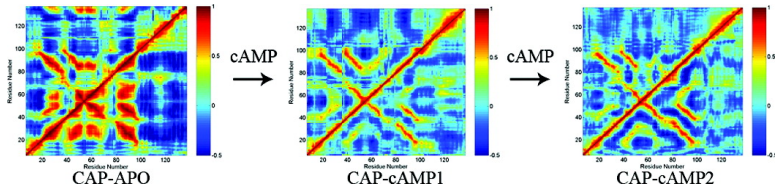
Article

A Computational Investigation of Allostery in the Catabolite Activator Protein

Liwei Li, Vladimir N. Uversky, A. Keith Dunker, and Samy O. Meroueh

J. Am. Chem. Soc., **2007**, 129 (50), 15668-15676 • DOI: 10.1021/ja076046a



More About This Article

Additional resources and features associated with this article are available within the HTML version:

- Supporting Information
- Links to the 1 articles that cite this article, as of the time of this article download
- Access to high resolution figures
- Links to articles and content related to this article
- Copyright permission to reproduce figures and/or text from this article

View the Full Text HTML

J|A|C|S
ARTICLES
Published on Web 11/28/2007

A Computational Investigation of Allostery in the Catabolite Activator Protein

Liwei Li,[†,‡] Vladimir N. Uversky,[†,‡,||] A. Keith Dunker,[†,‡] and Samy O. Meroueh*[,†,‡,§]

Contribution from the Department of Biochemistry and Molecular Biology, Indiana University School of Medicine, Indianapolis, Indiana 46202, Center for Computational Biology and Bioinformatics, Indiana University Purdue University Indianapolis, Indianapolis, Indiana 46202, Department of Chemistry, Indiana University Purdue University Indianapolis, Indiana, 46202, and Institute of Biological Instrumentation, Russian Academy of Sciences, Pushchino 142290, Russia

Received August 10, 2007; E-mail: smeroueh@iupui.edu

Abstract: The catabolite activator protein is a dimer that consists of two cAMP-binding subunits, each containing a C-terminus DNA-binding module and a N-terminus ligand binding domain. The system is well-known to exhibit negative cooperativity, whereby the binding of one cAMP molecule reduces the binding affinity of the other cAMP molecule by 2 orders of magnitude, despite the large separation between the cAMP binding pockets. Here we use extensive explicit-solvent molecular dynamics simulations (135 ns) to investigate the allosteric mechanism of CAP. Six trajectories were carried out for apo, singly liganded, and doubly liganded CAP, both in the presence and absence of DNA. Thorough analyses of the dynamics through the construction of dynamical cross-correlated maps, as well as essential dynamics analyses, indicated that the system experienced a switch in motion as a result of cAMP binding, in accordance with recent NMR experiments carried out on a truncated form of the protein. Analyses of conformer structures collected from the simulations revealed a remarkable event: the DNA-binding module was found to dissociate from the N-terminus ligand binding domain. An interesting aspect of this structural change is that it only occurred in unoccupied subunits, suggesting that the binding of cAMP provides additional stability to the system, consistent with the increase in entropy that was observed in our calculations and from isothermal titration calorimetry. Analysis of the distribution of intrinsic disorder propensities in CAP amino acid sequence using PONDR VLXT and VSL1 predictors revealed that the region connecting ligand-binding and DNA-binding domains of CAP have the potential to exhibit increased flexibility. We complemented these trajectories with free energy calculations following the MM-PBSA approach on more than 2000 snapshots that included 880 normal mode analysis. The resulting free energy differences between the singly liganded and doubly liganded states were in excellent agreement with isothermal titration calorimetry data. When the free energy calculations were carried out in the presence of DNA, we discovered that a switch in cooperativity occurred, so that the binding of the first cAMP promoted the binding of the other cAMP. The components of the free energy reveal that this effect is mainly entropic in nature, whereby the DNA reduces the degree of tightening that is observed in its absence, thereby promoting binding of the second cAMP. This finding prompted us to propose a new mechanism by which CAP triggers the transcription activation that is based on an order to disorder transition mediated by cAMP binding as well as DNA.

Introduction

In systems that exhibit allostery, a triggering event at one site of a macromolecule leads to a corresponding effect at a distal site. This phenomenon was highlighted more than 30 years ago, when it was found that the binding of the first oxygen molecule to the hemoglobin tetramer led to an increase in the affinity for oxygen in the remaining subunits.[1] Since then, numerous macromolecules have been found to display cooperativity between sites with a large separation between them.[2−11] However, a consensus on the mode of communication between

[†] Department of Biochemistry and Molecular Biology, Indiana University School of Medicine.
[‡] Center for Computational Biology and Bioinformatics, Indiana University Purdue University Indianapolis.
[§] Department of Chemistry, Indiana University Purdue University Indianapolis.
[||] Russian Academy of Sciences.

(1) Tyuma, I.; Imai, K.; Shimizu, K. *Biochemistry* **1973**, *12* (8), 1491−8.
(2) Daily, M. D.; Gray, J. J. *Proteins: Struct., Funct., Bioinf.* **2007**, *67* (2), 385−399.
(3) Maler, L.; Blankenship, J.; Rance, M.; Chazin, W. J. *Nat. Struct. Biol.* **2000**, *7*, (3), 245−250.
(4) Lee, A. L.; Kinnear, S. A.; Wand, A. J. *Nat. Struct. Biol.* **2000**, *7*, (1), 72−77.
(5) Volkman, B. F.; Lipson, D.; Wemmer, D. E.; Kern, D. *Science* **2001**, *291* (5512), 2429−2433.
(6) Kern, D.; Zuiderweg, E. R. *Curr. Opin. Struct. Biol.* **2003**, *13* (6), 748−57.
(7) Gunasekaran, K.; Ma, B. Y.; Nussinov, R. *Proteins: Struct., Funct., Bioinf.* **2004**, *57* (3), 433−443.
(8) Ma, J. P.; Karplus, M. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95* (15), 8502−8507.
(9) Harris, S. A.; Gavathiotis, E.; Searle, M. S.; Orozco, M.; Laughton, C. A. *J. Am. Chem. Soc.* **2001**, *123* (50), 12658−12663.

10.1021/ja076046a CCC: $37.00

these sites remains elusive. It has been suggested that cooperativity arises from incremental changes in structure propagated along a channel that connects the sites (induced fit).[12−14] Another model, known as the pre-existing equilibrium or conformational selection, suggests that the triggering effect that leads to allostery *selects* specific conformations from an ensemble of structures, effectively shifting the equilibrium toward the final state.[1,15] Each of these mechanisms assumes a shift in conformation (enthalpic effect). But there is growing evidence that structural change alone cannot explain cooperative effects that are exhibited by some systems.[16]

One example of an allosteric system is the catabolite activator protein (CAP), a transcription activator that associates with DNA in the presence of cyclic adenosine monophosphate (cAMP). CAP controls the transcription for more than 100 genes. In solution, the 47 kDa dimer exists as a homodimer and exhibits homotropic negative cooperativity, whereby the binding of the first cAMP molecule reduces the affinity of the second cAMP by nearly 2 orders of magnitude.[17−19] The basis for this allosteric behavior is not well understood as the crystal structure of CAP in its doubly liganded form reveals the same binding mode for both cAMP molecules. The distance between cAMP binding sites is 24 Å, effectively ruling out electrostatics or bonding interactions as the cause for the negative cooperativity. A recent NMR study carried out on a truncated form of the protein reports that binding of a ligand does not cause conformational change across the dimer interface but alters the dynamics of the system, suggesting that the negative cooperative effect is due to a change in motion.[16]

CAP's function centers on its ability to associate to DNA and initiate transcription activation. But the protein is known to exhibit allostery in the absence of DNA. A pertinent question can be then posed as to whether or not cooperativity is essential for its function. Biochemical analyses have convincingly shown that CAP exists in three distinct conformational states (only the doubly liganded conformation has been structurally characterized). The switching between these states is controlled by the concentration of cAMP: at millimolar, micromolar, and low concentrations of cAMP, CAP exists in its doubly, singly, and unliganded forms, respectively.[17,20] X-ray structures reveal that CAP is always in its doubly liganded form when associated to DNA.[21−23] But biochemical studies have shown that the singly liganded form of CAP binds more favorably to DNA than the doubly liganded form.[17,20] A question then arises as to how CAP has evolved to significantly impair the binding of a second cAMP, when this event is necessary for proper binding to DNA.

