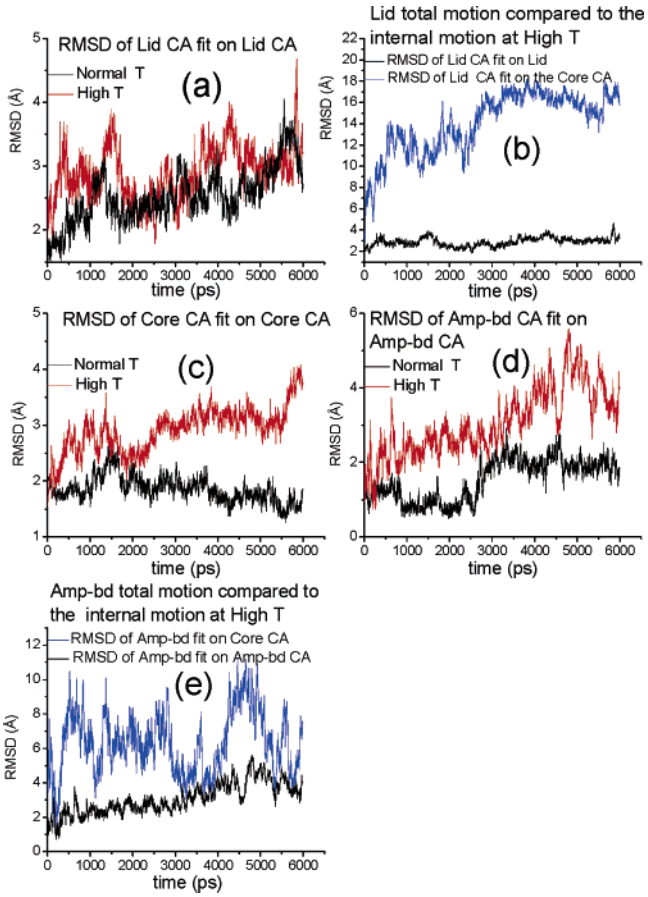

Figure 4. RMSD from the crystal structure of domains Lid, Core, and Amp-bd with different superpositions.

corresponding to the *i*th mode, $\mathbf{r}_i^j(t)$, cannot describe a finite rotation of the atoms. Indeed, any linear transformation will have this feature as is evident from eq 2.3, because vector $\mathbf{m}_i^j$ points in a fixed direction. This feature also implies that the bond vectors predicted from one mode are not of fixed length. Indeed, consider the bond vector $\mathbf{b}(t) = \mathbf{r}_i^{j+1}(t) - \mathbf{r}_i^j(t)$ between atoms $j + 1$ and j. Then

$$\mathbf{b}(t) = \mathbf{b}(0) + (p_i(t) - p_i(0))(\mathbf{m}_i^{j+1} - \mathbf{m}_i^j) \qquad (2.4)$$

and, because $(\mathbf{m}_i^{j+1} - \mathbf{m}_i^j)$ is a constant vector, the magnitude $b(t) = |\mathbf{b}(t)|$ must change with time. The change in bond lengths will usually be modest because, on average, $|\mathbf{m}_i^j| \sim 1/\sqrt{3N}$. However, when $(p_i(t) - p_i(0))$ has a significant excursion, as in a conformational change, the bond length change can be noticeable. Of course, if a sufficient number of terms are included in the mode sum in eq 2.1, the correct atom locations will be regenerated and the distortion will vanish.

The PCA method orders the modes starting from the one with largest eigenvalue. An implication of this decomposition is that

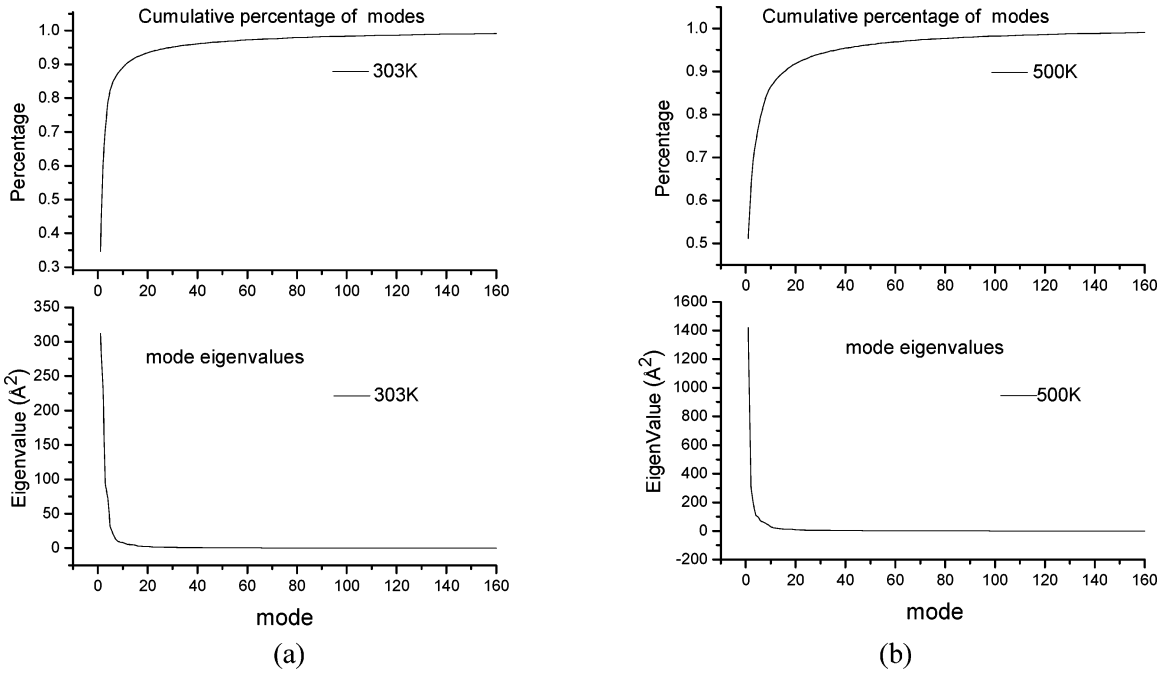

Figure 5. Cumulative percentage and magnitudes of the PCA modes for (a) normal *T* and (b) high *T*.

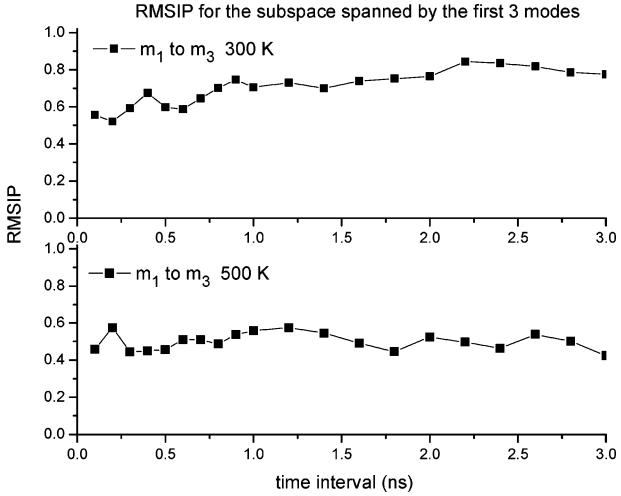

Figure 6. RMSIP (see eq 2.5) of the subspace spanned by the first three modes for time intervals in the first half overlapped with time intervals from the second half of the trajectory.

the largest eigenvalue mode corresponds to the slowest motion, and so forth. Thus, it is important to consider the convergence of the first few modes with the length of the simulation. This issue has been addressed by Amadei and co-workers[34] and by Hess;[35,36] they suggest various tests for convergence. Amadei and co-workers[34] introduced a root-mean-square inner product (RMSIP) measure

$$\mathrm{RMSIP} = \left[\frac{1}{n}\sum_{k=1}^{n}\sum_{i=1}^{n}\mathbf{m}_k(t)\cdot\mathbf{m}_i(t')\right]^{1/2} \qquad (2.5)$$

that is the overlap of a subset of *n* modes, where the modes are obtained from different time intervals taken from the total trajectory. For example, time intervals t' could be taken from the second half of the trajectory starting from the end and the other intervals of length *t* taken from the first half of the trajectory starting from the beginning.[34] Convergence can be assessed in this manner. Furthermore, the (converged) degree of overlap with respect to a random vector's projection onto a subspace (dimension *m*) of the full space (dimension *M*) was analytically obtained.[34] It provides a measure for what values of the RMSIP calculated from trajectory data can be considered sufficiently large to be statistically nonrandom. All the overlaps presented here are sufficiently large to not correspond to a random vector's subspace projection. For sampling from a fully equilibrated, sufficiently long trajectory, the RMSIP measure must, of course, converge to RMSIP = 1. Hess also introduced an RMSIP measure[35] and another one that involves the eigenvalues in addition to eigenvectors.[36]

Before the diagonalization of the covariance matrix is carried out, we first remove the overall translational and rotational motion of the protein by superimposing all the trajectory snapshots onto the Core CA atoms of the X-ray structure. The Core is defined as residues (1−29, 60−121 and 160−214), leaving the Amp-bd domain as (30−59) and the Lid domain as (122−159).

