Protein Science (1996), *5*:62–71. Cambridge University Press. Printed in the USA.

Simulations of CRP:$(cAMP)_2$ in noncrystalline environments show a subunit transition from the *open* to the *closed* conformation

ANGEL E. GARCÍA[1] AND JAMES G. HARMAN[2]

[1] Theoretical Biology and Biophysics, Los Alamos National Laboratory, Los Alamos, New Mexico 87545
[2] Department of Chemistry and Biochemistry, Texas Tech University, Lubbock, Texas 79409-1061

(RECEIVED June 20, 1995; ACCEPTED October 26, 1995)

Abstract

The CRP:cAMP complex functions as a transcription factor that facilitates RNA polymerase recognition of several bacterial promoters. Detailed crystal structure information is available for CRP:$(cAMP)_2$ and for CRP:$(cAMP)_2$ complexed with DNA. In the crystalline environment, CRP:$(cAMP)_2$ subunits are asymmetrically related; one subunit has a *closed* conformation and the other has an *open* conformation. The CRP:$(cAMP)_2$ complexed with DNA shows both subunits in a *closed* conformation. We have studied the molecular dynamics of CRP:$(cAMP)_2$ in noncrystalline environments. CRP:$(cAMP)_2$ was simulated for 625 ps in vacuo and for 140 ps in solution. The crystal structure of CRP:$(cAMP)_2$ in the absence of DNA was used as the initial conformation. Molecule optimal dynamic coordinates (MODCs) (García A, 1992, *Phys Rev Lett 68*:2696) were used to analyze protein conformations sampled during the course of the simulations. Two MODCs define a transition of the *open* subunit to a *closed* subunit conformation during the first 125 ps of simulation in vacuo; the resulting subunit conformation is similar to that observed in CRP:$(cAMP)_2$:DNA crystals. Simulation of CRP:$(cAMP)_2$ in solution showed that a transition from the *open* to the *closed* state also occurs when water is explicitly included in the calculations. These calculations suggest that the asymmetric conformation of CRP:$(cAMP)_2$ is stabilized by crystal lattice interactions. The predicted solution conformation is more symmetric, with both subunits in a *closed* conformation.

Keywords: allosteric interactions; cAMP; CRP; molecular dynamics simulation in solution; molecular dynamics simulation in vacuo; molecule optimal dynamic coordinates; nonlinear dynamics; subunit reorientation

The cyclic 3′:5′-adenosine monophosphate (cAMP) receptor protein (CRP) plays an important role in mediating transcription activation of several genes in enteric bacteria. Much is known of the physiology of CRP expression and CRP-dependent gene expression and of the biochemistry of CRP-dependent gene expression (Botsford & Harman, 1992); however, details of the mechanism by which cAMP effects allosteric control over CRP activity remain unclear.

The CRP:cAMP complex binds to specific sequences in several *E. coli* promoter regions (Botsford & Harman, 1992; Heyduk, 1989; Kolb et al., 1993). The results of a variety of biochemical analyses show that cAMP binding to CRP produces a structural change in the protein that leads to the activation of CRP as a transcription factor (extensively reviewed by Weber & Steitz, 1987; Botsford & Harman, 1992; Kolb et al., 1993). The activation process is complex, involving not only the binding of cAMP (Weber & Steitz, 1987; Botsford & Harman, 1992; Kolb et al., 1993), but also the interaction of CRP with DNA (Schultz et al., 1991; Gunasekera et al., 1992) and with RNA polymerase (Ebright, 1993; Kolb et al., 1993; Busby & Ebright, 1994).

The primary structure of CRP is known (Aiba et al., 1982; Cossart & Giquel-Sanzey, 1982). Detailed secondary and tertiary structure of CRP:$(cAMP)_2$ were determined from X-ray analysis of CRP crystals grown in the presence of cAMP (Weber & Steitz, 1987). CRP is a dimer of two identical 209-amino acid subunits (47,238 Da). Each subunit can bind one molecule of cAMP (Botsford & Harman, 1992). Each subunit folds into two domains; the N-proximal domain contains the cyclic-nucleotide-binding pocket and the C-proximal domain contains the DNA-binding helix-turn-helix motif. The two domains are connected by a hinge region 8–10 amino acids long (Weber & Steitz, 1987). Conformational differences between the two hinge regions result in structural asymmetry of the otherwise identical CRP:$(cAMP)_2$ subunits in the crystalline environment; one subunit has a *closed* conformation and the other has an *open* conformation.

A CRP:$(cAMP)_2$:DNA cocrystal has been analyzed (Schultz et al., 1991). In this environment, CRP subunit asymmetry is not observed; instead, both subunits have a *closed* conforma-

Reprint requests to: Angel E. García, Theoretical Biology and Biophysics, Group T10, Mail Stop K710, Los Alamos National Laboratory, Los Alamos, New Mexico 87545; e-mail: angel@agua.lanl.gov.

tion. It has been postulated that formation of the CRP:DNA complex results in a change in CRP subunit conformation from the asymmetric *open-closed* conformation to the symmetric *closed-closed* conformation (Kolb et al., 1993; Spolar & Record, 1994). Alternatively, differences in crystal lattice interactions could produce differences in CRP:(cAMP)$_2$ subunit conformation (Schultz et al., 1991; Kolb et al., 1993). Studies of the conformation and dynamics of CRP:(cAMP)$_2$ in a noncrystalline environment will help elucidate the influence of crystal lattice interactions on CRP conformation.

Molecular dynamics (MD) simulations of biomolecules, when combined with experimental techniques, have proved useful in studying the roles of specific amino acid residues in determining the stability and specificity of biomolecules (Brooks et al., 1988; Levitt, 1989). Development of reliable force fields and software during the last 15 years has made it possible to simulate large biomolecules in solution over extended times (Weiner et al., 1984, 1986). These MD studies have the advantages that many nonlinearities and the complicated energy landscape of the system may be explicitly included (García, 1992).

In describing the dynamics of proteins, we must consider the wealth of dynamics phenomena present in complex systems. The dynamics of a protein can be classified into three types (Frauenfelder et al., 1991): unimodal, multimodal fluctuating, and relaxation.

1. Unimodal dynamics describes small fluctuations on a single local minimum; these are easily explained by harmonic and quasi-harmonic dynamics (Karplus & Kushik, 1981; Brooks et al., 1988).

2. Fluctuation dynamics describes motions close to equilibrium that involve many local minima that are close in configuration space and energy. The energy barriers separating these states are of the same order of magnitude as KT, where K is the Boltzmann constant and T is the temperature in kelvins. Harmonic and quasi-harmonic dynamics fail to properly describe these motions because the system behaves nonlinearly.

3. Relaxation dynamics, which is observed in proteins that are initially metastable, has the characteristics of fluctuation dynamics but also exhibits a drift from the initial metastable state to a set of states close to equilibrium.

A characteristic of nonlinear dynamics is that the protein remains within a limited region of configuration space for a long time (relative to the characteristic vibrational periods of the system), then migrates out of this region, diffuses for a short time, and finds another local minimum (García, 1992, 1995). Relaxation dynamics shows a drift in configuration space from an initial metastable state to a set of states that are within a few KT of an equilibrium state. After these states are sampled, the dynamics of the system is characteristic of fluctuation dynamics. In the present study, relaxation dynamics is shown to play an important role in describing the transition of the CRP:(cAMP)$_2$ dimer from an asymmetric *open-closed* state to a *closed-closed* state.