To address these questions, and to gain a deeper understanding of the structural and dynamical basis for allostery in CAP, we used a combination of explicit-solvent molecular dynamics simulations and free energy calculations. Three runs of 40 ns each for CAP and 15 ns for the CAP/DNA complex were collected. Three separate simulations were carried out in each of the aforementioned cases, with CAP in its unliganded apo (CAP-APO), singly liganded (CAP-cAMP$_1$), and doubly liganded (CAP-cAMP$_2$) states. Extensive analysis of conformational change and motion was conducted through the computation of root-mean-square (rms) deviation, order parameters, dynamical cross-correlation maps, and essential dynamics. These analyses were supplemented with extensive free energy calculations following the MM-PBSA[24−26] approach on a total of 2200 snapshots, which included 880 normal mode analysis. We also conducted free energy calculations for the binding of cAMP in the presence of DNA. Our data support the results of a recent NMR study that was conducted on a truncated version of CAP, namely that cooperativity was mediated by a switch in motion. However, we also find that minute changes in structure, likely propagated over the long-range, are also responsible for cooperativity. Another significant finding of this work is that the binding of cAMP is highly coupled to the motion of the DNA-binding module, stemming from the observation that the DNA-binding module of unoccupied CAP subunits detached from the main N-terminus ligand-binding domain. The observation that a switch in the cooperativity occurs upon CAP binding to DNA led us to propose a new mechanism by which CAP mediates the initiation of transcription activation. Finally, this work shows that computational tools can be an effective means for predicting allosteric effects mediated by conformational and dynamical changes, which could be valuable in drug design efforts.

Methods

The X-ray crystal structures provided initial coordinates of CAP-cAMP$_2$ (PDB code: 1G6N)[27] and CAP-cAMP$_2$/DNA complex (PDB code: 1ZRC). The apo and singly liganded form of CAP and CAP/DNA complex were obtained by stripping off one and two cAMP's from the crystal structure, respectively. The following protocol for setting up and running the molecular dynamics simulations was followed in all cases. Hydrogen atoms were added to the protein using the Protonate program, which is part of the AMBER 9[28] suite of programs. The AMBER force field parameters were assigned to all atoms using the "parm99" set of parameters. The Sybyl 7.3 program (Tripos Inc., St. Louis, MO) was used for the manipulation and visualization of all structures and for protonation of the bound ligands. The atomic charges of cAMP were determined by using the AM1-BCC methodology[29,30] that is implemented in the Antechamber program, which is part of the AMBER 9 package. The program LEaP was used to neutralize the complexes. The complexes of CAP and CAP/DNA

(10) Jusuf, S.; Loll, P. J.; Axelsen, P. H. *J. Am. Chem. Soc.* **2003**, *125* (13), 3988−3994.
(11) Hawkins, R. J.; McLeish, T. C. B. *J. R. Soc. Interface* **2006**, *3* (6), 125−138.
(12) Bray, D.; Duke, T. *Annu. Rev. Biophys. Biomol. Struct.* **2004**, *33*, 53−73.
(13) Changeux, J. P.; Edelstein, S. J. *Science* **2005**, *308* (5727), 1424−1428.
(14) Koshland, D. E. *Nat. Med.* **1998**, *4* (10), 1112−1114.
(15) Kumar, S.; Ma, B.; Tsai, C. J.; Sinha, N.; Nussinov, R. *Protein Sci.* **2000**, *9* (1), 10−9.
(16) Popovych, N.; Sun, S. J.; Ebright, R. H.; Kalodimos, C. G. *Nat. Struct. Mol. Biol.* **2006**, *13* (9), 831−838.
(17) Heyduk, T.; Lee, J. C. *Biochemistry* **1989**, *28* (17), 6914−6924.
(18) Leu, S. F.; Baker, C. H.; Lee, E. J.; Harman, J. G. *Biochemistry* **1999**, *38* (19), 6222−6230.
(19) Takahashi, M.; Blazy, B.; Baudras, A.; Hillen, W. *J. Mol. Biol.* **1989**, *207* (4), 783−796.
(20) Takahashi, M.; Blazy, B.; Baudras, A.; Hillen, W. *J. Mol. Biol.* **1989**, *207* (4), 783−96.
(21) Parkinson, G.; Wilson, C.; Gunasekera, A.; Ebright, Y. W.; Ebright, R. E.; Berman, H. M. *J. Mol. Biol.* **1996**, *260* (3), 395−408.
(22) Napoli, A. A.; Lawson, C. L.; Ebright, R. H.; Berman, H. M. *J. Mol. Biol.* **2006**, *357* (1), 173−183.
(23) Chen, S. F.; Vojtechovsky, J.; Parkinson, G. N.; Ebright, R. H.; Berman, H. M. *J. Mol. Biol.* **2001**, *314* (1), 63−74.
(24) Srinivasan, J.; Cheatham, T. E.; Cieplak, P.; Kollman, P. A.; Case, D. A. *J. Am. Chem. Soc.* **1998**, *120* (37), 9401−9409.
(25) Massova, I.; Kollman, P. A. *Perspect. Drug Discovery Des.* **2000**, *18*, 113−135.
(26) Wang, W.; Lim, W. A.; Jakalian, A.; Wang, J.; Wang, J. M.; Luo, R.; Bayly, C. T.; Kollman, P. A. *J. Am. Chem. Soc.* **2001**, *123* (17), 3986−3994.
(27) Passner, J. M.; Schultz, S. C.; Steitz, T. A. *J. Mol. Biol.* **2000**, *304* (5), 847−59.
(28) Case, D. A., et al. *AMBER 9*; University of California−San Francisco: San Francisco, CA, 2006.
(29) Jakalian, A. B. L. B.; Jack, D. B.; Bayly, C. I. Fast, efficient generation of high-quality; charges., a. *J. Comput. Chem.* **2000**, *21*, 15.
(30) Jakalian, A. D. B. J. a. C. I. B. *J. Comput. Chem.* **2002**, *23*, 18.

were immersed in a box of TIP3P[31] water molecules such that no atom in the complex was within 12 Å for CAP and 15 Å for CAP/DNA from any side of the box. All bonds involving hydrogen atoms were constrained by using the SHAKE[32] algorithm, affording the use of a 2 fs time step. The particle mesh Ewald[33] (PME) method was used to treat long-range electrostatics. Water molecules were first energy minimized and equilibrated by running a short simulation with the complex fixed by using Cartesian restraints. This was followed by a series of energy minimizations in which the Cartesian restraints were gradually relaxed from 500 kcal/Å^2 to 0 kcal/Å,^{2} and the system was subsequently gradually heated to 300 K via a 48 ps molecular dynamics run. Another 2 ns simulation was carried out at 300 K for further equilibration.