3. Results and Discussion

A. Stability of the Simulation. The root-mean-square deviations (RMSDs) from the crystal structure for the two simulations are presented in Figure 2. For normal *T*, the protein has overall a quite stable structure in terms of the RMSD; this indicates that for the normal *T* (303 K) simulation the protein stays around one major conformation that, in terms of RMSD, is about 2.5 Å away from the X-ray structure. For high *T* (500 K), the protein deviates more significantly from the X-ray structure. As will be shown below, the first deviation (0 to 620 ps) is due to a Lid closure motion. The second major change, between 2360 and 3798 ps, will be traced, by the PCA, to a conformational change in the high *T* simulation.

As shown in Figure 3, the residue RMSF for normal and high *T* are systematically larger than those obtained from the *B* factors,[37] with the differences concentrated in the Lid (120−159) and AMP-bd (30−60) domains and the differences between normal and high *T* also concentrated in these regions. This may indicate that the X-ray-structure is more constrained due to crystallization forces.[38] The similarity in terms of the Pearson correlation coefficient is summarized in Table 2. The high *T*

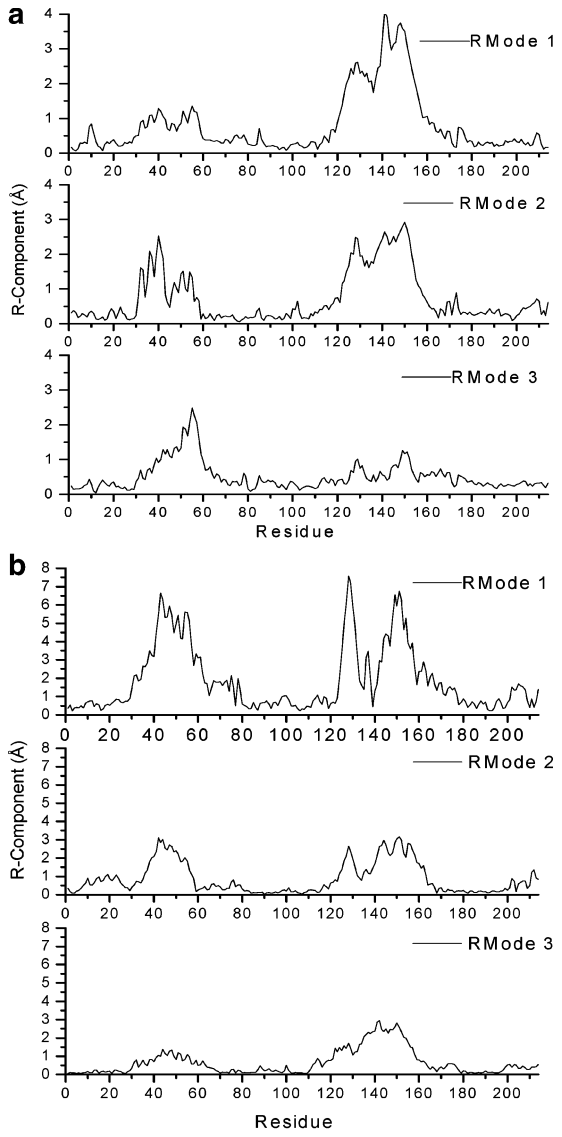

Figure 7. (a) Decomposition of normal *T* modes 1−3 into the contributions from each residue's CA fluctuation. (See eq 2.2 for the definition of R-component). (b) Decomposition of high *T* modes 1−3 into the contributions from each residue's CA fluctuation. (See eq 2.2 for the definition of R-component.)

RMSF is more similar with the X-ray-structure than the normal *T*, as well as the RMSF in the Amp-bd domain. In the Lid, both are the same. Most of the similarity with the X-ray-structure is from the Core, which agrees with the feature that the Core is more rigid than the Lid and Amp-bd domains.

At high *T*, the various RMSDs presented in Figure 4 suggest that the internal structures of the Lid, Core, and Amp-bd domains do not change much compared to their counterparts at normal temperature. The protein maintains its basic integrity and exhibits motions that correspond, in essence, to rigid body rotations of domains around hinges. Figure 4a shows that the amplitude of the internal motions at high and normal temperatures in the Lid is almost the same as that measured by the CA RMSD, when the trajectories are fit on the Lid X-ray structure. For the Amp-bd and Core domains (parts c and d of Figure 4), the CA RMSDs after fitting, respectively, to the AMP-bd and Core domains are somewhat larger at high compared with the normal temperature but are still modest. In contrast, when fit on the Core, the motions of the Lid (Figure 4b) and Amp-bd domains (Figure 4e), relative to their internal motions,

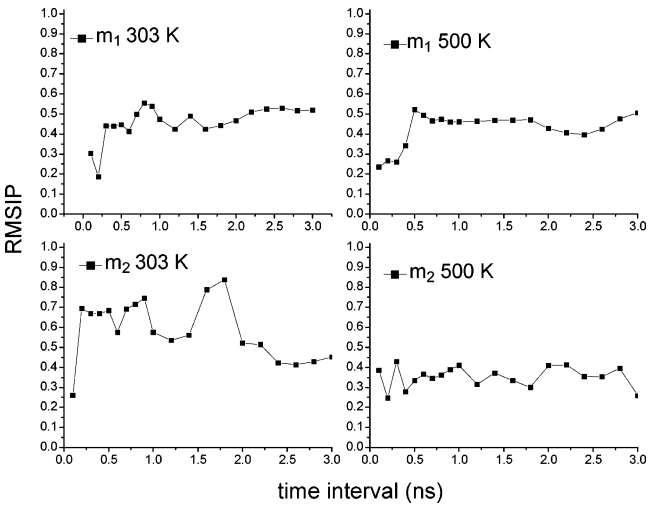

Figure 8. RMSIPs (see eq 2.5) of the first two modes for time intervals in the first half overlapped with time intervals from the second half of the trajectory.

are much larger as measured by RMSDs. In particular, Figure 4b shows most clearly that at high *T* the Lid undergoes a rigid body motion while its internal structure is well preserved. These results suggest that the high *T* simulation is dominated by rigid body motion of domains, in accord with various hinge-bending hypotheses for AKE[2,39] and related enzymes.[14]

An examination of snapshots (not shown) of the Lid, Amp-bd, and Core that correspond to their *extreme* RMSDs from the crystal structure over the entire trajectory (fit on their respective domains in the crystal structure) also shows that there are minimal perturbations from the protein secondary structure within these domains. Thus, we may consider the high *T* trajectory as suggestive of possible domain motions of AKE though, naturally, a simulation at 500 K cannot be viewed as physically realistic.

In this regard, McCammon and co-workers[40] carried out a simulation of Sac7d, a DNA-binding protein, at 550 K. Similar to what we find, while there were larger atom fluctuations relative to their lower temperature simulations, all the secondary structure elements characteristic of the native state were still present at the end of the simulation. They note that this stability may be a property of the protein. They also point out that previous high *T* unfolding studies were carried out with cutoff methods, as opposed to the Ewald-based methods that they (and we) use, and suggest that Ewald methods may artificially enhance stability.