Studies of the dynamics and relaxation of proteins have shown that transitions among conformational states are important in protein function (Frauenfelder et al., 1991). Molecule optimal dynamic coordinates (MODCs) exhibiting nonlinear motion have been shown to describe most of the Cartesian coordinate fluctuations of protein atoms during MD simulation (García, 1992). It is our hypothesis that long-term, nonlocal (i.e., involving all regions of the protein), large-scale structural changes are related to functionally important motions in a protein. Motion along these modes involves low energy barriers and large structural variations. Allosteric changes are the best known example of such motion.

Here we report the results of MD simulations of CRP:(cAMP)$_2$ (see Kinemages 1, 2, and 3). Our goal in this study was to determine the flexibility and stability of the CRP:(cAMP)$_2$ crystal structure in a noncrystalline environment, paying particular attention to the flexibility of the protein domains that play an important role in the binding of CRP to DNA (the E-helix-turn-F-helix) and in the interaction of CRP with RNA polymerase (the 156–162 loop) (Weber & Steitz, 1987; Schultz et al., 1991; Zhou et al., 1993). Motions that involve the relative conformations of the CRP subunit domains relate to the transition from the asymmetric *open-closed* dimer structure observed in CRP:(cAMP)$_2$ crystals (Weber & Steitz, 1987) to a *closed-closed* dimer structure, similar to that observed in CRP:(cAMP)$_2$:DNA crystals (Schultz et al., 1991).

Results

Figure 1 shows the projection of the MD trajectory of CRP:(cAMP)$_2$ along the five principal MODCs. MODCs 1–5, respectively, describe 33%, 30%, 7.8%, 4.6%, and 3.9% of the

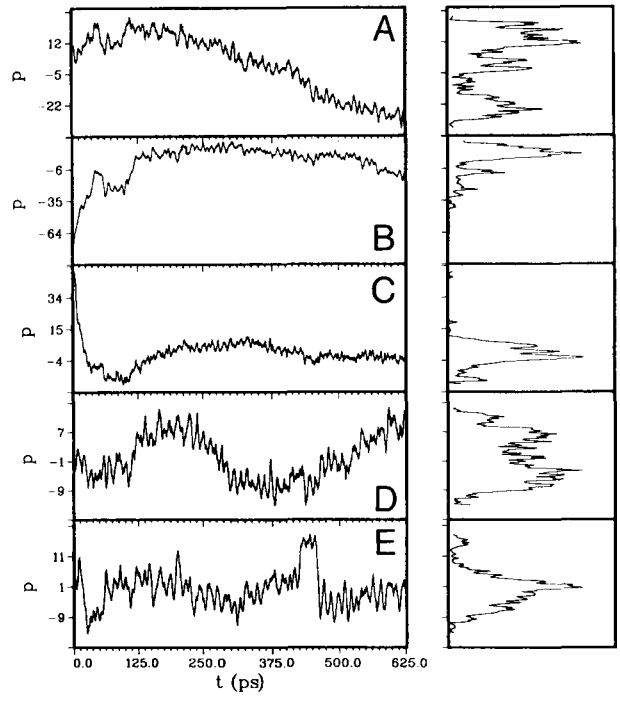
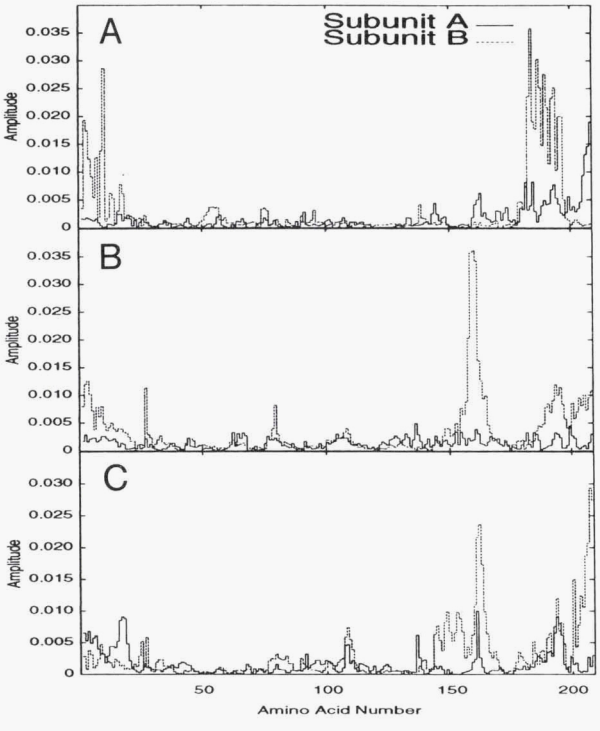


Fig. 1. Projection (p_α) of the CRP:(cAMP)$_2$ MD trajectories, in Å, as a function of time, in ps, along MODCs 1–5 (labeled A–E, from top to bottom), which best describe the fluctuations from the average conformations in the system. Eigenvalues were: MODC 1, 277.9 Å^2; MODC 2, 250.7 Å^2; MODC 3, 66.13 Å^2; MODC 4, 38.8 Å^2; MODC 5, 32.88 Å^2. The mean square fluctuation along each MODC is proportional to its eigenvalue. MODCs 1–5 describe 33%, 30%, 7.8%, 4.6%, and 3.9%, respectively, of the displacements of the Cα atoms. Panels on the right show histograms of the occurrence of projection values along the trajectory. Note that the projection-value distributions for each mode are neither unimodal nor Gaussian.



mean square displacements (MSDs) (McLachalan, 1979) of the Cα atoms, representing 80% of the Cartesian coordinate fluctuations of the Cα atoms; an additional 1,219 vectors describe the remaining 20% of these fluctuations. The right-hand panels in Figure 1 display histograms of the occurrence probabilities of different amplitudes for each of these five MODCs. MODCs 1–5 show multimodal distributions that represent oscillations around more than one equilibrium position; transitions from one minimum well to another occur often. In contrast, vectors 6–1,224 exhibit unimodal distributions that represent atomic fluctuations that correspond to harmonic or quasi-harmonic motions around a single equilibrium point.

The MSDs along the three principal MODCs of every Cα atom in CRP:$(cAMP)_2$ are shown in Figure 2. The motions described by each of these MODCs are nonlocal, representing the collective motion of groups of atoms throughout the protein. Figure 3 and Kinemage 1 show three superimposed conformations (Cα atoms only) of the structures along MODC 1. MODC 1, with an approximate 500-ps oscillation time, describes the concerted motion of the A helix of each protein subunit and the F helix of subunit B. This motion is responsible for most of the protein fluctuations during the 625 MD trajectory, but MODCs 2 and 3 are responsible for most of the fluctuations occurring during the first 125 ps.