The method for determining the binding free energy following the MM-PBSA approach has been described in the past.[24−26] It combines molecular mechanics, Poisson−Boltzmann electrostatics, surface-accessible calculations, and normal mode analyses for the entropy. These are carried out on a series of snapshots collected from a molecular dynamics simulation. The binding free energy is expressed as

$$\Delta G_{\text{bind}} = \Delta G_{\text{MM}} + \Delta G_{\text{solv}}^{\text{PL}} - \Delta G_{\text{solv}}^{\text{P}} - \Delta G_{\text{solv}}^{\text{L}} \quad (1)$$

where ΔG_{bind} is the binding free energy, ΔG_{MM} is the free energy of association of the protein and ligand in the gas phase, and $\Delta G_{\text{solv}}^{\text{P}}$, $\Delta G_{\text{solv}}^{\text{L}}$, and $\Delta G_{\text{solv}}^{\text{PL}}$ correspond to the solvation free energies of the CAP (P), cAMP (L), and the CAP-cAMP complex (PL), respectively.

$$\Delta G_{\text{MM}} = \Delta E_{\text{MM}} - T\Delta S \quad (2)$$

where ΔE_{MM} is the difference in the sum of bond, angle, dihedral, electrostatic, and van der Waals energies between products and reactants computed with the AMBER force field. The solvation free energy is composed of two terms, namely, electrostatics and nonpolar, and is given by

$$\Delta G_{\text{solv}} = \Delta G_{\text{PB}} + \Delta G_{\text{nonpolar}} \quad (3)$$

In this work, a total of 2200 snapshots were extracted from the aforementioned molecular dynamics simulations of CAP and CAP/DNA complexes for binding energy calculation, and a subset of 880 snapshots were used for normal mode analysis. The electrostatic contribution to the solvation free energy is determined using the PBSA program in the AMBER 9 suite which numerically solves the Poisson−Boltzmann equations to determine the electrostatic contribution to the solvation free energy. A 0.5 Å grid size was used, and the dielectric constant for the solute and solvent were set to 1 and 80, respectively. The optimized atomic radii set in AMBER 9 were used, and partial charges were taken from Cornell et al.[34] for standard residues. The nonpolar contribution to the solvation free energy was determined using the Molsurf program, which is part of the AMBER 9 suite of programs. The entropy was computed for each snapshot from normal mode analyses by using the Nmode program within the AMBER package.

The generalized order parameters S^2, defined as the long time limit of internal motion autocorrelation function $C_i(t)$,[35,36] were evaluated from the trajectories using the N−H vectors of all backbone N atoms based on the following expression:

$$S^2 = \lim_{t\to\infty} C_i(t) = \lim_{t\to\infty} \langle P_2[\vec{\mu}_{\text{NH}}(0)\,\vec{\mu}_{\text{NH}}(t)]\rangle$$

P_2 is the second rank Legendre polynomial, and $\vec{\mu}_{NH}$ is the unit vector of N−H bond. P_2 was determined using the program Ptraj from the AMBER suite of programs. The residue based cross-correlation map was also determined using Ptraj. Essential dynamics analysis of CAP motion was carried out using Ptraj, and the motions were projected onto the CAP crystal structure thereafter. Animations were created using the program VMD.[37] Images were created using PyMOL[38] and VMD.

Distribution of the intrinsic disorder propensities within the CAP sequence was estimated by a set of PONDR algorithms (*P*redictor *O*f *N*atural *D*isordered *R*egions), VLXT and VSL1. PONDR VLXT is a set of neural network predictors of disordered regions on the basis of local amino acid composition, flexibility, hydropathy, coordination number, and other factors. These predictors classify each residue within a sequence as either ordered or disordered. PONDR VL-XT integrates three feed forward neural networks: the Variously characterized Long, version 1 (VL1) predictor from Romero et al.,[39] which predicts nonterminal residues, and the X-ray characterized N- and C-terminal predictors (XT) from Li et al.,[40] which predicts terminal residues. Output for the VL1 predictor starts and ends 11 amino acids from the termini. The XT predictors' output provides predictions up to 14 amino acids from their respective ends. A simple average is taken for the overlapping predictions and a sliding window of 9 amino acids is used to smooth the prediction values along the length of the sequence. Unsmoothed prediction values from the XT predictors are used for the first and last 4 sequence positions. The recently developed Various Short−Long, version 1 (PONDR-VSL1) algorithm is an ensemble of logistic regression models that predict per residue order−disorder.[40,41] Two models predict either long or short disordered regions—greater or less than 30 residues—based on features similar to those used by VL-XT. The algorithm calculates a weighted average of these predictions, where the weights are determined by a meta-predictor that approximates the likelihood of a long disordered region within its 61-residue window. Predictor inputs include PSI-blast[42] profiles and PHD[43] and PSI-pred[44] secondary structure predictions.

Results

Conformational Change in CAP. Molecular dynamics simulations provide coordinates with respect to time and hence enable a detailed analysis of the conformational evolution of a system. Starting with the crystal structure of CAP,[27] three separate 40 ns of explicit-solvent molecular dynamics simulations were carried out. This consisted of three separate 40 ns trajectories for the protein in its three liganded states, namely apo (CAP-APO), singly liganded (CAP-cAMP$_1$), and doubly liganded (CAP-cAMP$_2$). An analysis of the conformations sampled by the systems was conducted by determining the root-mean-square (rms) deviation of snapshots from the crystal structure. Larger rms deviations signify a greater degree of structural deviation from the original crystal structure. They were determined for (i) the N-terminus ligand-binding domain (Figure 1A) and (ii) the entire protein (Figure 1B). Figure 1A reveals

(31) Jorgensen, William L.; Jeffry, J. C.; Madura, D.; Impey, R. W.; Klein, Michael L. *J. Chem. Phys.* **1983**, *79* (2), 10.
(32) Ryckaert, J. P.; Ciccotti, G.; Berendsen, J. J. C. *J. Comput. Phys.* **1977**, *23*, 15.
(33) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98* (12), 10089−10092.
(34) 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* (9), 2309−2309.
(35) Wong, K. B.; Daggett, V. *Biochemistry* **1998**, *37* (32), 11182−11192.
(36) Rabah, G.; Popescu, R.; Cox, J. A.; Engelborghs, Y.; Craescu, C. T. *FEBS J.* **2005**, *272* (8), 2022−2036.
(37) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graphics* **1996**, *14* (1), 33−38.
(38) DeLano, W. L. *PyMOL*; DeLano Scientific: Palo Alto, CA, U.S.A. 2002.
(39) Romero, P.; Obradovic, Z.; Li, X.; Garner, E. C.; Brown, C. J.; Dunker, A. K. *Proteins* **2001**, *42* (1), 38−48.
(40) Li, X.; Romero, P.; Rani, M.; Dunker, A. K.; Obradovic, Z. *Genome Inform Ser Workshop Genome Inform* **1999**, *10*, 30−40.
(41) Obradovic, Z.; Peng, K.; Vucetic, S.; Radivojac, P.; Dunker, A. K. *Proteins* **2005**, *61 Suppl 7*, 176−82.
(42) Altschul, S. F.; Madden, T. L.; Schaffer, A. A.; Zhang, J.; Zhang, Z.; Miller, W.; Lipman, D. J. *Nucleic Acids Res.* **1997**, *25* (17), 3389−402.
(43) Rost, B.; Sander, C.; Schneider, R. *Comput. Appl. Biosci.* **1994**, *10* (1), 53−60.
(44) Jones, D. T.; Ward, J. J. *Proteins* **2003**, *53 Suppl 6*, 573−578.