B. Principal Component Analysis Results. 1. Fluctuation of Each Mode. The PCA decomposes the overall protein motion over the simulation time into a set of modes that can be ordered from largest to smallest contributions to the protein RMSF, as measured by the eigenvalues $\lambda_i^2 (i = 1, ..., 3N - 6)$ of the covariance matrix. Fluctuations of the modes are shown in Figure 5 for normal and high *T*. For normal *T*, 12 out of 642 modes contribute 90% of the protein's total RMSF. Modes 1/2/3 contribute 35%/26%/10% of the total, with fluctuations of 312.08 Å^2/232.31 Å^2/92.40 Å^2, respectively. Mode 1 is almost as important as mode 2 in terms of fluctuation, noting that their ratio of fluctuations is 1.3. However, as shown below, mode 3 is as important as mode 2 in describing the motion of the Amp-bd site. For high *T*, 16 out of 642 modes take 90% of protein's motion. Modes 1/2/3 contributes 51%/11.4%/ 7% of the total, with a fluctuation of 1420.5 Å^2/317.77 Å^2/195.06 Å^2, respectively. Mode 1 for high *T* is dominant due to a large conformational change, while the second and third modes are

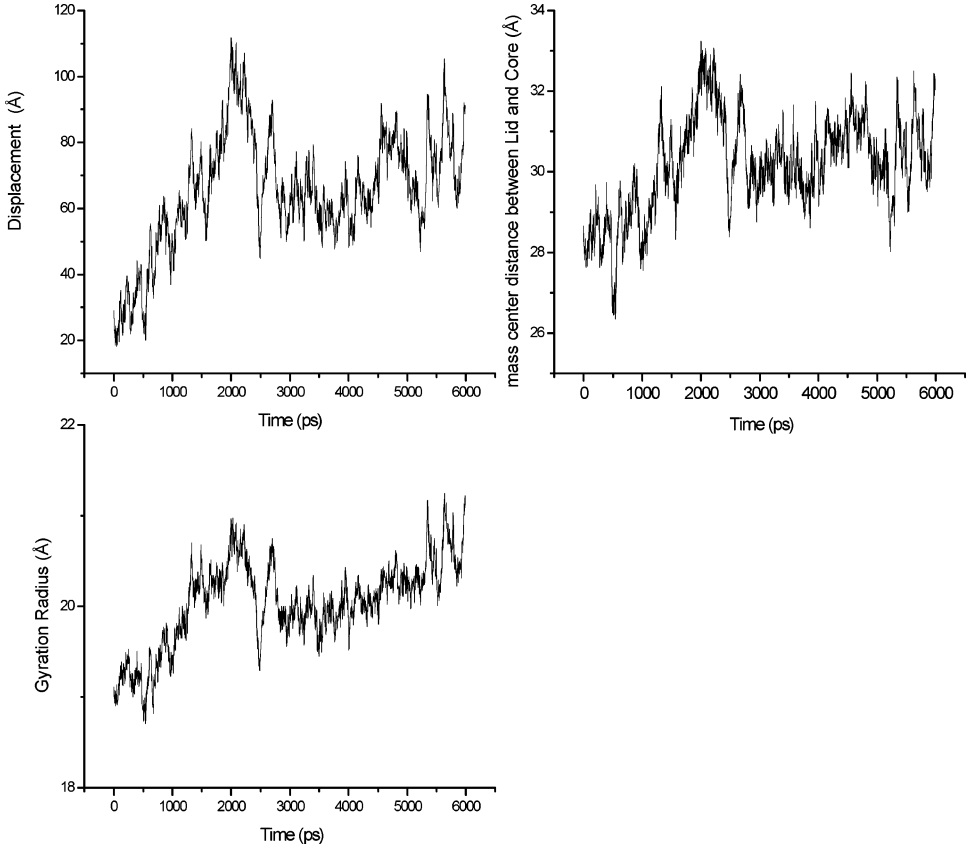

Figure 9. The mode 1 displacement, distance between mass centers of Lid and Core, and the radius of gyration as a function of time.

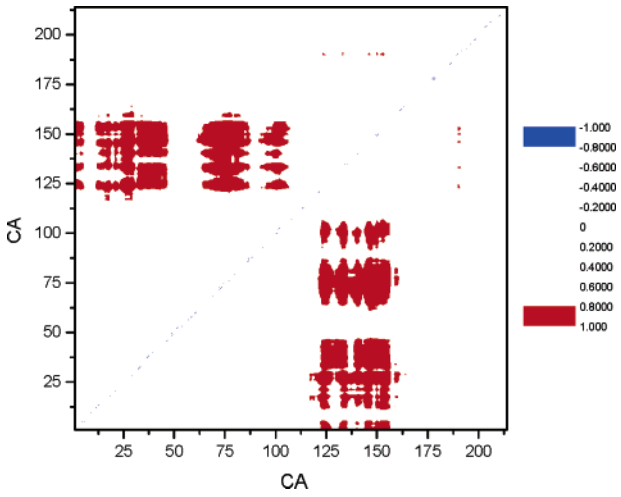

Figure 10. The contour of the Pearson correlation between the time series of the mode 1 displacement and the distance between any pair of CAs, for normal *T*.

comparable to the first and second mode for normal *T* in terms of their size of fluctuation.

The first three modes account for ~70% of the total fluctuation at both temperatures. It is important to check that these large magnitude, and therefore slow, modes are reasonably converged for the simulation length. One method of convergence check, developed by Amadei and co-workers[34] and by Hess,[35] measures the root-mean-square inner product (RMSIP) of PCA modes obtained from one part of the trajectory projected onto another part of the trajectory. Figure 6 displays the overlap of the subspace spanned by the first three modes ($n = 3$ in eq 2.5) obtained from time intervals of varying lengths, as described in the Methodology section. The results for the normal temperature are reasonably converged, though not fully reaching unity. For high temperature, it is clear that the simulation is not long enough for convergence. In view of the transition behavior of the high-temperature simulation for the RMSD (see Figure 2) convergence is not expected for this length of simulation.

2. Atom Fluctuations for Modes 1−3. The contributions to the RMSF, denoted as R-components (see eq 2.2), from each residue CA for the first three modes for normal and high *T* are shown in parts a and b of Figure 7, respectively. Note that, as defined in eq 2.2, the R-Component$_i^j$ incorporates the corresponding λ_i value. Thus, the R-components permit comparison of fluctuations across different modes. For normal *T*, mode 1 has two major regions, spanning, respectively, residues 30−60 that contain the Amp-bd and 120−160 that contain the Lid domain. There are three peaks in the Lid, with the highest at Lys141 having R-component 4.07 Å. In the Amp-bd domain, the highest peak, at Lys40, has an R-component of 1.29 Å. Thus, the contribution of the Lid is about 9 times that of the Amp-bd in terms of the highest peak, and fluctuations of the Lid dominate the mode 1 motion. In mode 2, there are three prominent regions, spanning residues 30−45, 45−60, and 120−160. For these three regions, the contributions to the fluctuations of mode 2 are almost equally important. In mode 3, only the AMP-bd region contributes significantly to the fluctuation, spanning residues 30−60 with a peak of 2.50 Å at Ala55. Because the R-component plots put the fluctuations in the different modes on the same scale, one can conclude that the contribution to the fluctuation of the Amp-bd domain is about four times as large in mode 2 as in mode 1, with a peak of 2.5 Å at Lys 40 in mode 2 and 1.4 Å at Ala 55 in mode 1. In mode 3, the same region contributes to the fluctuation as much as the corresponding one in mode 2 with an R-component peak of 2.48 Å at Ala55. However, the contribution of the Lid from mode 2, with

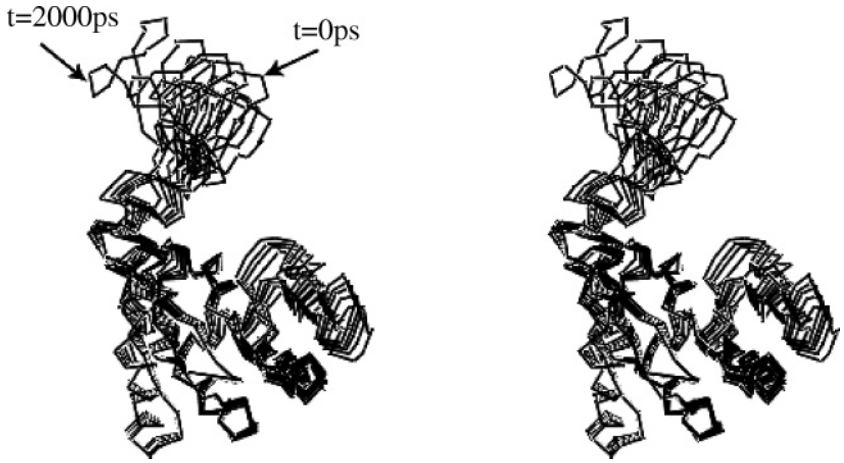

Figure 11. Stereoview of snapshots of the CA atom projection of mode 1, for normal *T*.