MODCs 2 and 3 of Figure 1 define a drift in the protein conformation away from the initial metastable state during the first 125 ps. This drift can be explained in terms of the differences

Fig. 2. Normalized MSDs of the Cα atoms, as a function of the amino acid number, in CRP:$(cAMP)_2$ along the three principal MODCs that best describe the fluctuations of the system from the average conformation. **A:** MODC 1. **B:** MODC 2. **C:** MODC 3. Solid lines show the MSD for atoms in one monomer (subunit A); dashed lines show the MSD for the other monomer (subunit B). **D:** The secondary structure of a CRP monomer consists of six helices and 12 β strands, which contain the following amino acids: 9–18, helix A; 99–110, helix B; 112–133, helix C; 140–151, helix D; 168–176, helix E; 180–191, helix F; 19–23, β 1; 26–33, β 2; 34–43, β 3; 46–52, β 4; 58–66, β 5; 68–70, β 6; 79–89, β 7; 90–99, β 8; 157–160, β 9; 161–165, β 10; 195–199, β 11; and 201–205, β 12 (Weber & Steitz, 1987). Amino acids 156–162 constitute the CRP activation region that interacts with the RNA polymerase α-subunit (Busby & Ebright, 1994). This region exhibits large amplitudes along all three MODCs for the *open* subunit.

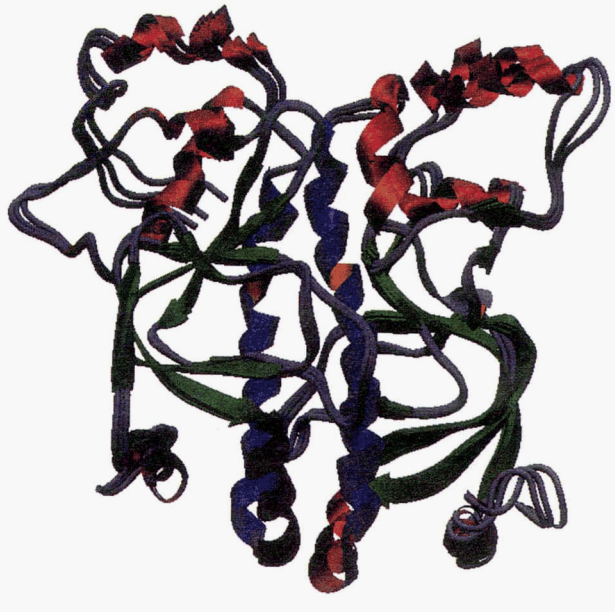


Fig. 3. Three superimposed conformations (Cα atoms only) of CRP:$(cAMP)_2$ along MODC 1. These three structures are obtained from the time-averaged structure by displacing by 12.0, 0.0, and −22.0 Å along MODC 1. These structures correspond to the configurations sampled near 50, 375, and 525 ps along the MD trajectory. This MODC oscillates with a periodicity greater than 500 ps (see Fig. 1) and describes a concerted motion involving the N terminus and A helix of each subunit and the F helix of subunit B. Helices A, B, D, E, and F are in red; helices C are in blue. Amino acid 124 (near the cAMP binding pocket) in the C helices are yellow. Beta strands are green. Coil and turn regions are gray. Amino acid secondary structures illustrated in Figure 2B are assigned according to Weber and Steitz (1987).

in potential energy between the two conformations. Figure 4A shows time-block-averages of the total potential energy (Weiner et al., 1984). Energies were averaged over 5-ps blocks. Vertical lines indicate 1 SD from the average for each 5-ps block. The changes in block-averaged potential energy as a function of time follow the changes in conformation described by MODC 2 in Figure 1. Figure 4B and C shows the nonbond electrostatic and the nonbond van der Waals energies, respectively, along the 625-ps MD simulation. The difference in block-averaged energies at $t = 5$ ps (representing the *open–closed* conformation) and $t = 150$ ps (representing the *closed–closed* conformation) were −491, −325, −152, and −52 kcal/mol for total, electrostatic, van der Waals, and torsional energies (not shown), respectively. The *closed–closed* state adopted by the protein after the first 125 ps is energetically favored over the asymmetric *closed–open* conformation. Other terms contributing to the total potential energy did not show differences greater than 1 SD (over a 5-ps block average). The average energies presented here correspond to differences in enthalpy between the two structures. The average energies obtained from a single MD simulation do not provide enough information to determine the protein energy surface; however, our calculations show that, for CRP:(cAMP)$_2$ isolated from the crystal lattice, the *open–closed* conformation is metastable.

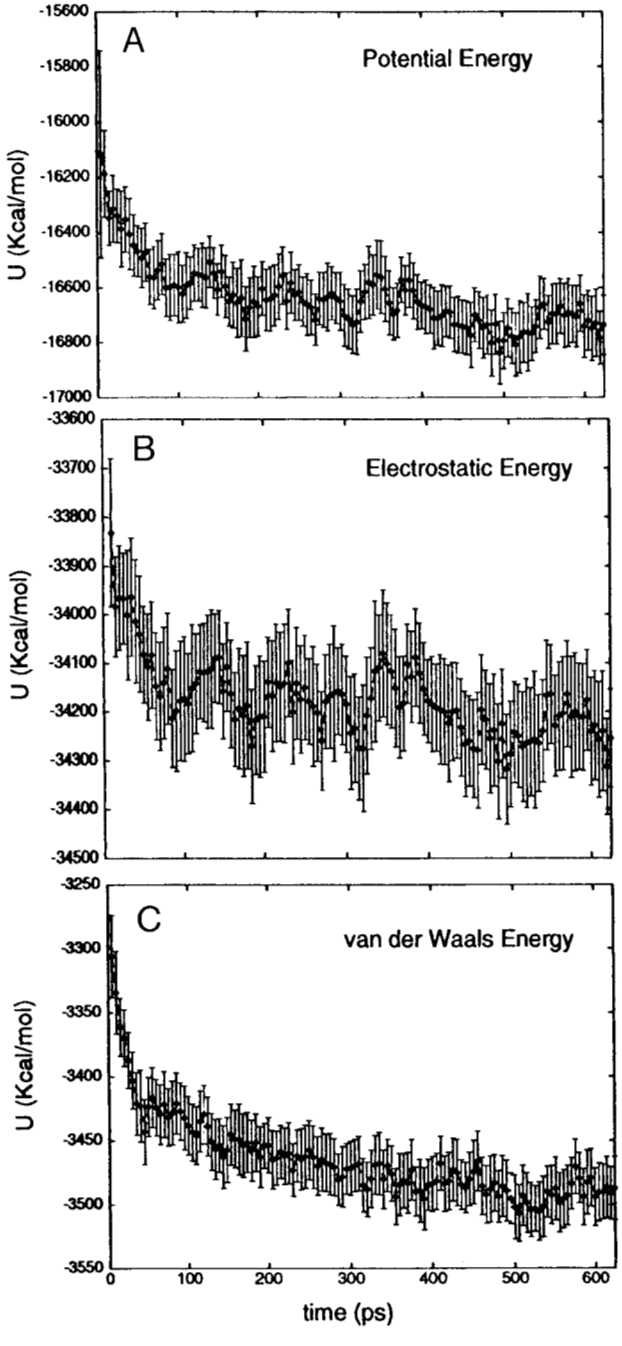


Fig. 4. Time-block-averaged energies of (**A**) the total potential energy (Weiner et al., 1984); (**B**) the nonbond electrostatic energy; and (**C**) the nonbond van der Waals energy for the 625-ps MD simulation. Energies were averaged over 5-ps blocks. Vertical lines indicate 1 SD for the average for each time block.