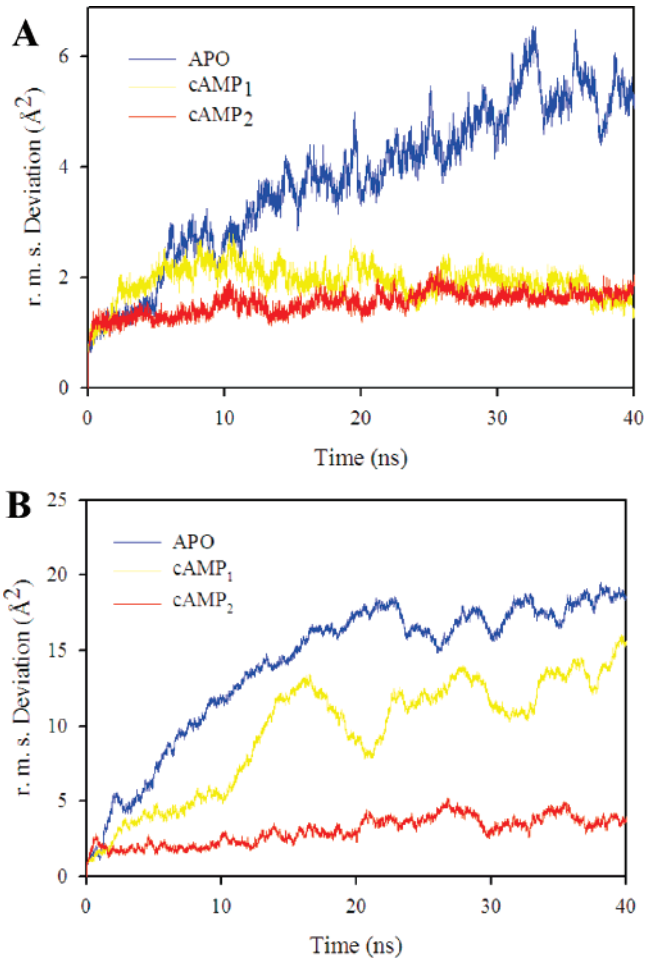


Figure 1. (A) Plot of the rms devation that is experienced by the N-terminus ligand binding domain over the course of the trajectories for CAP-APO (blue), CAP-cAMP1 (yellow), and CAP-cAMP2 (red). (B) Same as (A), except that the rms deviation values are for the entire protein.

that the N-terminus domains of CAP-cAMP$_1$ and CAP-cAMP$_2$ remain structurally stable over the course of the trajectory, as evidenced by rms deviation values that consistently fall within 2 Å. The structure of the apo protein, on the other hand, showed larger structural variation with a final rms deviation reaching nearly 4 Å.

In the case of the full-length protein, the rms deviations suggested that large conformational change was occurring in the CAP-APO and CAP-cAMP$_1$ systems, with values reaching 15 Å, while the CAP-cAMP$_2$ structure remained relatively stable with ~2−3 Å values. The stark contrast between the structural flexibility of the N-terminus ligand binding domain and the full protein suggests that the conformational change either was concentrated on the DNA-binding module or perhaps occurring in a region between the C- and N-terminus. To gain better insight into this structural change, the trajectories CAP-APO, CAP-cAMP$_1$, and CAP-cAMP$_2$ were rendered into animations, which were captured in MPEG files that are provided in the Supporting Information. Visualization of the CAP-cAMP$_2$ trajectory revealed that CAP retained a conformation that is similar to its crystal structure over the course of the trajectory, albeit the DNA binding module appears to experience a greater degree of motion relative to the N-terminus domain (Figure 2A). In sharp contrast, results from the CAP-cAMP$_1$ trajectory revealed a remarkable event: a region that serves to connect

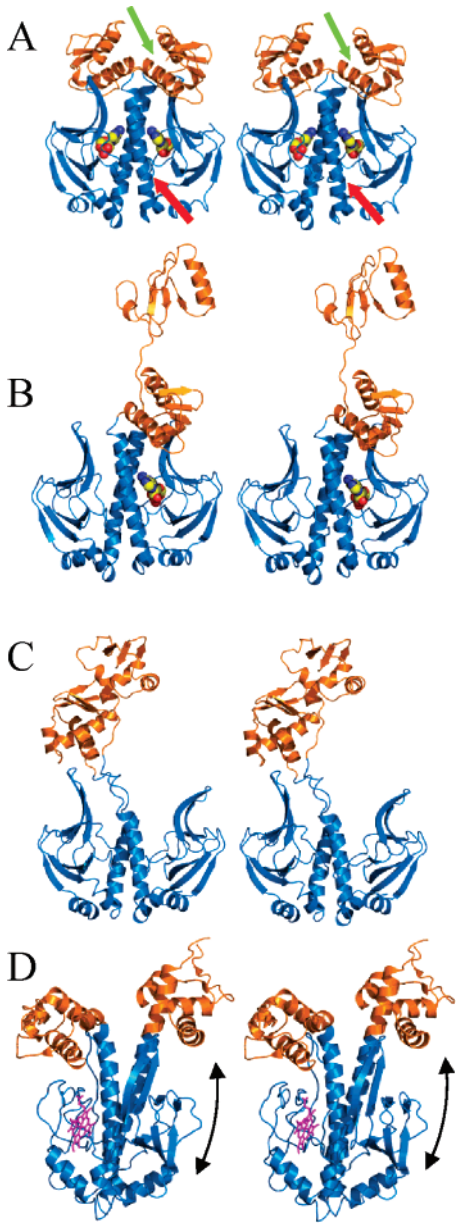


Figure 2. Stereoview of the three-dimensional structure of (A) doubly liganded CAP with the red arrow pointing to α-helix C and the green arrow pointing to α-helix D; (B) singly liganded CAP; (C) unliganded CAP. In each case, the protein is shown in ribbon representation (blue for the N-terminus and orange for the C-terminus), while the ligand is shown in sphere representation (yellow, blue, and red correspond to carbon, nitrogen, and oxygen, respectively). (D) Stereoview of the crystal structure of heme-binding catabolite activator protein CooA (PDB code 1FT9). The protein is depicted in ribbon representation (blue for the N-terminus and orange for the C-terminus), and the heme domain is shown in capped-sticks representation and color-coded pink.

the N- and C-termini of CAP known as the hinge domain (residues 130−139) suddenly unravels, leading to the detachment of the C-terminus DNA-binding module and its separation from the N-terminus ligand-binding domain (Figure 2B). In the final structure, the hinge domain is seen as unstructured, but an α-helix (α-helix D, residues 140−150) from the C-terminal DNA binding module also unraveled (see Figure 2A for location of α-helix D). The N-terminus modules largely maintain their structural integrity during the process. Visualization of the CAP-APO trajectory reveals the same remarkable dissociation of the C-terminus DNA-binding module, but unlike the CAP-cAMP$_1$

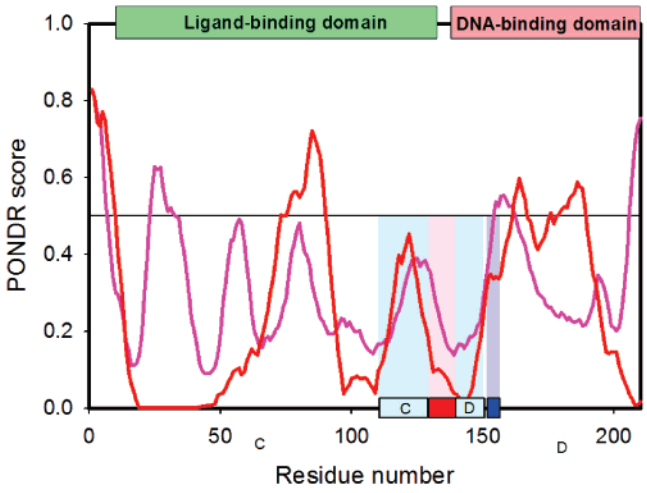


Figure 3. PONDR VLXT (red line) and VSL1 (pink line) score distributions for CAP. Localizations of ligand-binding and DNA-binding domains are indicated as green and pink boxes, respectively. Structural elements discussed in the text are also shown: α-helix C (residues 111−129, cyan box marked C and cyan shading), hinge region (residues 130−139, red box and light pink shading), α-helix D (residues 140−150, cyan box marked D and cyan shading), and *β*-hairpin (residues 152−156 blue box and light blue shading).

simulation, *both* modules dissociate, again due to the unraveling of the hinge region that connects the N- and C-terminus domains. However, more significant unraveling than the CAP-$cAMP_1$ is observed here. In addition to α-helix D, the α-helix along the interface between the CAP monomers (α-helix C, residues 111−129) also undergoes some degree of unraveling. Furthermore, another conformational change is observed, where the *β*-hairpin (residues 152−156) separates from α-helix C, leading to additional exposure of the cAMP binding pocket to solvent. This opening is likely necessary to enable cAMP to access the active site, as the crystal structure in the doubly liganded form does not show a suitable path that the ligand can follow to adopt its final bound state.