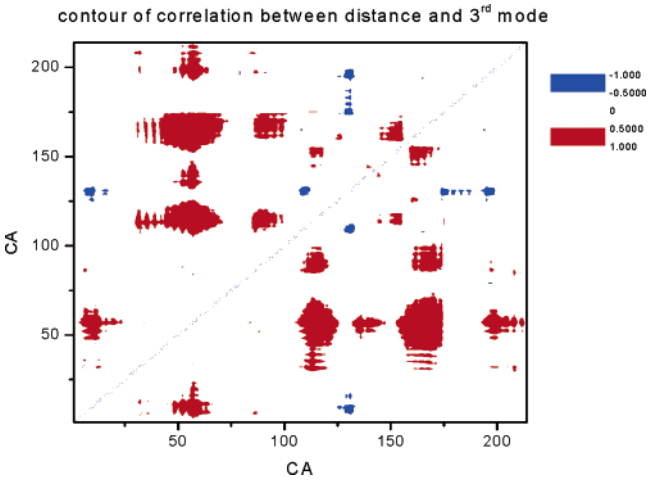

Figure 12. The contour of the Pearson correlation between time series of the mode 3 displacement and the distance between any pair of CAs, for normal *T*.

a peak of 2.9 Å at Gly150, is about 58% of the one from mode 1 with its peak of 4.0 Å at Lys141, while the contribution of the Lid from mode 3 is ignorable. In short, for normal *T*, in terms of fluctuations, mode 1 is suitable to describe the Lid's motion, while mode 2 describes a collective motion of the Lid and Amp-bd domains. Moreover, mode 3 describes the motion of the Amp-bd domain, and this is the only significant contribution to this mode.

The motion described by mode 3 may be related to the first step in the induced-fit movements proposed by Schulz.[2] In their scheme, based on a comparison of apo, AMP-bound and AP5A-bound crystal structures, the first conformational change is dominated by a closing motion of the AMP-bd domain toward the Lid. Sinev et al. studied the large domain movements in AKE by measuring the distance distribution between Ala55 and Val169 with time-resolved dynamic nonradiative excitation energy transfer.[7] In the MD simulation, Ala55 has its largest fluctuation in the third mode, while Val169 has no big fluctuation in any of the major modes. (Of course, a large distance fluctuation can be consistent with a significant motion by only one residue of the pair.) Thus, the motion described by mode 3 is consistent with the experimental results.

The simulation also suggests that a residue pair better directed to featuring Lid motions would be Lys40 and Lys141. These residues have peak R-Components in mode 1 (and are also prominent in mode 2). However, in contrast with mode 1, the correlation between the distance between these two residues and the displacement of mode 2 is close to zero. These mode 1 and 2 motions are more important, at least in terms of fluctuation size, than that of mode 3.

For high *T*, mode 1 has three equally important regions spanning residues 30−60, 120−139, and 140−160 with peaks of 6.78 Å at Ser43, 7.53 Å at Pro128, and 6.78 Å at Glu151. Mode 2 has two regions spanning residues 30−59 and 115−168 with peaks of 3.11 Å at GLY42 and 2.65 Å at Pro128 and 3.15 Å at Glu151. Two important regions are identified in mode 3; regions spanning residues 28−73 and 108−168 with the later being more significant. Comparing the magnitudes of the R-components across the different modes suggests that the mode 1 is dominant not only in the Lid but also in the Amp-bd domain. Thus, this mode represents the major motion for high *T*. The data in parts a and b of Figure 7 indicate that the Amp-bd and Lid domains tend to move more collectively at the higher *T*.

As pointed out by Amadei and co-workers,[34] the individual eigenvectors may not be converged for finite time simulations. Thus, we have calculated the RMSIP of eq 2.5 (with $n = 1$) between a particular eigenvector from the analysis of the last 3 ns of the trajectory and the one it overlaps the most, selected from the first three eigenvectors $\{m_1, m_2, m_3\}$, from the first 3 ns, evaluated over various time subintervals. Figure 8 shows that the RMSPs for the first two vectors for both temperatures are not approaching unity over the nanosecond time scale. Despite the lack of convergence, the functional motions implied by these modes are still of interest. They must be interpreted as holding over a limited time range and not true equilibrium results. For the third mode at 500 K, the convergence is even weaker, and we do not discuss it further.

3. The Characterization of Four Modes for Normal and High *T*. *Mode 1, Normal T.* Mode 1 for normal *T* is dominated by fluctuations in the Lid region. In a proposed mechanism of the ATP-AMP reaction,[2], first the AMP-bd domain partly closes, which is associated with the binding of AMP. Then, there is a large motion of the Lid, along with additional AMP-bd domain motion to bind the substrates (AMP and MgATP) and exclude (most) waters from the active site. From the X-ray structure of 4AKE, one may infer that the radius of gyration of this protein should have a strong dependence on the Lid's position relative to the Core.

Figure 9 displays time series of the displacement of mode 1, the distance between the Lid and Core mass centers (based on the atoms in the respective regions), and the radius of gyration. It is clear that these three quantities are strongly correlated, with a Pearson correlation coefficient of 0.91 between the mode 1 displacement and the radius of gyration and 0.92 between the

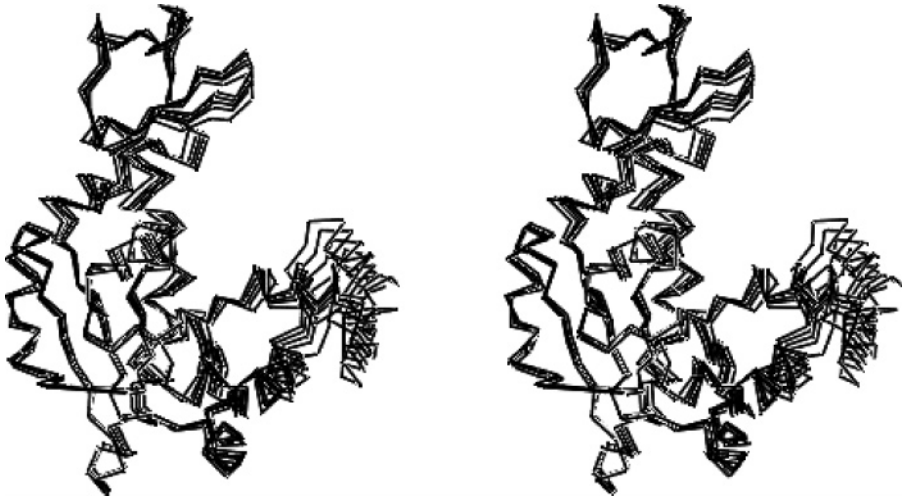

Figure 13. Stereoview of snapshots of the CA atom projection of mode 3 for normal *T*. The motion of the Amp-bd domain is evident in this projection.

mode 1 displacement and the mass center distance of Lid and Core regions. This roughly shows that the mode 1 motion is related to the closing or opening of the Lid. From the time course of the center of mass distance, the Lid first opens up until around 2 ns; after that a closure follows and then the distance begins to fluctuate around 30 Å. The standard deviation of the mass center distance is 1.20 Å with minimum of 26.34 Å and maximum at 33.23 Å.

Figure 10 presents a contour plot of the Pearson correlation coefficients between the time series of the mode 1 displacement, $p_1(t)$, and the distances between any pair of CAs. Correlation coefficients bigger than 0.8 are found between the Lid (122−159) and regions spanning residues 15−47, 65−89, 96−106, as well as the region around residue 190. Residues 15−47 contain α-helix 1, α-helix 2, β-strand SA2, and loops between them. Residues 65−89 contain α-helix 4, α-helix 5, and β-strand SA3. Residues 96−106 define the loop between β-strand SA3 and α-helix 6. These three regions represent the neighborhood and part of the Amp-bd domain. The residues around residue190 are at the end of α-helix 8. The correlation pattern confirms that mode 1 characterizes fluctuations in the Lid-to-Amp-bd domain distances.

The lack of correlation of the distance between residues 55 and 169 with the displacement of mode 1 again suggests that the probe residue selection in the energy transfer experiment[7] is not appropriate to capture the motion described by mode 1. The strong correlation of distances between residues in the range 15−47 and Lid residues (whose motion dominates mode 1) with mode 1, reinforces our suggestion of the choice of residue pair Lys40 and Lys141 as a probe for AKE's dominant motion.

In Figure 11, snapshots of the CA coordinates from mode 1 are displayed for normal *T*. The pattern shows how the Lid and Amp-bd domains move relative to each other.

Mode 3, Normal T. Mode 3 mainly involves the Amp-bd region, as indicated in Figure 7a. Medium correlation is found between some CA distance pairs and the mode 3 displacement, as mapped in the contour in Figure 12. The region spanning residues 53−63 (loop region) has medium correlation with regions spanning residues 5−15 (first loop and part of α-helix 1), 108−124 (α-helix 7 and part of a loop), 133−147 (SB4 and the loop between SB3 and SB4), 153−174 (a loop and α-helix 8), and 197−214 (α-helix 9 and the last loop).