The protein approached a noncrystalline equilibrium structure after 125 ps of MD simulation. Figure 5 and Kinemage 2 show two conformations obtained by displacing atoms along a linear combination of MODCs 2 and 3. The structure on the left represents a conformation near the initial state (0 ps $< t <$ 125 ps), and the structure on the right is representative of the conformations adopted during the last 500 ps. The drift defined by MODCs 2 and 3 results in the transition of CRP subunit B from an *open* state to a *closed* state. Figure 6 shows stereo plots of the CRP:(cAMP)$_2$ crystal structure (GAP3), the average conformation sampled during the 625-ps MD simulations, and the CRP dimer in the CRP:(cAMP)$_2$:DNA complex (CGP1; see Kinemage 3). Figure 6 clearly shows that the average MD conformation adopts a *closed–closed* conformation similar to that of the CRP:(cAMP)$_2$:DNA complex.

Figure 7 shows the projection of the MD trajectory of CRP:(cAMP)$_2$ in solution along the five principal MODCs. MODCs 1–5, respectively, describe 47%, 16%, 9%, 5%, and 4% of the MSDs of the Cα atoms, representing 81% of the Cartesian coordinate fluctuations of the Cα atoms. The right-hand panels display histograms of the occurrence probabilities of different amplitudes for each of these five MODCs. Here, MODC 1 defines a drift in the protein conformation away from the initial X-ray structure during this 140-ps trajectory. This result reproduces the observation made for the system modeled in vacuo and shows that the conformation of CRP:(cAMP)$_2$ in an aqueous noncrystalline environment is similar to the *closed–closed* conformation observed in CRP:(cAMP)$_2$:DNA crystals; association of DNA with CRP:(cAMP)$_2$ is not required to induce this transition. These results strongly suggest that the subunit conformation of the *open–closed* dimer is stabilized by crystal lattice forces (Shultz et al., 1991; Kolb et al., 1993).

Projections of the MD trajectories along higher dimensional space show that the MODCs of CRP:(cAMP)$_2$ were not independent, but rather were strongly correlated. Figure 8 shows the projection of the MD trajectory on a three-dimensional space and on three two-dimensional planes defined by MODCs (1,2), (1,3), and (2,3). Sampling of conformational space shows that the system underwent motions characteristic of relaxation dynamics during the first 125 ps of simulation. The system drifted from the initial metastable conformation to a set of conformations near equilibrium. The remainder of the trajectory is representative of the fluctuation dynamics that results in the sampling of multiple local minima. Five distinct local minima were sampled by the protein during the course of the MD simu-

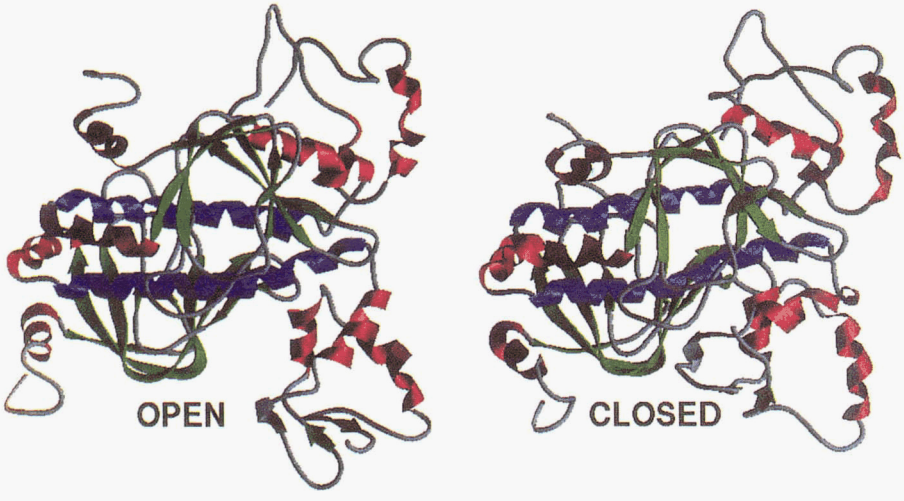


Fig. 5. Two conformations obtained by displacing the Cα atoms along a linear combination of MODCs 2 and 3. The left-hand structure represents a conformation near the initial state (0 ps < t < 125 ps). The right-hand structure represents a conformation adopted during the last 500 ps. The drift described by MODCs 2 and 3 results in a transition of subunit B from an *open* state to a *closed* state. Color coding is as in Figure 3.

lation. Transitions from one minimum-well to another occurred in periods of the order of 150 ps. Representative structures around each basin were obtained at times 0, 175, 275, 425, and 575 ps. Variations among these conformations are largely represented by displacements along the three principal MODCs described above. The analysis presented here shows that the dynamics of the protein cannot be described in terms of quasi-harmonic dynamics around a single equilibrium point; the protein sampled multiple local equilibrium points. Transitions from one equilibrium point to another occurred rapidly, but only after residence periods of approximately 100 ps. These transitions involved the coherent motion of many structural domains in the protein, as shown in Figures 3 and 5.

Overall conformational changes in CRP:$(cAMP)_2$

Figure 9A shows the time evolution of the RMS distance (McLachalan, 1979), in Å, between the crystal structure of CRP:$(cAMP)_2$ and all conformations sampled during the in vacuo MD simulation. The averaged sampled conformation for CRP:$(cAMP)_2$ is 3.5 Å from the crystal structure. This large deviation from the crystal structure is the result of a slow (~125 ps) but steady drift in the structure from an asymmetric *open–closed* conformation to a *closed–closed* conformation, as explained above. The RMS variation (2 Å) during the last 500 ps of simulation was smaller than for the first 125 ps, indicating that a new local equilibrium state has been reached. This new equilibrium is characterized by a *closed–closed* state. However, within this state there are transitions between energy minima. These motions are characteristic of fluctuation dynamics, as described in the Introduction. Given the fact that the protein samples multiple minima (as shown in Fig. 8), we can expect the RMS distance between two conformations sampled at two different times, $d(t,t')$ (shown in Fig. 9B), to exhibit different behaviors depending on the difference in time, $t^* = |t - t'|$. For short t^*, $d(t^*)$ will increase exponentially to a value determined by the width of the energy minimum and the temperature. As t^* approaches the mean time between transitions (approximately 100 ps), $d(t^*)$ will increase to a new value that is characteristic of the RMS distance between neighboring local energy minima. The RMS distance curve labeled "last 500 ps" in Figure 9 shows jumps at $t \sim$ 275 ps, $t \sim$ 425 ps, and $t \sim$ 575 ps. The RMS distance curve for the entire in-vacuo simulation does not show these jumps because the distance between the initial state and the t = 150 ps conformation is larger than the characteristic distance between minima sampled during the last 500 ps. This is clearly shown in Figure 8A. For large t^*, $d(t^*)$ reached a value characteristic of the total fluctuation of the system (i.e., the available configuration space at a given temperature). Our 625-ps simulation did not show the large t* behavior because the simulation time was too short. Experiments on myoglobin suggest that this time is as large as seconds (Frauenfelder et al., 1991).