To understand whether the mechanisms of the observed behavior are encoded in the CAP primary structure, we have evaluated the distribution of intrinsic disorder propensities in CAP amino acid sequence using two predictors of intrinsic disorder, members of the PONDR family, VLXT and VSL1. PONDR VLXT was chosen because this predictor is known to find some function-related features in proteins (e.g., potential protein−protein interaction sites),[45−47] whereas the choice of PONDR VSL1 was determined by the fact that this algorithm is the most accurate predictor of intrinsic disorder developed so far.[48] Figure 3 shows that α-helix C (residues 111−129) and *β*-hairpin (residues 152−156) are located in regions with increased flexibility as predicted by both algorithms, whereas the hinge region (residues 130−139) and α-helix D (residues 140−150) have boundary positions and are located at "slopes" of PONDR curves. The localization of α-helix C and *β*-hairpin to regions with increased probability to be disordered suggests that these segments potentially possess internal flexibility. It seems that the entire structural element "α-helix C−hinge−α-helix D−*β*-hairpin" is stabilized via the interaction of α-helix D with the remainder of the molecule. This interaction is responsible for holding the DNA-binding domain bound to the ligand-binding domain. When it is lost, the DNA-binding domain is set free, and modules dissociate.

Ligand Binding Induces Change in Motion. In NMR, the order parameter, S^2, is often used to probe motion in the ps/ns regime. The values for S^2 typically range from 0 to 1. When $S^2 = 1$, the vectors that represent NH bonds from the protein backbone have little motion, and when $S^2 = 0$, the bond vectors rapidly sample multiple orientations. Since molecular dynamics simulations sample motions in the ps/ns regime and provide snapshots with respect to time, it is possible to use coordinates from these snapshots to estimate the order parameter.[49] We computed S^2 values for the CAP-APO, CAP-$cAMP_1$, and CAP-$cAMP_2$ trajectories to assess the effect of ligand binding on the ps/ns regime. We find that the binding of the first ligand resulted in an increase of the mean order parameter ($\langle \Delta S^2 \rangle = 0.15$). This suggests a tightening of the protein structure. The binding of the second cAMP molecule also resulted in an increase in S^2, but not as significant as the binding of the first ligand ($\langle \Delta S^2 \rangle = 0.11$). The systematic increase of S^2 that we find from our trajectories is consistent with what is observed from S^2 values from NMR.[16] The NMR results, however, showed a larger increase in the order parameters upon binding of the second ligand. This difference could be due to the fact that our simulations are based on the full structure of CAP, while the NMR experiments were conducted on the truncated structure that does not include the DNA-binding domain.

To gain further insight into the changes in the motions due to ligand binding, we investigated the correlation between the motions of residues in all three CAP trajectories. A convenient way to illustrate these correlations is through the construction of a dynamical cross-correlation map (normalized covariance matrix). These maps were constructed for the CAP-APO, CAP-$cAMP_1$, and CAP-$cAMP_2$ trajectories, as shown in Figure 4. We exploited the symmetric nature of these maps and used the upper quadrant to show correlation coefficients of one monomer, while the lower quadrant corresponds to the correlation coefficients of the other monomer. These coefficients provide information about the correlation between the fluctuations of the positions of the residues and secondary structure elements in the protein; it is worth mentioning that these coefficients do not reflect on the degree of motion experienced by residues. These values range from −1 (anticorrelated motion) to 1 (highly correlated motion). Negative values correspond to anticorrelated motion, where residues are generally moving in opposite directions, and positive values correspond to correlated motions, during which residues are generally moving in the same direction. All three maps show correlations that range from highly anticorrelated (blue) to highly correlated (red). In the case of the CAP-APO simulation (Figure 4A), a greater degree of correlated motion is observed, as evidenced by the large number of red patches on the surface. This is in contrast to the CAP-$cAMP_1$ and CAP-$cAMP_2$ maps (Figure 4B and 4C) that indicate a dampening of the highly correlated motions that were observed in the CAP-APO simulation. Closer scrutiny of the CAP-$cAMP_1$ trajectory resulted in a notable observation: the

(45) Garner, E.; Romero, P.; Dunker, A. K.; Brown, C.; Obradovic, Z. *Genome Inform Ser Workshop Genome Inform* **1999**, *10*, 41−50.
(46) Oldfield, C. J.; Cheng, Y.; Cortese, M. S.; Romero, P.; Uversky, V. N.; Dunker, A. K. *Biochemistry* **2005**, *44* (37), 12454−70.
(47) Mohan, A.; Oldfield, C. J.; Radivojac, P.; Vacic, V.; Cortese, M. S.; Dunker, A. K.; Uversky, V. N. *J. Mol. Biol.* **2006**, *362* (5), 1043−59.
(48) Peng, K.; Radivojac, P.; Vucetic, S.; Dunker, A. K.; Obradovic, Z. *BMC Bioinformatics* **2006**, *7*, 208.
(49) Case, D. A. *Acc. Chem. Res.* **2002**, *35* (6), 325−331.