The medium correlation found between residue pairs in the ranges 53−63 and 153−174 with mode 3 shows that the Sinev et al.[7] choice of residue pair (55 and 169) is a good indicator for the motion captured in mode 3. The mode 3 projection displayed in Figure 13 makes it clear that part of the Amp-bd site approaches the Lid and helix 8, which is similar to the suggested first step in the induced-fit mechanism,[2] where AMP binds to the protein and the Amp-bd site closes somewhat.

Mode 1 for High T. The high *T* trajectories were run with the idea of increasing the atom fluctuations in order to lead to a clearer picture of how AKE can close. Certainly, the total PCA fluctuation of 2776.98 Å^2 for high *T* is considerably larger than the normal *T* value of 899.03 Å^2. PCA for high *T* reveals a quite different picture of the dynamics. The distances between the Lid and Core mass centers in the X-ray structures of the closed and open forms are 20.1 Å[41] and 30.2 Å,[30] respectively. Figure 14 shows that the mass center distance does come close to 20 Å, which indicates closure of the Lid. The mode 1 and 2 displacements shown in Figure 14 indicate a conformational change that involves the Amp-bd and Lid domains.

The large, two conformational state motion evident in mode 1 suggests that the corresponding atom displacements may be similar to the Lid rotation (see Figure 1) proposed by Schulz et al.[2] as the second step in closing AKE to its catalytic conformation. That turns out to not be the case, as we now show. Figure 15 displays the Cartesian atom displacements for mode 1 at times along the trajectory spanning before and after the large transition behavior evident in Figure 14. The blue traces are for the first 3 ns while the red ones are for the last 3 ns. After the transition, the Lid is squashed into a small region, an apparent distortion of the protein. We noted in the methods section that bond lengths are not necessarily preserved in a PCA analysis when the atom motions are reconstructed from just one mode and finite rotations are not describable by one mode. Figure 15 plots only CA atoms, but the neighboring CA distances will manifest the bond length changes. The squashing evident in Figure 15 is a manifestation of this behavior. Nevertheless, the main features of the mode 1 motion can be obtained, and compared with the suggested Lid rotation. First, note that the Lid residues (residues 122−159) have only medium CA RMSDs (data not shown), averaging 2.8 Å with standard deviation 0.4 Å and a maximum RMSD of 4.6 Å. Also, the fluctuation of the radius of gyration of the Lid (data not shown) is only about 1 Å while the width of the Lid is about 23 Å. These facts show that the Lid still holds its basic shape during the high *T* simulation. In other words, the Lid can only do, roughly, a large rigid rotation or, more likely, a combination of a large rotation and translation.

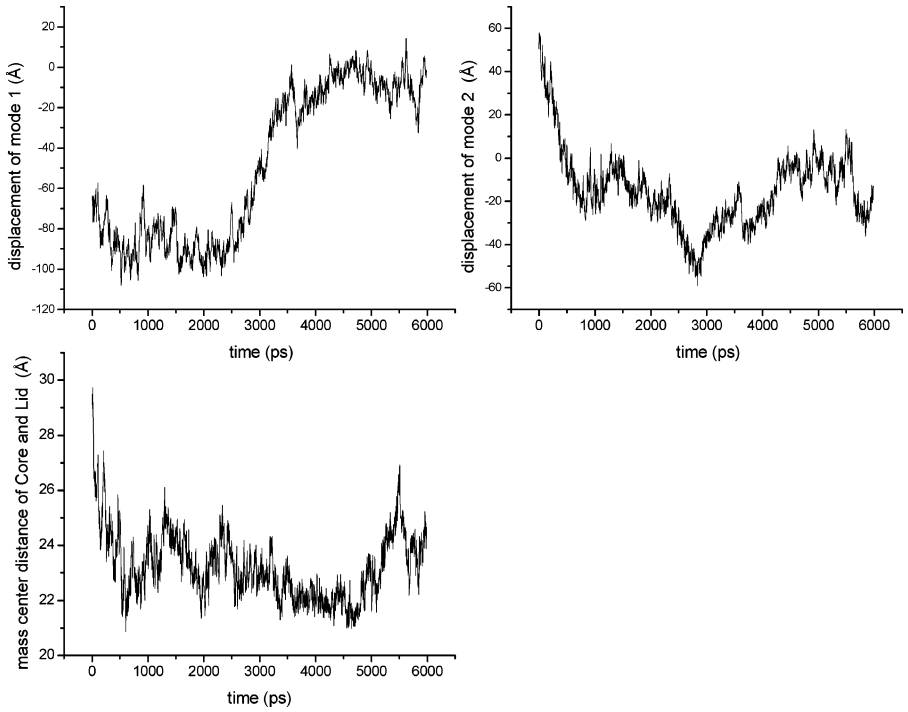

Figure 14. The displacements of mode 1, mode 2, and the distance between the mass centers of the Core and Lid, for high *T*.

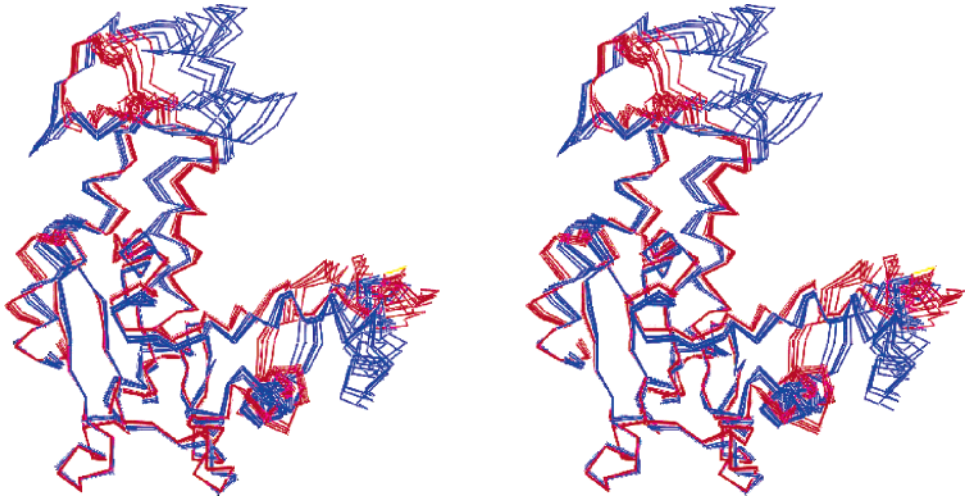

Figure 15. Stereoview of snapshots of the CA atom projection of mode 1, for high *T*. The blue traces are for the first 3 ns and the red ones are for the last 3 ns.

To explore if the mode 1 motion is a projection of the suggested Lid rotation, define a plane, which we will refer to as plane 1, formed by the component vectors $\mathbf{m}_1^j$ (cf. before eq 2.2) of two CA atoms, $j = 124$ and $j = 143$ in mode 1. If one connects the CA 120 and CA 133 atoms in the X-ray structure (see Figure 1), then the resulting vector between these two atoms will be almost perpendicular to plane 1 (about 85°). Angles between component vectors of CAs in mode 1 in the Lid and plane 1 are calculated. Most of the component vectors of mode 1 for the other Lid CAs are almost parallel to plane 1 (<10°), with the exception of the component vectors of CAs 120−123, 138, and 157−160. This shows that the motion described by mode 1 is mainly parallel to plane 1. The axis by which the protein closes the Lid in the Schulz et al.[2] scheme lies almost parallel to this plane. If it were the case that mode 1 is the projection of the rotation that closes the Lid, then all its component vectors would be perpendicular to this rotation axis which, combined with the fact that the axis is almost parallel to the plane 1, means that all component vectors would have to be in the same direction. Calculation of the angles between the component vectors shows this to not be the case. Consequently, mode 1 is not completely related to the closure of the Lid. Two snapshots, one before (602 ps) and one after (5402 ps) the transition, are shown in Figure 16. From these snapshots, one can easily see that this motion is not related to the closure of the Lid. Rather, it corresponds to a rotation of the Lid to a new conformation. That the MD of the apo form samples more than one distinct Lid conformation does not contradict the Lid rotation path found by Schulz et al.,[2] since they base their path on the results of the apo, AMP, and AP_5A bound X-ray structures.