The simulation of CRP in solution exhibited a drift in the RMS distance similar to that observed in the in-vacuo simulation. Figure 10A shows the average MSD fluctuations of the Cα atoms. For most of the cAMP-binding domain of the protein (amino acid residues 25–141), the structure was reasonably conserved, with an MSD = 1.75 Å^2. The DNA-binding domain showed large conformational variations. Notice that the fluctuations described by MODCs 1, 2, and 3 describe large MSD variations for this domain and the A helix of the cAMP binding domain. The A helix of one CRP subunit is in close contact with the boundary between the two domains of the same CRP subunit in another molecule in the crystal. Figure 10B shows the crystal *B* factors for the Cα atoms (Weber & Steitz, 1987). For the region of each subunit that contains amino acids 25–141, there is a good correlation between the *B* factors and the MSDs. This correlation is absent for residues 183–200 (F helix) in the *closed* subunit. The large RMS displacement of the F helix during the dynamics trajectory results primarily from rigid-body displacement of the D-E-F helices relative to the rest of the protein and, to a large degree, are described by MODC 1. The MSDs obtained from the MD simulation in solution correlate better with the crystal *B* factors (Weber & Steitz, 1987) than the corresponding values for MD simulation in vacuo. The large MSD values observed for the D-E-F helices in vacuo were not present in solution; however, the characteristic duration of the motion described by MODC 1 in Figure 1 is on the order of 500 ps. A 140-ps simulation in solution does not contain enough information to allow us to extract this mode from the trajectory, if indeed it is present.

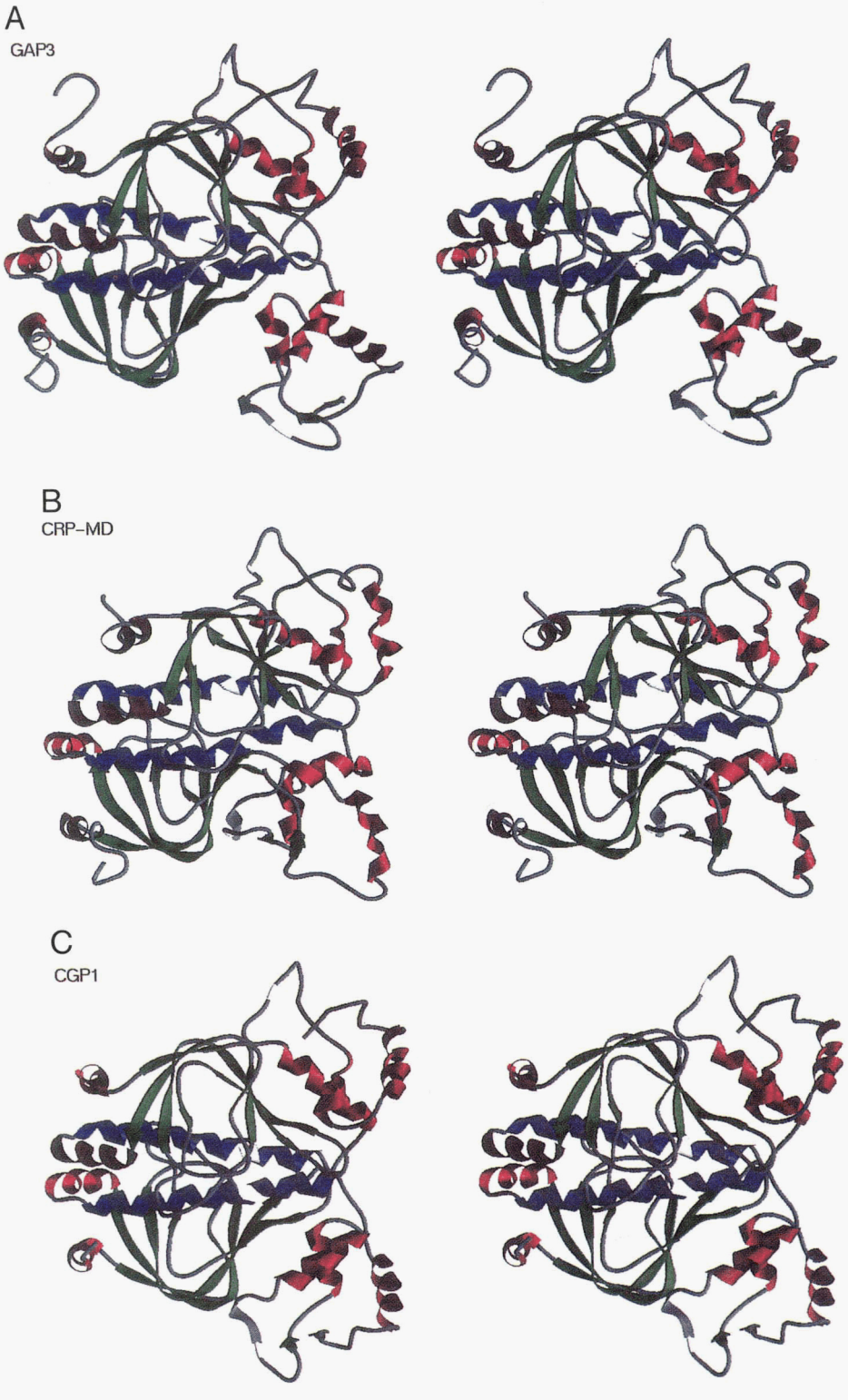


Fig. 6. Stereo plots showing (**A**) the crystal structure of CRP:$(cAMP)_2$ (GAP3), (**B**) the average structure sampled during the 625-ps MD simulation, and (**C**) the crystal structure of the CRP dimer in the CRP:$(cAMP)_2$:DNA complex (CGP1).

Discussion

The simulation of CRP:$(cAMP)_2$ presented here illustrates that current simulation methods, force fields, and advanced simulation analysis techniques yield descriptions of CRP structure and dynamics that are in good agreement with available experimental data. Simulations of an all-atom model of this system in aqueous solution provide a view of the dynamics and stability of CRP:$(cAMP)_2$ similar to that obtained from the simulation in vacuo. A drift in the state of the protein from an initial *open–closed* metastable state to a more stable *closed–closed* state occurs within the first 125 ps of the MD trajectory. Fluctuations

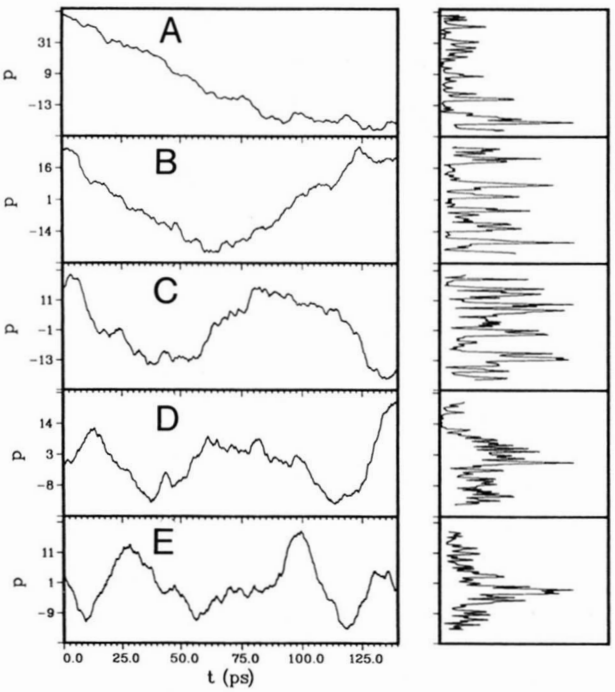


Fig. 7. Projection of the CRP:$(cAMP)_2$ solution MD trajectories along the five MODCs that best describe the fluctuations from the average conformations in the system. Eigenvalues were as follows: MODC 1, 647.0 Å^2; MODC 2, 223.0 Å^2; MODC 3, 128.2 Å^2; MODC 4, 69.2 Å^2; MODC 5, 56.1 Å^2. The MSD along each MODC is proportional to this eigenvalue ($MSD_i = \lambda_i/N$, where λ_i is the eigenvalue of the ith MODC, labeled A–E, and N is the number of amino acids in the CRP dimer). MODCs 1–5 describe 47%, 16%, 9%, 5%, and 4%, respectively, of the displacements of the Cα atoms. The right-hand panels show histograms of the occurrence of projection values along the trajectory. The statistics of this 140-ps simulation do not allow for a clear definition of these histograms as either unimodal or multimodal. The top plot describes a drift of the conformation of CRP:$(cAMP)_2$ from an asymmetric *open–closed* conformation to a *closed–closed* conformation. A similar drift was described by MODCs 2 and 3 in Figure 1.