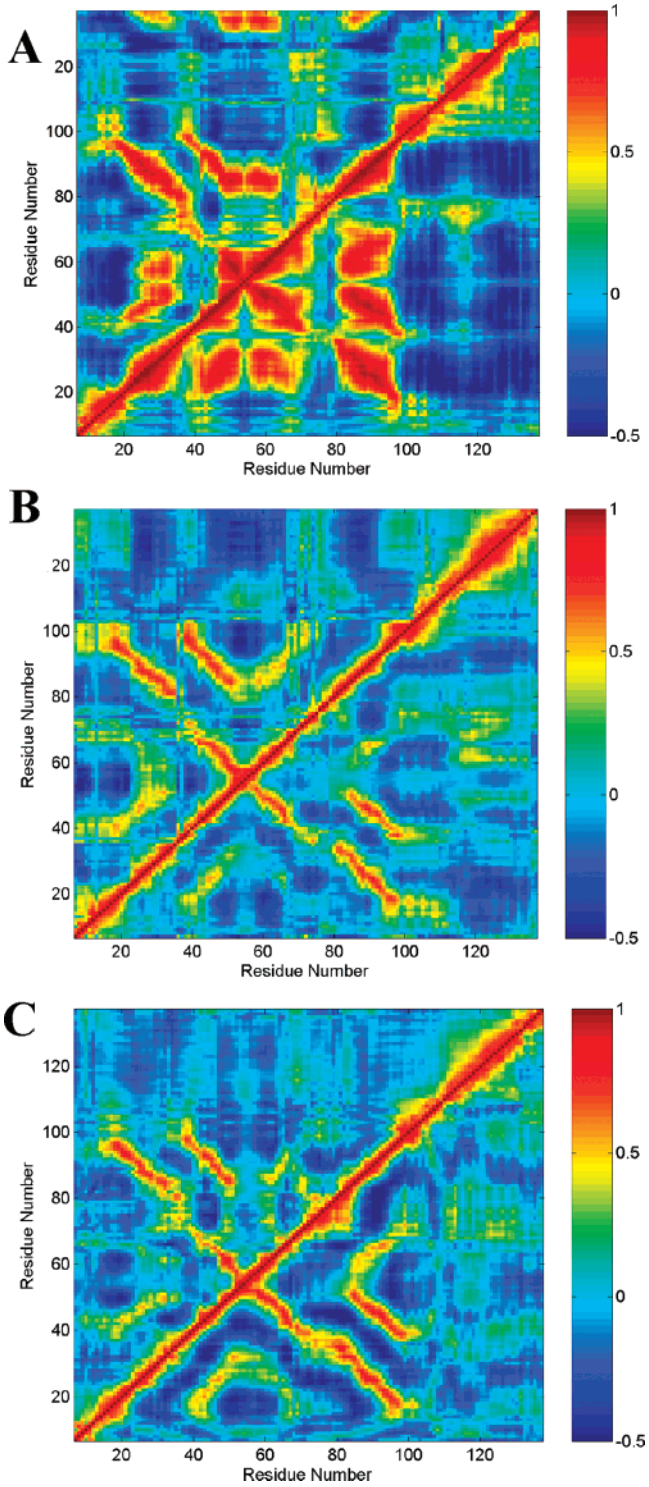


Figure 4. Dynamical cross-correlation maps illustrating the correlation of motion between residues in (A) CAP-APO, (B) CAP-cAMP$_1$, and (C) CAP-cAMP$_2$.

dampening of the correlation coefficients occurred in *both* quadrants of the CAP-cAMP$_1$ map, despite the fact that binding of cAMP occurs in only one of the monomers. These results indicate that the binding of a single cAMP is able to propagate change in motion across the dimeric interface and hence is a direct reflection of the cooperative nature of the CAP system. They also suggest that ligand binding to one monomer is capable of altering the motion of the unliganded monomer across the dimeric interface.

Further analysis of the dynamics of the CAP systems was conducted by using the three 40 ns nanosecond trajectories to project the type of motions that are likely experienced by the system beyond the nanosecond regime. The approach we follow is to extract collective motions with the use of principal component analysis (PCA), which consists of the diagonalization of the covariance matrix and the projection of the resulting eigenvectors onto the structure of the protein. This approach has been widely used in the past.[50] Recently, an extensive study showed that short nanosecond trajectories were sufficient to reproduce most of the motions that occur at longer time scales.[51] It is worth noting that such an analysis results in a large number of possible motions (N), as it is based on an $N \times N$ covariance matrix, where N corresponds to the number of heavy atoms in the system. It has been shown that 90% of the total atomic motion is described by less than 5% of all degrees of freedom.[52−54] Consequently, we focused on the motions corresponding to the lowest frequency from the CAP-APO, CAP-cAMP$_1$, and CAP-cAMP$_2$ trajectories since those motions make the most significant contribution to the dynamics. We color-coded the residues in the structures using the magnitude of the eigenvectors averaged over all heavy atoms in each residue to illustrate the degree of motion experienced by each residue (Figure 5). A general inspection of the images in Figure 5 clearly shows that the three states of CAP experience different types of motions on the longer time scale, which is in full agreement with a recent NMR study showing that ligand binding results in a change in these motions. In the apo structure, the large number of red residues near the hinge domain confirms the observations from the molecular dynamics simulations of an unstructured hinge domain. Figure 5B shows that these motions are significantly dampened upon binding of the first ligand, as evidenced by the reduction in the red color-coded regions in the protein. Further dampening of the motion is observed as a result on the binding of the second ligand.

To provide insight into the direction of motion that the residues in each system were experiencing, the essential dynamics were captured into MPEG animations, which are made available in the Supporting Information. These animations were created by mapping the eigenvectors obtained from the essential dynamics analysis onto the crystal structure of CAP. Visualization of the animations confirmed the differences in the motions that were initially suggested by Figure 5. These differences clearly indicate that ligand binding has an effect not only on ps/ns motions but also on the longer time scale motions.

Binding Thermodynamics. The availability of calorimetry data for the binding of the first and second cAMP molecule to CAP provided an impetus to carry out post-trajectory free energy calculations in the spirit of the MM-PBSA methodology.[24−26] This method is highly suitable as it can be implemented on a standard molecular dynamics simulation. It is also highly insightful as it provides the various components of the binding free energy, including the energy due to Coulombic and van der Waals interactions, electrostatics and nonpolar components

(50) Amadei, A.; Linssen, A. B. M.; Berendsen, H. J. C. *Proteins: Struct., Funct., Genet.* **1993**, *17* (4), 412−425.
(51) Lange, O. F.; Grubmuller, H. *J. Phys. Chem. B* **2006**, *110* (45), 22842−22852.
(52) Hayward, S.; Kitao, A.; Hirata, F.; Go, N. *J. Mol. Biol.* **1993**, *234* (4), 1207−1217.
(53) Kitao, A.; Go, N. *Curr. Opin. Struct. Biol.* **1999**, *9* (2), 164−169.
(54) Berendsen, H. J. C.; Hayward, S. *Curr. Opin. Struct. Biol.* **2000**, *10* (2), 165−169.