From the contour of the Pearson correlation between the time series of the mode 1 displacement and the distance between any pair of CAs, for high *T* (not shown), one notices that the distance between residues in the regions 163−174 (α-helix 8) and 28−86, which incorporates the AMP-bd domain, is anti-

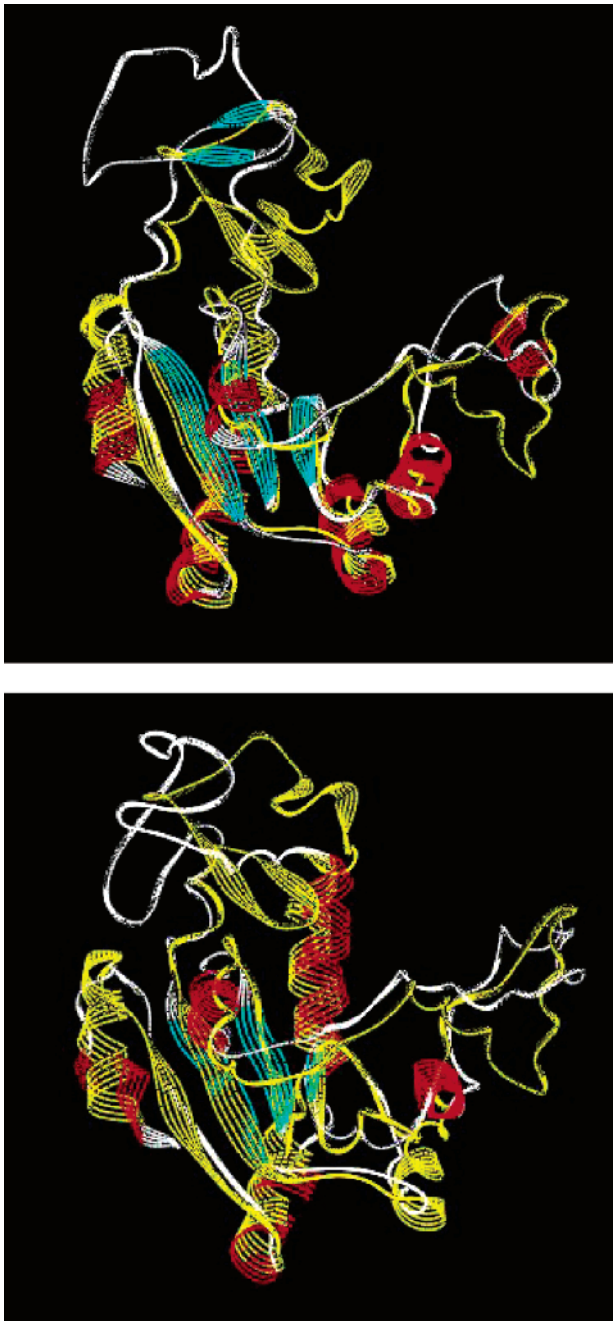

Figure 16. (top panel) Snapshots at 2 and 602 ps (yellow is 602 ps). The approach of Lid and Core is apparent. (bottom panel) Snapshots at 602 and 5402 ps (yellow is 602 ps).

correlated (<-0.8). This shows that the AMP-bd site approaches Helix 8 when the transition in the Lid occurs, as is apparent in Figure 15. Thus, the Lid rotation and the AMP-bd motions are correlated.

Mode 2, High T. Figure 14 shows that the mass center distance does come close to 20 Å, which indicates closure of the Lid. Two snapshots are shown in Figure 16 (top panel), which illustrates the closure of the Lid toward the Core. Snapshots of the CA atom displacements obtained from mode 2 are presented in Figure 17, confirming that the Lid is closing to the Core region.

The time courses shown in Figure 14 of the Lid-to-Core mass centers distance and mode 2 appear correlated over the first 2000 ps (before the mode 1 transition takes place), but the correlation coefficient is only 0.76. When the mode 1 transition takes place, the Lid-to-Core mass centers distance change is not only related to mode 2 but also related to mode 1; thus the Lid-to-Core and mode 2 correlation is poor. The changes in the mode 2 trajectory and the distance of mass centers also coincide in time with the first transition in the RMSD, as shown in Figure 2, which shows that the first transition in RMSD is due to the Lid closure.

The just-enumerated features of the high T data permit a description of the important protein motions. First, the Lid closes toward the Core region. This closure corresponds to the first changes in RMSD (0−600 ps). In terms of RMSD relative to the AP_5A-bound X-ray structure,[41] the Lid has closed in a fashion similar to the AP_5A-bound form. Then, with the Lid remaining closed, it performs a conformational change, reflected in mode 1. The resulting conformation is different from the X-ray structure conformation of both closed and open forms. This transformation corresponds to the second change in RMSD in Figure 2. The process is represented by the snapshots in Figure 16.

4. Histograms of the Mode Displacements. The time course of the projections of the deviation from the average of the system vector onto the $\mathbf{m}_i$ directions, $p_i(t) = \mathbf{r}^{3N}(t)\cdot\mathbf{m}_i$, can be made into histograms that form probability distributions of various orders

$$P_1(x_i)\,\mathrm{d}x_i = \langle\Delta(x_i - p_i(t))\rangle$$

$$P_2(x_i,x_j)\,\mathrm{d}x_i\,\mathrm{d}x_i = \langle\Delta(x_i - p_i(t))\Delta(x_j - p_j(t))\rangle$$

$$(i \neq j)$$

where Δ is an indicator function with value 1 when $p_i(t) = x_i$ on the mesh size $\mathrm{d}x_i = \mathrm{d}x$ and 0 otherwise, and the bracket indicates a trajectory average. Potentials of mean force[42] W can be defined from these histograms according to, e.g., in the two-dimensional case $W(x_1,x_2) = -k_BT \log P_2(x_1,x_2)$, providing insight to the stable and transition state structures in these reduced spaces. The connection between these probabilities and potentials of mean force are valid when sampling from an equilibrium ensemble. Because the convergence of the individual eigenvectors is not complete, based on the RMSIP measure, the one-dimensional histograms must be interpreted as measures of sampled states rather than (equilibrium) potentials of mean force.

We check convergence of the subspaces spanned by mode pairs by evaluating the RMSIPs in eq 2.5 for the three pairwise combinations, $\mathbf{m}_1$, $\mathbf{m}_2$, $\mathbf{m}_1$, $\mathbf{m}_3$, and $\mathbf{m}_2$, $\mathbf{m}_3$, of the first three modes. Again, the later parts of the trajectory are overlapped with the earlier, as described in the methodology section. For normal T the RMSIP averaged over the three mode combinations approaches ~0.6 while for high T it approaches ~0.4 indicating a lack of convergence in the pair subspace. Thus, even in the pair space, the two-dimensional histograms do not provide accurate potentials of mean force but can be used to indicate states (conformations) that are sampled during the simulation.

For normal T, the histograms in Figure 18a show that the first three modes are approximately Gaussian, with mode 2 not fitting quite as well as the others. For the high T data, Figure 18b, shows that mode 1 is a two-peak Gaussian distribution with the separation between the peaks larger than the peak widths. The remaining two modes are also Gaussian distributions with approximately the same width of those for normal T.

Figure 19 presents two-dimensional histograms in the planes spanned by the first three modes. Peaks indicate a relatively stable conformation; the higher the peak, the more stable the conformation. For normal T, there are three close peaks for the

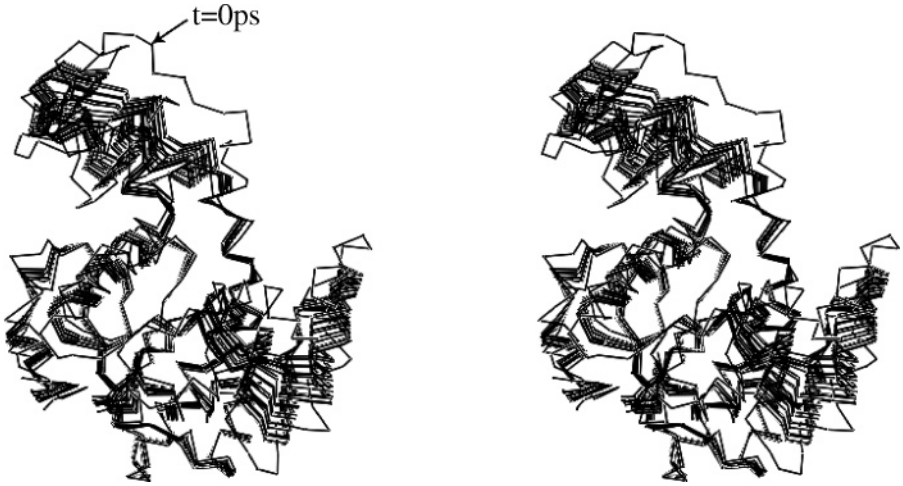

Figure 17. Stereoview of snapshots of the CA projection of mode 2 for high *T*.