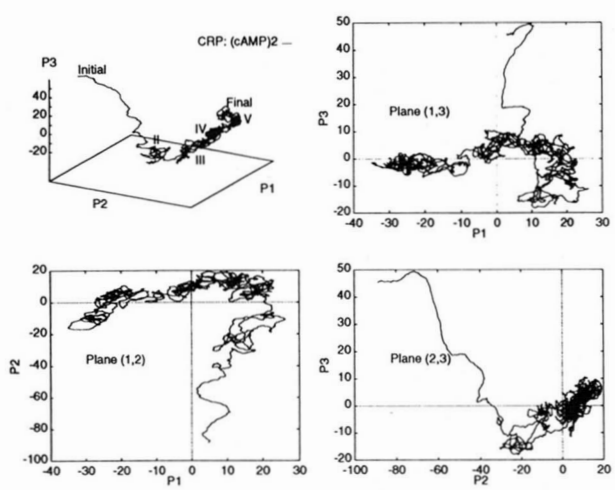


Fig. 8. Projection of the MD trajectories, in Å, in three-dimensional space and in three two-dimensional planes defined by MODCs (1,2), (1,3), and (2,3). Five local minima were sampled by the protein during the course of the 625-ps MD simulation. Transitions from one minimum to another occurred in time periods of the order of 100 ps.

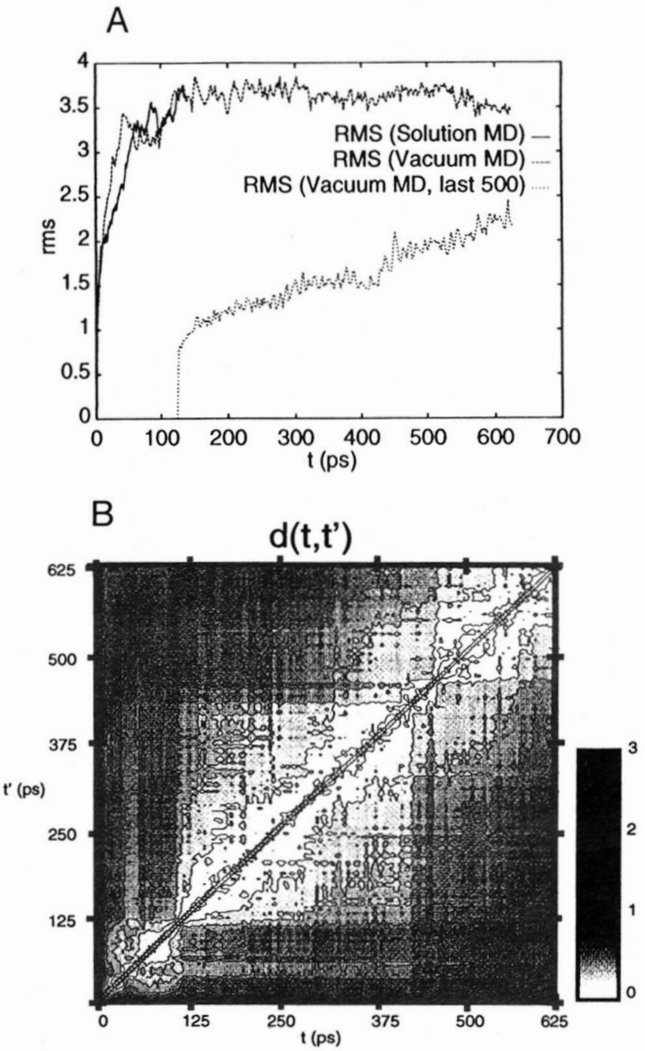


Fig. 9. A: Evolution of the RMS distance (McLachalan, 1979), in Å, between the crystal structure of CRP:$(cAMP)_2$ and all conformations sampled during the in vacuo MD simulation (solid line), and the MD simulation in solution (dashed line). The RMS distance between the conformation after 125 ps in the in-vacuo simulation (i.e., the *closed–closed* conformation) and all subsequent conformations is shown by the dotted line. The RMS variation (2 Å) during the last 500 ps of simulation is smaller than for the first 125 ps, suggesting that a new equilibrium conformation is reached. **B:** Contour plot of the distance matrix, $d(t,t')$, between all pairs of conformations adopted at times t and t' along the 625-ps MD simulation.

on a longer time scale reveal that the protein samples multiple minima. Transitions between minima are rapid but occur only after residence periods of 100 ps.

A comparison of the RMS displacement of Cα atoms for each of the two monomers with the crystal B factors shows that these displacements are correlated for the amino acids 25–141 but not for the A helix or the DNA-binding domain. The large RMS displacements of the F helix during the dynamics trajectory results primarily from rigid-body displacements of the D-E-F helices relative to the rest of the protein and, to a large degree, are described by a single MODC, MODC 1. Binding of CRP to DNA will damp the rigid-body displacements of the D-E-F helices, be-

