

Structural Basis for Cooperative DNA Binding by CAP and *Lac* Repressor

Alexander Balaeff,^{1,4} L. Mahadevan,³
and Klaus Schulten^{1,2,*}

¹Center for Biophysics and Computational Biology
and Beckman Institute

²Department of Physics
University of Illinois at Urbana-Champaign
Urbana, Illinois 61801

³Harvard University Division of Engineering
and Applied Sciences
29 Oxford Street
Cambridge, Massachusetts 02138

Summary

Catabolite gene activator protein (CAP) and *lac* repressor (LR) are celebrated transcription-regulating proteins that bind to DNA cooperatively forming a ternary complex with the promoter loop. Here we present a multiscale model of the ternary complex derived from crystal structures of the proteins and a continuous structure of the DNA loop built using the theory of elasticity. We predict that the loop is underwound in the binary complex with the LR, whereas in the ternary complex with the LR and CAP, the loop is overwound and extended due to an upstream relocation of a DNA binding hand of LR. The computed relocation distance matches the experimental observations and the energy balance of the system explains the cooperativity effect. Using the multiscale approach, we build an all-atom model of the ternary complex that suggests a series of further experimental investigations.

Introduction

Catabolite gene activator protein (CAP) and *lac* repressor (LR) are two well-known *E. coli* proteins regulating the level of transcription from *lac* operon, a textbook example of a genetic control system (Alberts et al., 2002; Berg et al., 2002) (illustrated in Figure 1A). CAP is a universal transcriptional activator that affects more than hundred *E. coli* promoters (Busby and Ebright, 1999), including that of the *lac* operon. CAP attaches itself to one of several specifically recognized 22 base pair (bp) long DNA sites upstream from a promoter and facilitates the subsequent binding of RNA polymerase, resulting in a manifold increase in the level of transcription. There are three CAP binding sites (CBSs) upstream from the *lac* operon promoter; the CBSs are centered near positions –61, –71, and –103 (Schultz et al., 1991). A bound CAP creates two symmetric kinks within the CBS (Figures 1B and 1C), resulting in the effective DNA bend of approximately 80° (Schultz et al., 1991; Parkinson et al., 1996; Kapanidis et al., 2001).

LR, on the other hand, shuts down the *lac* operon

when the bacterial environment contains no lactose (Alberts et al., 2002; Berg et al., 2002). The repressor functions as a dimer of two “hand” domains (Figure 1D). Each hand binds with a high specificity to a 21 bp operator DNA (Lewis et al., 1996; Barkley and Bourgeois, 1980). One hand always binds to an operator centered at position 11 (O₁), and the other hand, to either of two operators centered at positions 412 (O₂) and –82 (O₃) (Oehler et al., 1990; Lewis et al., 1996). As a result, the DNA between the bound operators folds into a loop (Oehler et al., 1990) (Figure 1A). The formation of the loop is critical for the *lac* operon repression (Matthews, 1992).

The primary roles of CAP and LR are the opposite of each other, yet the two proteins bind to DNA cooperatively (Hudson and Fried, 1990; Lewis et al., 1996). CAP and LR form a ternary complex with the O₁-O₃ DNA loop (which contains a CBS inside [Lewis et al., 1996]; cf. Figure 1A) and increase each other’s affinity to the DNA by 4- to 11-fold, which corresponds to a free energy gain of –0.8 to –1.4 kcal/mol, or –1 to –2 kT (Hudson and Fried, 1990). This cooperative behavior led to a reinterpretation of the role of CAP as that of not simply a transcriptional activator, but also a factor that amplifies the response of the *lac* operon genetic switch to the level of lactose in the environment (Hudson and Fried, 1990; Alberts et al., 2002).

In this manuscript, we present a multiresolution model of the ternary complex of CAP, LR, and the O₁-O₃ DNA loop and use the model to explain the structural basis for the cooperative DNA binding by CAP and LR. The model consists of all-atom structures of CAP- and LR-DNA complexes and a continuous structure of a 76 bp long section of the O₁-O₃ loop, built using the theory of elasticity (Olson and Zhurkin, 2000; Balaeff et al., 1999; Olson, 1996). The binding of CAP within the loop is mimicked via “intrinsic” curvature and twist terms in the equations of elasticity (Westcott et al., 1997; Yang et al., 1995). The multiresolution models of the ternary complex are built for different lengths of the DNA loop and the free energy of cooperation between CAP and LR is estimated for each model.