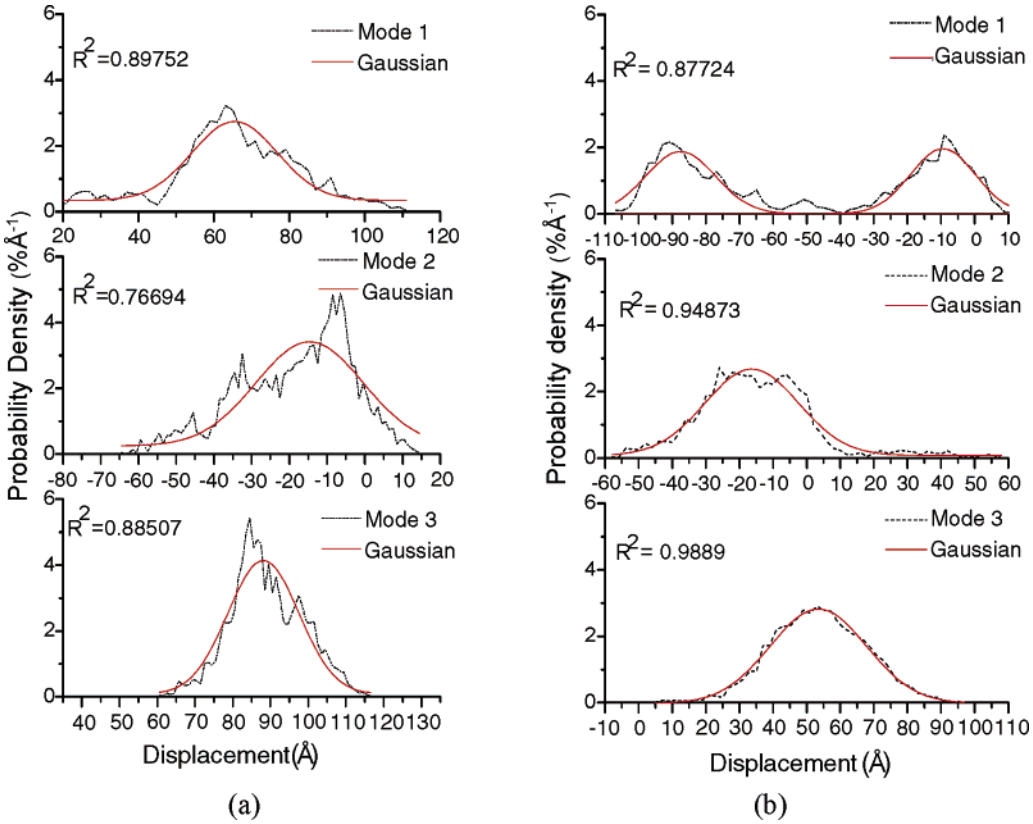

Figure 18. Histograms of the displacements of the first three modes for (a) normal *T* and (b) high *T*. R^2 is the squared Pearson correlation coefficient between the fitted curve and the original data.

plane spanned by modes 1 and 2, two close peaks for the plane spanned by modes 1 and 3, and three close peaks for the plane spanned by modes 2 and 3. However, the latter peaks are near to each other and do not have deep valleys between them. Thus they represent one major conformation. The trajectory producing the modes 1 and 3 contour starts in the bottom left corner and transits more-or-less diagonally before sampling the indicated probability maxima. All three two-dimensional probability distributions are quite broad indicating that in the space spanned by these three modes, which are dominated by Lid and Amp-bd motions, there is extensive sampling of configuration space.

For high *T*, planes spanned by modes 1 and 2 are obviously split into two different regions due to the two-peak Gaussian distribution for mode 1 in Figure 18b. Compared to the conformations appearing for normal *T*, these two peaks are relatively remote and with a relatively deeper valley between them. In the plane spanned by modes 1 and 2, the regions on the left and right have the same height and thus the same stability, while the right is split into two relatively equal, stable subconformations. In the plane spanned by modes 1 and 3, the left one is relatively less stable than the right one, and the right one is split into two equally stable subconformations. In the plane spanned by modes 2 and 3, similar to normal *T*, one region is found.

If the two-dimensional probability distributions arose from separable degrees of freedom in the $\mathbf{m}_i$ basis, then $P_2(x_i,x_j) = P_1(x_i)\,P_1(x_j)$. Thus, probabilistic independence or dependence can be inferred from a comparison of the data in Figures 18 and 19, and it is evident that the modes are dependent. The dependence between modes indicates that the motions described

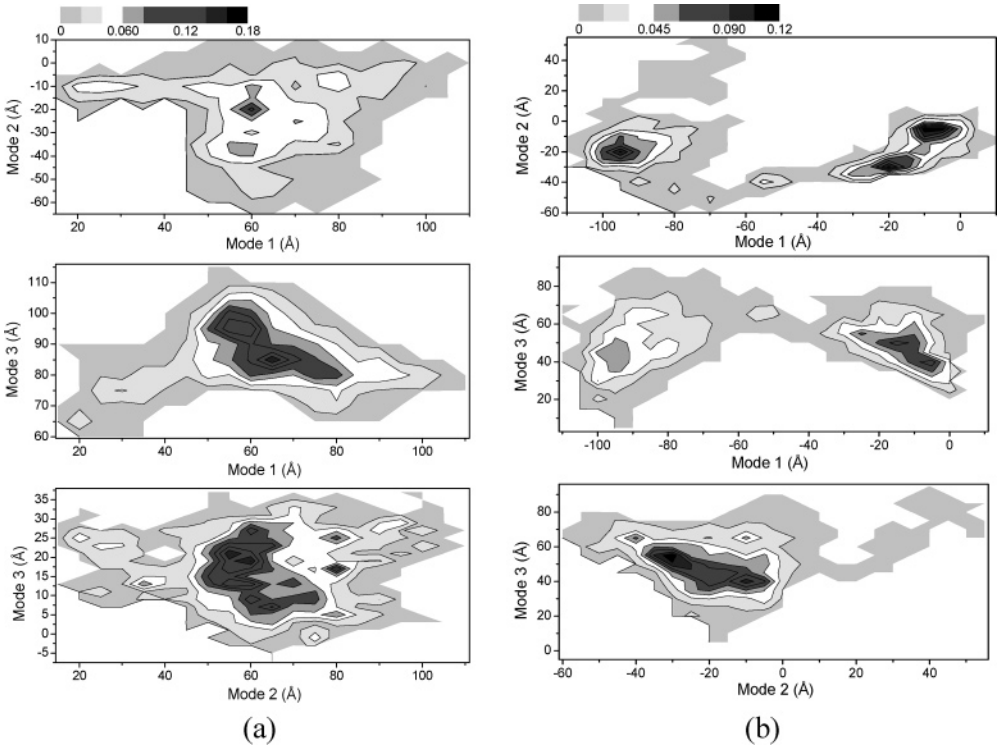

Figure 19. Distribution of the planes spanned by the first three modes for (a) normal *T* and (b) high *T*.

by these modes are not harmonic because, if the atom motions were strictly harmonic, then they would correspond to true normal coordinates. These are, by definition, independent, and their joint probability distribution would be the product of the single mode distributions. Along with the multiple conformations found at high *T*, this indicates that a harmonic analysis, even at the normal *T*, would not be appropriate to characterize the motion of AKE.

4. Concluding Remarks

The PCA method applied to the open form of AKE shows that approximately 70% of the fluctuations in both normal and high T MD trajectories are captured by just the first three modes, indicating that a good account of the motion should be interpretable from an analysis of these modes. The motions described by these modes are dominated by the Lid and Amp-bd dynamics that are essential for closure. Even without substrates present, there are indications of how open AKE can fluctuate toward its closed form. The difference in fluctuation between normal and high *T* (899.03 Å^2 vs 2776.98 Å^2) is dominated by the mode 1 differences (312.08 Å^2 vs 1420.5 Å^2). This arises from the conformational change at high *T* that dominates mode 1. Excluding the large effect on the total fluctuation from the two-state motion, the widths of the distributions displayed in Figures 18 and 19 imply that the normal *T* sampling is comparable to that at high *T*. Thus, the Lid and Amp-bd domains sample large conformational spaces, indicating rather flat potentials of mean force in the space of the first three PCA modes.

For normal *T*, mode 1 is dominated by Lid motion, and Figure 11 shows that the Lid and Amp-bd domain move relative to each other. Mode 2 describes a collective motion of the Lid and Amp-bd domains. Mode 3 describes motion of the Amp-bd domain, and this is the only significant contribution to this mode. The mode 3 motion may be related to the first step in the induced-fit movements proposed by Schulz et al.,[2] whereby the AMP-bd domain closes toward the Lid and helix 8 to form what will be the AMP binding site.