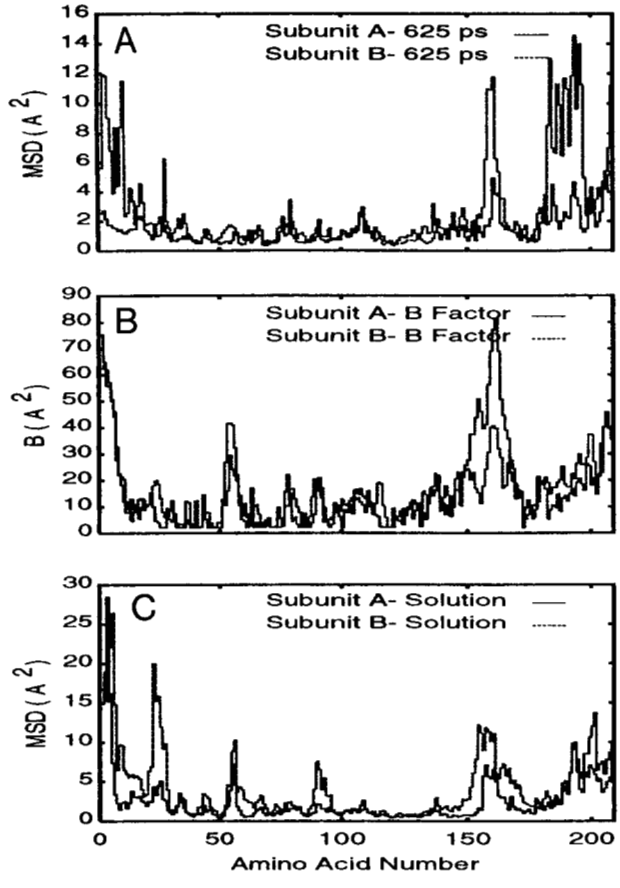


Fig. 10. A: Average MSDs for the Cα atoms of the CRP:(cAMP)$_2$ dimer. The mean square fluctuations for most of the cAMP-binding domain (residues 25–141) of the dimer are small. The structure of this region is reasonably conserved and shows a MSD of 1.75 Å^2. **B:** *B* factors for the Cα atoms reported by Weber and Steitz (1987). Compared with A and C, there is a good correlation between large *B* factors and large MSDs. The large MSDs of the F helix during the dynamics trajectory result primarily from rigid-body displacement of the D-E-F helices relative to the cAMP-binding domain. This motion is described to a large degree by MODC 1, as shown in Figures 1 and 3. **C:** Average MSDs for the Cα atoms of the CRP:(cAMP)$_2$ dimer in solution. These MSDs are better correlated with the *B* factors shown in B than the corresponding values from MD simulation in vacuo. The large MSD values observed in A for the D-E-F helices are smaller in solution; however, the characteristic period of the motion described by MODC 1 is ~500 ps. Amino acids 156–164, which contain the RNA polymerase-binding site (see Fig. 2), show large MSDs and *B* factors.

cause a large DNA surface will further bridge the DNA-binding domain of each CRP monomer. Spolar and Record (1994) suggested that key regions of the CRP-binding interface, helices D-E-F, are disordered in solution, even though these regions are folded in the crystal. This disorder is described in part by the MODC 1 of this study (Fig. 2A).

The MODC analysis presented here is capable of quantitatively representing global motions of large structural subdomains of the protein complex. We have shown that two MODCs describe the transition of one subunit of CRP from an *open* to a *closed* conformation. Examples of both the *open-closed* and the *closed-closed* subunit conformations exist for the CRP:(cAMP)$_2$ dimer in a crystalline environment. We have shown here that the subunit conformation of CRP:(cAMP)$_2$ in noncrystalline environments is the *closed-closed* conformation, similar to that observed in CRP:(cAMP)$_2$:DNA crystals. The presence of DNA is not required to induce this transition in CRP:(cAMP)$_2$. We interpret our results as evidence that the subunit asymmetry observed in CRP:(cAMP)$_2$ crystal structure is stabilized by crystal lattice interactions (Schultz et al., 1991; Kolb et al., 1993). Our results illustrate that CRP is a highly flexible molecule. Transition from a *open-closed* to a *closed-closed* state occurs rapidly. This flexibility may be important for allosteric changes upon binding of cAMP and for DNA binding, as suggested by Spolar and Record (1994).

Methods

Simulation of CRP:(cAMP)$_2$ in vacuo

Because of the large size of the system and the enormous computational demand of MD simulations of large systems in solution, we initially studied this system in its simplest form, neglecting solvent interactions. The atomic coordinates of CRP:(cAMP)$_2$ were obtained from the reported crystal structure coordinates (Weber & Steitz, 1987). The C-terminal residue R209 of the *closed* subunit (subunit A), and amino acids 206–209 of the *open* subunit (subunit B) were not resolved in the diffraction data. These regions were modeled using the coordinates of the *open* subunit when available, and by using an extended conformation for R209. The N- and C-terminal amino acids were modeled in their charged states. All R, K, and E residues were modeled in their fully charged states. The charge states of H residues were modeled according to the protonation states described by Clore and Gronenborn (1982); H19A and H19B were fully protonated (i.e., in the +1 state), whereas all other H residues were δ-protonated only. The total charge of the system, including two cAMP anions, was +2e. Bonding and van der Waals interactions were modeled using the united atom force field of Weiner et al. (1984). Only the hydrogen atoms of polar and charged groups were explicitly included. The force field parameters for cAMP were taken to be those of nucleic acids. Nonbonding interactions involving atoms separated by one or two bonds were not considered, whereas those for atoms separated by three bonds were scaled to 0.5 of their value. Electrostatic interactions between amino acid residues closer than 10.5 Å were explicitly included using a dielectric constant of 1.0. In-vacuo calculations represent gross models of a system and, considering the objectives of our calculations, a dielectric constant of 1.0 was used, in as much as it allowed these results to be compared with the results of solution simulations. A dielectric constant of 1 has been used in modeling proteins (Jorgensen & Tirado-Rives, 1988), although values of 2 or 4 could have been equally acceptable. The agreement between the in-vacuo and solution simulations presented here supports our choice of the simplest model for the dielectric constant. The system was simulated at a constant temperature of 300 K (Berendsen et al., 1984) for 625 ps, with an integration step of 0.001 ps. All bonds involving hydrogen atoms were constrained to a fixed length using SHAKE (van Gunsteren & Berendsen, 1977). All simulations were done using AMBER 4.0 (Pearlman et al., 1991). CRP:(cAMP)$_2$ consists of 418 amino acids and 2 cAMP molecules totaling 4,156 atoms. Amino acids 1–209 correspond to subunit A, and amino acids 210–418 correspond to subunit B. The initial system configura-

tion was energy minimized for 2,000 cycles of steepest descent and conjugate gradient energy minimization. Random (Gaussian-distributed) velocities were assigned to the energy-minimized structure to simulate an instantaneous temperature of 200 K, which was maintained for 2 ps of MD simulation. Over the following 2 ps of MD simulation, the system was heated to 300 K. Rotational and translational velocities of the center of mass were subtracted from the initial dynamics simulations and at 1-ps intervals throughout the simulation. A temperature of 300 K was maintained throughout the following 625 ps of MD simulation. Configurations were saved at a rate of 20/ps. The time required for the system to reach equilibrium was determined by analysis of the MODC trajectories described in the Results; equilibrium was reached after 125 ps of MD simulation.

Simulation of CRP:(cAMP)$_2$ in solution

To evaluate the results of the MD simulations of CRP:(cAMP)$_2$ in vacuo, we simulated this system in solution. We explicitly included all atoms in the CRP:(cAMP)$_2$ complex surrounded by a free 6-Å water shell. This water shell was formed by including an equilibrated water box in the system. All water molecules less than 2.7 Å or more than 6 Å from the closest protein atom were removed from the system. The resulting system consisted of 6,794 atoms and 2,443 TIP3P (Jorgensen et al., 1983) water molecules; a total of 14,123 atoms. Bonding and nonbonding interactions were modeled using the all-atom force field of Weiner et al. (1986). The simulation parameters used for the in-vacuo calculation were precisely followed for the solution simulation. The system was simulated for 140 ps at 300 K. This duration is similar to that (125 ps) in which large conformation changes were observed in the in-vacuo simulation.

Collective motions of CRP:(cAMP)$_2$

We analyzed the MD trajectories of CRP:(cAMP)$_2$ in terms of the collective motions that best represent the fluctuations of the system. The derivation of methodology that accomplishes this purpose, and its application to describing the dynamics of proteins and DNA molecules, has been presented (García, 1992, 1995; García et al., 1994). Our analysis of the data presented here, based on a generalized least-squares fitting of the protein fluctuations along vectors, reveals that a small set of MODCs describe most of the protein fluctuations during the course of MD simulations. In addition, the MODCs reveal a trajectory that exhibits a migration of the protein conformation from one minimum well to another. A generalization of the MODCs that allows two- and three-dimensional representations of the protein conformational space clearly shows a trajectory around multiple basins of attraction (Clarage et al., 1995; García, 1995).