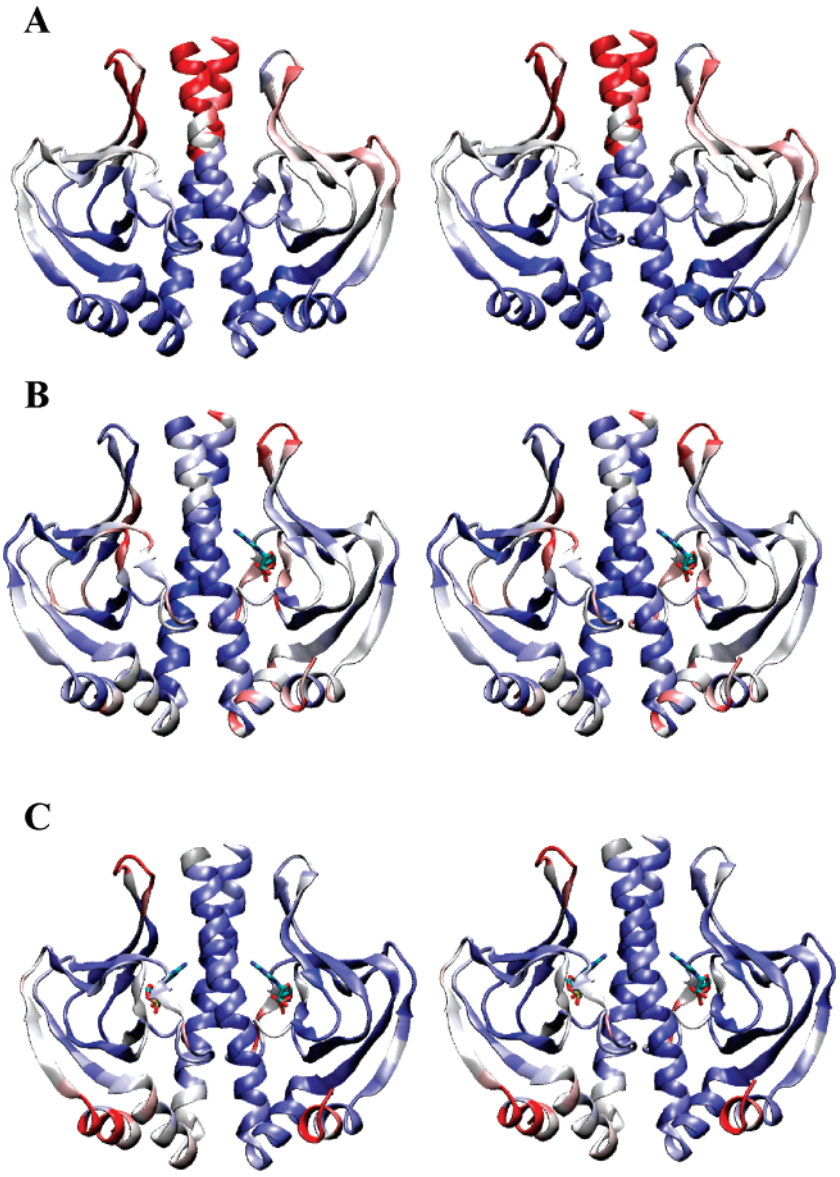


Figure 5. Stereoview of CAP colored-coded based on the magnitude of the eigenvector corresponding to the largest eigenvalue obtained from diagonalizing the covariance matrix of (A) APO-CAP, (B) CAP-cAMP$_1$, and (C) CAP-cAMP$_2$. Red corresponds to high flexibility while blue corresponds to low flexibility.

Table 1. Free Energy Calculations for the Binding of cAMP to CAP and CAP/DNA Complex[a]

	ΔE^{vdw}	ΔE^{elec}	ΔG^{PB}	ΔG^{SA}	$T\Delta S^{nmode}$	ΔG^{calc}	$T\Delta\Delta S^{nmode}$	$T\Delta\Delta S^{exp}$	$\Delta\Delta G^{calc}$	$\Delta\Delta G^{exp}$
CAP-cAMP$_1$	-35.0 ± 0.7	-77.3 ± 1.5	77.0 ± 1.5	-3.9 ± 0.1	-20.4 ± 1.2	-18.8 ± 1.5				
CAP-cAMP$_2$	-37.3 ± 0.5	-68.2 ± 1.0	71.2 ± 1.0	-4.2 ± 0.1	-22.3 ± 1.3	-16.1 ± 1.4	-1.9 ± 1.8	-3.9	2.7 ± 2.1	2.8
CAP/DNA-cAMP$_1$	-41.9 ± 0.2	-107.9 ± 0.4	98.0 ± 0.3	-4.7 ± 0.1	-20.1 ± 1.0	-36.4 ± 1.1				
CAP/DNA-cAMP$_2$	-40.5 ± 0.1	-92.6 ± 0.5	74.8 ± 0.3	-4.7 ± 0.1	-20.3 ± 1.1	-42.7 ± 1.2	-0.2 ± 1.5		-6.3 ± 1.6	

[a] Units in kcal/mol; ΔE^{vdw}, van der Waals potential energy; ΔE^{elec}, electrostatic potential energy; ΔG^{PB}, electrostatic contributions to the solvation free energy calculated with Poisson−Boltzmann equation; ΔG^{SA}, nonpolar contributions to solvation free energy; ΔS^{nmode}, entropy calculated with normal mode analysis; $T\Delta\Delta S^{nmode}$, change in the calculated entropic contribution to the free energy of binding. $T\Delta\Delta S^{exp}$, change in the experimentally determined entropic contribution to the free energy of binding. $\Delta\Delta G^{calc}$, change in the calculated free energy of binding; $\Delta\Delta G^{exp}$, change in the experimentally determined free energy of binding.

of the solvation free energy, and the entropy. The results from this exercise are provided in Table 1. These calculations were carried out for the three liganded states of CAP, in both the presence and absence of DNA. In the absence of DNA, the calculated free energy change due to binding of the second cAMP is 2.7 kcal/mol, in excellent agreement with the free energy change measured from isothermal titration calorimetry.[16] The calculations also reveal an entropy change of 1.9 kcal/mol. It is worth mentioning that, by definition, the thermodynamic cycle that is constructed to carry out the MM-PBSA calculations assumes that the calculated entropy is that of the system in the absence of solvent. Hence, it is not a surprise that the computed value does not show a high degree of agreement with the ITC measured value of 3.9 kcal/mol. Nevertheless, it appears that

both free energy changes due to entropy changes are positive, suggesting that the binding of the second cAMP results in a tightening of the overall structure. The molecular dynamics simulations also made it possible to compute the conformational entropy change due to ligand binding and compare with values that were computed from NMR.[16] Order parameters that were computed from the trajectories revealed a systematic decrease as a result of each binding event. The results are consistent with the trend observed from NMR.[16]

An attractive feature of MM-PBSA calculations is that it provides insight into the components of the free energy. Table 1 reveals that the binding of the second ligand results in a non-negligible change in the van der Waals energy (ΔE^{vdw}), nearly 2 kcal/mol. This change is likely due to minor alterations in the structure that are introduced as a result of the ligand occupying the binding site on the second subunit. The structural change is likely occurring away from the ligand site, since cAMP binds to both CAP subunits in an identical manner.[27] The calculations revealed little change in the nonpolar component of the free energy of solvation (ΔG^{SA}). Despite the slightly favorable contributions from nonpolar effects ($\Delta E^{vdw} + \Delta G^{SA}$), the electrostatic contributions as a result of binding of a second cAMP molecule play a destabilizing role, mainly due to unfavorable contributions from the Coulombic interactions (ΔE^{elec}). The increase in the free energy due to ΔE^{elec} is compensated by a decrease in the desolvation energy ΔG^{PB} as a result of binding of a second cAMP.

To determine the role of DNA in the binding of cAMP, we conducted a series of 5 ns simulations in the presence of DNA for CAP-APO, CAP-$cAMP_1$, and CAP-$cAMP_2$. Results from free energy calculations from snapshots collected from these simulations are presented in Table 1. The results were highly intriguing as the free energy change due to binding of a second cAMP was very favorable. This change suggests that, in the presence of DNA, cAMP binding to CAP is positively cooperative. The van der Waals energy appears to be slightly unfavorable when two cAMP molecules are bound, suggesting less contacts. The electrostatic interactions were also highly unfavorable in the doubly liganded form. But these unfavorable interactions are compensated by the favorable electrostatics component of the solvation free energy. The nonpolar component of the free energy showed no change. Similarly, there was no observed change in the entropy. It thus appears that the presence of DNA leads to less tightening of the structure and hence a more favorable contribution from the entropy.

Discussion

The large separation between the cAMP binding sites in the CAP dimers has made it difficult to identify the causes for the negative cooperativity between the sites, as electrostatic or bonding interactions are ruled out as the cause. It has been proposed that the basis of cooperativity is enthalpic in nature,[12−14] suggesting changes in the internal energy of the system due to structural alterations. Others have suggested that these changes are due to alterations in the dynamics of the system.[16] Our simulations of the CAP system in its various liganded states have revealed both structural and dynamical changes. The structural changes occur not only in the immediate region surrounding the binding sites of cAMP as illustrated by changes in the van der Waals and Coulombic terms of the free energy (Table 1) but also in the overall structure of the ligand binding domain as detected by rms deviation analysis of the trajectories. Additional flexibility was detected with a remarkable dissociation of the DNA binding module of the ligand binding module (Figure 2). This dissociation was found to be directly coupled to the presence of cAMP, despite the lack of direct interaction between the two. This strongly suggests an entropic effect, whereby the presence of the cAMP stabilizes the system, as evidenced by a reduction in the entropy upon binding of a second cAMP (Table 1). Entropic effects were also observed from the dynamical changes that were detected within the ligand-binding module of CAP. For example, a shift in the correlation of motions is seen as a result of binding of cAMP on both subunits of CAP, despite the fact that cAMP only binds to one of the subunits. In addition, essential dynamics analyses indicate changes in motion in the longer time scale. The preponderance of our data supports a more complex view of allostery in the CAP system, revealing an interplay between enthalpic and entropic effects that are responsible not only for the cooperative effect of cAMP binding but also for CAP's ability to bind DNA and perform its function as a transcriptional activator.