In the energy transfer studies that measure distance distributions, Ala55 and Val169 were used.[7] Our simulations show that Ala55 has the largest fluctuation in mode 3. Thus, the simulation does indicate that an important motion of AKE is being accessed with this choice of labeled residues. However, mode 1 has greater fluctuations and, based on the peaks in Figure 7a, we suggest choosing the residue pair Lys40 and Lys141 as probes that would emphasize the most dramatic motions associated with the Lid.

The behavior at high *T* introduces several new features; the most significant one is the presence of two distinct conformational states. Note that the configurations sampled at normal *T* cannot be considered as a subset of the higher *T* configurations. During the MD trajectory, first the Lid closes toward the Core region. This closure corresponds to the first changes in RMSD (0−600 ps). Modes 1 and 2 characterize this closing motion. Then, with the Lid remaining closed, it performs a conformation change, reflected in mode 1. The resulting conformation is different from the X-ray structure conformation of both closed and open forms. This transformation corresponds to the second change in RMSD in Figure 2. The overall process is represented by the snapshots in Figure 16.

Thus for normal *T*, the protein samples one main conformation, while for high *T* the protein samples two relatively quite different conformations. As discussed before, one of these conformations is quite different from both the closed and open X-ray structures. Zhang[4] proposed an isorandom bi−bi mechanism for the reaction of AK and concluded that there are at least two open conformations in solution, with each one sampling a broad range of conformers.[3,5] The high *T* simulation suggests that there are indeed at least two distinct native conformations in solution. The mechanism of sampling different conformational states for the high *T* simulation consists of more or less rigid body domain motions around hinges. This result agrees with the suggested rigid body domain motions invoked

in AKE originally, based on an examination of several closed and the open form crystal structures,[2,30] though the particular motions seen in the MD are not the same as those suggested in the crystal structure studies. The specifics of these domain motions are of course influenced by the bound ligands and, therefore, comparison with the ligand free simulations can only be qualitative.

The RMSIP analysis shows that, especially for high T, a 6 ns simulation is not long enough to converge the results for all the modes. Thus, the generated data do not correspond to a trajectory describing equilibrium fluctuations. This is certainly to be expected for the high T simulation where a large conformational transition is observed. For the normal T, lack of convergence indicates that as the simulation progresses, new states may be sampled. Also, the proportions of states sampled that are separated by relatively high barriers (measured by the simulation time scale) can be evolving. These features are of course endemic to protein simulations, with their rough energy landscapes. The dominant PCA modes reflect the longer time scale motions and, while standard measures of convergence such as the total protein potential energy might appear converged, other measures such as the first few PCA modes may not be converged.

Acknowledgment. This work is supported by the NIH (Grant No. GM62790).The simulations were carried out on the Quantitative Biology and Modeling Initiative and the Michigan Center for Biological Information Linux clusters at Michigan State University.

References and Notes

(1) Walsh, C. *Enzymatic reaction mechanisms*; W. H. Freeman: San Francisco, CA, 1979.
(2) Schulz, G. E.; Muller, C. W.; Diederichs, K. *J. Mol. Biol.* **1990**, *213*, 627−630.
(3) Sheng, X. R.; Li, X.; Pan, X. M. *J. Biol. Chem.* **1999**, *274*, 22238.
(4) Zhang, H. J.; Sheng, X. R.; Niu, W. D.; Pan, X. M.; Zhou, J. M. *J. Biol. Chem.* **1998**, *273*, 7448.
(5) Han, Y.; Li, X.; Pan, X. M. *FEBS Lett.* **2002**, *528*, 161.
(6) Burlacu-Miron, S.; Gilles, A. M.; Popescu, A.; Barzu, 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) Kern, P.; Brunne, R. M.; Folkers, G. *J. Comput.-Aided Mol. Des.* **1994**, *8*, 367.
(9) Krishnamurthy, H.; Lou, H. F.; Kimple, A.; Vieille, C.; Cukier, R. I. *Proteins* **2005**, *58*, 88.
(10) Elamrani, S.; Berry, M. B.; Phillips, G. N.; McCammon, J. A. *Proteins: Struct., Funct., Genet.* **1996**, *25*, 79.
(11) Temiz, N. A.; Meirovitch, E.; Bahar, I. *Proteins* **2004**, *57*, 468.
(12) Gerstein, M.; Lesk, A. M.; Chothia, C. *Biochemistry* **1994**, *33*, 6739.
(13) Kumar, S.; Sham, Y. Y.; Tsai, C.-J.; Nussinov, R. *Biophys. J.* **2001**, *80*, 2439.
(14) Hayward, S. *J. Mol. Biol.* **2004**, *339*, 1001.
(15) Eisenmesser, E. Z.; Bosco, D. A.; Akke, M.; Kern, D. *Science* **2002**, *295*, 1520.
(16) García, A. E. *Phys. Rev. Lett.* **1992**, *68*, 2696.
(17) García, A. E.; Blumenfeld, R.; Hummer, G.; Krumhansl, J. A. *Physica D* **1997**, *107*, 225.
(18) Amadei, A.; Linssen, A. B. M.; Berendsen, H. J. C. *Proteins: Struct., Funct., Genet.* **1993**, *17*, 412.
(19) Cox, T. F.; Cox, M. A. A. *Multidimensional scaling*, 2nd ed.; Chapman & Hall: Boca Raton, FL, 2001.
(20) Romo, T. D.; Clarage, J. B.; Sorensen, D. C.; Phillips, G. N. *Proteins: Struct., Funct., Genet.* **1995**, *22*, 311.
(21) Wlodek, S. T.; Clark, T. W.; Scott, L. R.; McCammon, J. A. *J. Am. Chem. Soc.* **1997**, *119*, 9513.
(22) Creveld, L. D.; Amadei, A.; van Schaik, R. C.; Pepermans, H. A. M.; de Vlieg, J.; Berendsen, H. J. C. *Proteins: Struct., Funct., Genet.* **1998**, *33*, 253.
(23) Arcangeli, C.; Bizzarri, A. R.; Cannistraro, S. *Biophys. Chem.* **2001**, *92*, 183.
(24) Garcia, A. E.; Sanbonmatsu, K. Y. *Proteins: Struct., Funct., Genet.* **2001**, *42*, 345.
(25) Grottesi, A.; Ceruso, M.-A.; Colosimo, A.; DiNola, A. *Proteins: Struct., Funct., Genet.* **2002**, *46*, 287.
(26) Teodoro, M. L.; Phillips, G. N.; Kavraki, L. E. *J. Comput. Biol.* **2003**, *10*, 617.
(27) Case, D. A.; Pearlman, D. A.; Caldwell, J. W.; III, T. E. C.; Wang, J.; Ross, W. S.; Simmerling, C. L.; Darden, T. A.; Merz, K. M.; Stanton, R. V.; Cheng, A. L.; Vincent, J. J.; Crowley, M.; Tsue, V.; Gohlke, H.; Radmer, R.; Duan, Y.; Pitera, J.; Massova, I.; Seibel, G. L.; Singh, C.; Weiner, P.; Kollman, P. A. AMBER7; University of California: San Francisco, 2002.
(28) 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**, *118*, 2309.
(29) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, G. L. *J. Chem. Phys.* **1995**, *103*, 8577.
(30) Müller, C. W.; Schlauderer, G. J.; Reinstein, J.; Schulz, G. E. *Structure* **1996**, *4*, 147.
(31) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327.
(32) Berendsen, H. H. C.; Postma, J. P. M.; Gunsteren, W. F.; DiNola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684.
(33) Elmaci, N.; Berry, R. S. *J. Chem. Phys.* **1999**, *110*, 10606.
(34) Amadei, A.; Ceruso, M. A.; Di Nola, A. *Proteins: Struct., Funct., Genet.* **1999**, *36*, 419.
(35) Hess, B. *Phys. Rev. E* **2000**, *62*, 8438.
(36) Hess, B. *Phys. Rev. E* **2002**, *65*, 031910.
(37) McCammon, A.; Harvey, S. C. *Dynamics of proteins and nucleic acids*; Cambridge University Press: Cambridge, 1987.
(38) Hunenberger, P. H.; Mark, A. E.; Vangunsteren, W. F. *J. Mol. Biol.* **1995**, *252*, 492.
(39) Gerstein, M.; Schulz, G.; Chothia, C. *J. Mol. Biol.* **1993**, *229*, 494.
(40) de Bakker, P. I. W.; Hunenberger, P. H.; McCammon, J. A. *J. Mol. Biol.* **1999**, *285*, 1811.
(41) Müller, C. W.; Schulz, G. E. *J. Mol. Biol.* **1992**, *224*, 159.
(42) Kirkwood, J. G. *J. Chem. Phys.* **1935**, *3*, 300.