The MODCs for CRP:(cAMP)$_2$ were obtained from all saved conformations of a trajectory by diagonalizing the resulting covariance matrix of selected dynamic variables. It is known that the conformations and conformational changes of proteins are well represented by the position of the Cα atoms in the protein backbone. Therefore, we performed the MODC analysis on a set of coordinates involving only the 418 Cα atoms of CRP:(cAMP)$_2$. A 1,254 × 1,254 matrix, $\{S_{ij}\}$, was evaluated by evaluating $S_{ij} = \langle (x_i - \langle x_i \rangle)(x_j - \langle x_j \rangle) \rangle$, where the bracket, $\langle\ \rangle$, represents a time average along the MD trajectory, and the variables x_i are the Cartesian components of all Cα atoms in the protein, after alignment (McLachalan, 1979). A set of 1,254 vectors, each with dimensionality equal to that of the matrix, was obtained by diagonalizing the matrix $\{S_{ij}\}$. These vectors are called MODCs. On the surface, the MODCs appear similar to a quasi-harmonic analysis (Karplus & Kushik, 1981; Brooks et al., 1988); however, in the MODC analysis, no assumption of (quasi) harmonicity (i.e., a single basin of attraction) is made, and the frequency spectrum of the system is not calculated. The capacity of each MODC to describe the atomic fluctuations of the protein is proportional to the corresponding eigenvalue; MODCs are ranked systematically according to eigenvalue, and the MODCs with the largest eigenvalues best describe the atomic fluctuations. We denote the projection of the conformation adopted at a given time t on the MODC α, by $p_\alpha(t)$. The best plane is spanned by the two MODCs with the largest eigenvalues. The best volume is spanned by the three MODCs with the largest eigenvalues (García, 1995). Projections of the trajectory in one-, two-, and three-dimensional spaces aid in determining both the equilibration and the character of the trajectory obtained in the MD simulation.

Acknowledgments

We acknowledge the financial support of the Welch Foundation grant D-1248, the Los Alamos National Laboratory LDRD award KAA 5, and the U.S. Department of Energy.

References

Aiba H, Fujimoto S, Ozaki N. 1982. Molecular cloning and sequencing of the gene for *E. coli* cAMP receptor protein. *Nucleic Acids Res 10*:1345–1362.

Berendsen HJC, Postma JPM, van Gunsteren WF, Kaak JR. 1984. Molecular dynamics with coupling to an external bath. *J Chem Phys 81*:3654–3690.

Botsford JL, Harman JG. 1992. cAMP in prokaryotes. *Microbiol Rev 56*: 100–122.

Brooks CL, Karplus M, Montgomery-Pettitt BM. 1988. Proteins: A theoretical perspective of dynamics, structure, and thermodynamics. In: *Advances in chemical physics, vol LXXI*. New York: Wiley.

Busby S, Ebright RH. 1994. Promoter structure, promoter recognition, and transcription activation in prokaryotes. *Cell 79*:743–746.

Clarage JB, Romo T, Andrews BK, Pettitt BM, Phillips GN. 1995. A sampling problem in molecular dynamics simulations of macromolecules. *Proc Natl Acad Sci USA 92*:3285–3292.

Clore GM, Gronenborn AM. 1982. Proton nuclear magnetic resonance study of the histidine residues of the *Escherichia coli* adenosine 3′,5′-phosphate receptor protein. pH titration behavior, deuterium exchange and partial assignments. *Biochemistry 21*:4048–4053.

Cossart P, Gicquel-Sanzey B. 1982. Cloning and sequencing of the *crp* gene of *Escherichia coli* K12. *Nucleic Acids Res 10*:1363–1378.

Ebright RH. 1993. Transcription activation at class I CAP-dependent promoters. *Mol Microbiol 8*:797–802.

Frauenfelder H, Sligar SG, Wolynes PG. 1991. The energy landscapes and motions in proteins. *Science 254*:1598–1601.

García AE. 1992. Non-linear dynamics of proteins. *Phys Rev Lett 68*:2696–2699.

García AE. 1995. Multi-basin dynamics of a protein in aqueous solution. In: Peyrard M, ed. *Nonlinear excitations in biomolecules*. Berlin: Les Editions de Physique, Springer. pp 191–206.

García AE. Soumpasis DM, Jovin TM. 1994. Dynamics and relative stability of parallel- and antiparallel-stranded DNA duplexes. *Biophys J 66*: 1742–1755.

Gunasekera A, Ebright YW, Ebright RH. 1992. DNA sequence determinants for binding of the *Escherichia coli* catabolite gene activator protein. *J Biol Chem 21*:14713–14720.

Heyduk T, Lee JC. 1989. *Escherichia coli* cAMP receptor protein: Evidence

for three protein conformational states with different promoter binding affinities. *Biochemistry 28*:6914–6924.

Jorgensen WL, Chandrasekhar J, Madura JD. 1983. Comparison of simple potential functions for simulating liquid water. *J Chem Phys 79*:926–935.

Jorgensen WL, Tirado-Rives J. 1988. The OPLS potential functions for proteins. Energy minimizations for crystals of cyclic peptides and crambin. *J Am Chem Soc 110*:1657–1666.

Karplus M, Kushick JN. 1981. Method for estimating the configurational entropy of macromolecules. *Macromolecules 14*:325–332.

Kolb A, Busby S, Buc H, Garges S, Adhya S. 1993. Transcriptional control by cAMP and its receptor protein. *Annu Rev Biochem 62*:749–795.

Levitt M. 1989. Molecular dynamics of macromolecules in water. *Chemica Scripta 29A*:197–203.

McLachalan J. 1979. Least square fitting of two structures. Appendix in gene duplication in the structural evolution of chymotrypsin. *J Mol Biol 128*: 49–79.

Pearlman DA, Case DA, Caldwell JC, Siebel GL, Singh UC, Weiner P, Kollman PA. 1991. *AMBER 4.0*. San Francisco: University of California.

Schultz SC, Shields GC, Steitz TA. 1991. Crystal structure of a CAP-DNA complex: The DNA is bent by 90 degrees. *Science 253*:1001–1007.

Spolar RS, Record MT. 1994. Coupling of local folding to site-specific binding of proteins to DNA. *Science 263*:777–784.

van Gunsteren WF, Berendsen HJC. 1977. Algorithms for macromolecular dynamics and constraint dynamics. *Mol Phys 34*:1311–647.

Weber IT, Steitz TA. 1987. Structure of a complex of catabolite gene activator protein and cyclic AMP refined at 2.5 Å resolution. *J Mol Biol 198*:311–326.

Weiner SC, Kollman PA, Case DA, Singh UC, Ghio C, Alagona G, Profeta S, Wiener P. 1984. A new force field for molecular mechanical simulations of nucleic acids and proteins. *J Am Chem Soc 106*:765–784.

Weiner SC, Kollman PA, Nguyen DT, Case DA. 1986. An all atom force field for simulations of proteins and nucleic acids. *J Comput Chem 7*: 230–252.

Zhou Y, Zhang X, Ebrigh RH. 1993. Identification of the activating region of catabolite gene activator protein (CAP): Isolation and characterization of mutants of CAP specifically defective in transcription activation. *Proc Natl Acad Sci USA 90*:6081–6085.