The remarkable detachment of the C-terminal DNA-binding domain from the N-terminus ligand binding domain that occurs in the unoccupied monomers of CAP during the molecular dynamics simulations leads to three distinct conformational states of CAP. This is in accordance with biochemical analyses that have shown that CAP exists in three different conformations depending on the concentration of cAMP.[17,20] Support for the conformational change can also be found in other experimental investigations. Proteolytic studies have revealed that exposure of CAP to proteases at micromolar concentrations of cAMP results in the separation of C- and N-terminus domains due to cleavage along the hinge domain.[55] This suggests a disordered hinge domain region, which is consistent with observations from the molecular dynamics simulations and intrinsic disorder predictions. Further evidence of the unraveling of the hinge region comes from NMR studies, which reveals significant differences in the structure between the apo-CAP and CAP-$cAMP_1$ hinge domains.[56] The most compelling experimental evidence of the conformational change experienced in the unoccupied CAP subunits stems from a crystal structure of another heme-binding catabolite activator protein, namely CooA, which was crystallized with a heme bound to only one subunit.[57] CooA shares a significant structural resemblance to CAP. The structure reveals the same remarkable displacement of the C-terminus DNA binding module, which occurs *only* in the unoccupied subunit (Figure 2D).[57]

At physiological conditions, the concentration of cAMP dictates that CAP exists in its apo *or* singly liganded state, suggesting that CAP binds to DNA while adopting one of these conformational states.[58] But crystal structures have unanimously shown that the doubly liganded form of the protein is required for binding to DNA.[27] This apparent contradiction led us to hypothesize that CAP first binds to DNA in its singly liganded

(55) Kolb, A.; Busby, S.; Buc, H.; Garges, S.; Adhya, S. *Annu. Rev. Biochem.* **1993**, *62*, 749−95.
(56) Won, H. S.; Yamazaki, T.; Lee, T. W.; Yoon, M. K.; Park, S. H.; Kyogoku, Y.; Lee, B. J. *Biochemistry* **2000**, *39* (45), 13953−13962.
(57) Lanzilotta, W. N.; Schuller, D. J.; Thorsteinsson, M. V.; Kerby, R. L.; Roberts, G. P.; Poulos, T. L. *Nat. Struct. Biol.* **2000**, *7* (10), 876−80.
(58) Harman, J. G. *Biochim. Biophys. Acta* **2001**, *1547* (1), 1−17.

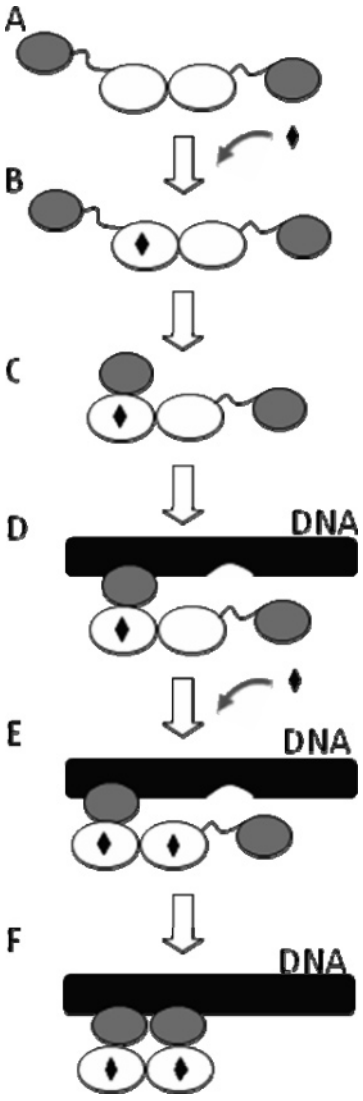


Figure 6. Schematic illustration of the new mechanism of CAP in the transcription activation process.

form, and DNA would favor the formation of the doubly liganded form of CAP. Results from our free energy calculations in the presence of DNA strongly supported this hypothesis, as we found that the affinity of the second cAMP to CAP increased significantly. Hence, CAP appears to switch from a negative cooperative system to a positive cooperative system in the presence of DNA.

These findings prompted us to propose a new mechanism for the role of CAP in the transcription activation process, as illustrated in Figure 6. In the apo form, both DNA modules of CAP are displaced (Figure 2C), and it is widely accepted that there is no binding to DNA in this state (Figure 6A). Upon an increase in concentration of cAMP to micromolar levels, due to a drop in glucose concentration, for example, one cAMP molecule will bind to each CAP resulting in the singly liganded state of CAP (Figure 6B). The lack of [experimentally observed] binding to DNA in the apo form and the strong binding in the singly liganded form, coupled with the observation from our simulations that the binding of cAMP to the N-terminus domain stabilizes the complex between the N− and C-termini (Figure 6C), suggests that binding to DNA only occurs when the C- and N-termini of CAP are bound. This is illustrated in Figure 6D, which shows that only the monomer with bound N- and C-termini binds to DNA, while the other remains free in solution. But the presence of DNA will strongly favor binding of another cAMP, as we have shown from free energy calculations in the presence and absence of DNA (Figure 6E). The binding of another cAMP to the disordered monomer will cause the N- and C-termini to come together and the second module to bind to DNA (Figure 6E). As our simulations have shown, the presence of cAMP has a tendency to stabilize the binding of the DNA binding module to the ligand-binding module of CAP (Figure 6F). The resulting complex then binds to DNA and leads to the necessary bending of DNA for RNA polymerase to bind and initiate transcription.

The extensive molecular simulations and free energy calculations carried out in this study show that molecular dynamics simulations and ensuing free energy calculations can be used as effective tools to detect and analyze allosteric events in macromolecular-ligand interactions. These results reveal that CAP's role as a transcription activator is inextricably linked to the cooperativity that the protein exhibits. Our results show, for the first time, that the experimentally observed disorder in the hinge region results in complete separation of the N- and C-terminus domains of CAP, which is tightly controlled by the binding of cAMP. We find that DNA binding to CAP in its singly liganded form leads to a switch in cooperativity that promotes binding of the other cAMP, leading to the essential events that are required for initiation of transcription activation. These results show that (i) allostery in CAP is due not only to alterations in dynamics but also to conformational changes that are controlled by cAMP binding and (ii) DNA plays an active role in the initiation of the activation process, rather than simply serving as a mere receptor for the doubly liganded form of CAP. While this study has shed light into the mechanism of cooperativity in the CAP system, it remains to be seen whether the conformational changes detected follow an induced-fit or pre-existing complex/conformational selection mechanism. Only an extensive free energy profile, such as a molecular dynamics-based potential of mean force analysis using a targeted dynamics approach, will identify the barriers that are required to interconvert between these conformations and hence support one of these mechanisms. This work is currently ongoing in our laboratory.

Acknowledgment. The research was supported by the INGEN grant from the Lilly Endowment, Inc. Computer time on the Big Red supercomputer at Indiana University is funded by the National Science Foundation under Grant Nos. ACI-0338618l, OCI-0451237, OCI-0535258, and OCI-0504075 as well as by Shared University Research grants from IBM, Inc. to Indiana University. This research was supported in part by the Indiana METACyt Initiative. The Indiana METACyt Initiative of Indiana University is supported in part by Lilly Endowment, Inc.

Supporting Information Available: Complete ref 19. MPEG files corresponding to animations discussed in the text. This material is available free of charge via the Internet at http://pubs.acs.org.

JA076046A