We find that two structures of the DNA loop can exist: an underwound one, preferred in the binary complex of the DNA loop with the LR, and an overwound one, preferred in the ternary complex with LR and CAP. Our results indicate that the conformational switch between the two states is accompanied by an upstream relocation of the O₃ binding “hand” of the LR; the predicted distance of relocation agrees well with an estimate from DNase I footprinting experiments (Perros et al., 1996). The free energy, released during the conformational switch, accounts for the cooperativity effect. Finally, we build an all-atom structure of the whole ternary complex on top of the multiresolution structure and discuss how that structure can be used in the design of further biochemical and computational experiments that would test our predictions and further advance the study of the CAP-LR-DNA ternary complex.

*Correspondence: kschulte@ks.uiuc.edu

⁴Present address: IBM Thomas J. Watson Research Center, 1101 Kitchawan Road, Route 134, Yorktown Heights, New York 10598.

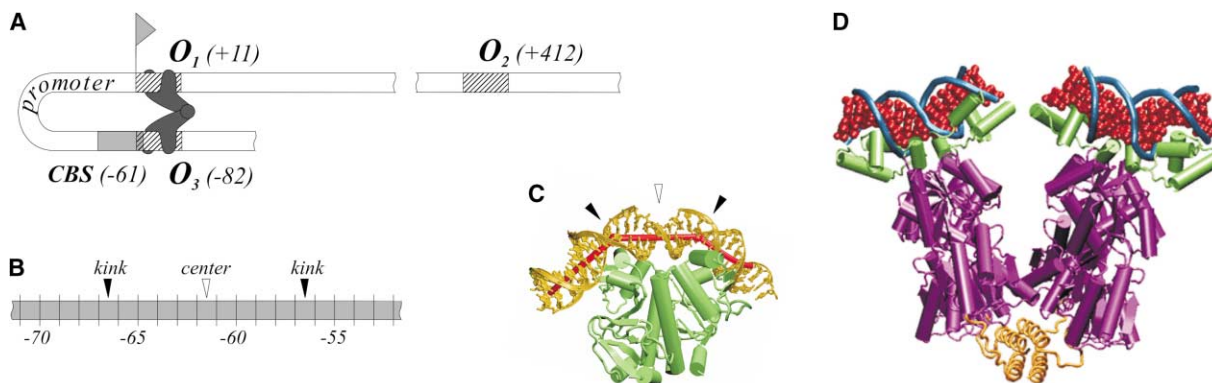


Figure 1. CAP and LR Structures and Binding Sites

(A) Diagram of the *lac* operon promoter bent by LR. Hashed bars indicate the three LR operators. The shaded bar represents the CAP binding site (CBS) centered at position -61.5 . LR is shown bound to O_1 and O_3 , folding the DNA between the operators into a 76 bp long loop. The flag indicates position $+1$ of the operon.

(B) The map of the CBS shows the location of the DNA steps kinked by a bound CAP.

(C) Crystal structure of the CAP-DNA complex from Parkinson et al. (1996). The elastic rod model of DNA in the CBS is shown as a tube fitted inside the DNA segment (cf. Figures 2A, 4, and 6).

(D) The hybrid all-atom structure of the LR complex with O_1 and O_3 , constructed from the available PDB structures, as detailed in Modeling Procedures (cf. Figure 7). Structures in (C) and (D) are not drawn to scale.

Results and Discussion

“U” and “O” Structures of the O_1 - O_3 Loop

Our model of the ternary complex of CAP, LR, and the O_1 - O_3 DNA loop is based on the elastic rod structure of the loop (cf. Figure 2). The latter is obtained by solving Kirchhoff equations of elasticity, for which the all-atom structures of the CAP-DNA and LR-DNA complexes (Figures 1C and 1D) provide parameters and boundary conditions. The specifics of the model and the structures used are detailed in Modeling Procedures.

Prior to studying the effect of CAP binding, the elastic rod model of the “empty” O_1 - O_3 loop, created by LR in the binary complex with the DNA (without CAP), is built. The equations of elasticity are solved with constant elastic moduli and intrinsic twist, and zero intrinsic curvature. The boundary conditions for the problem are obtained from the hybrid all-atom structure of the LR-DNA complex, constructed, as detailed in Modeling Procedures, from the available X-ray and NMR structures including the structure from Lewis et al. (1996) used in

our previous work (Balaeff et al., 1999). As before, the boundary conditions are derived from the closest ends of the DNA segments bound to the LR. While other loop topologies are conceivable (Friedman et al., 1995), the chosen boundary conditions result in a wide open DNA loop pointing away from the LR that offers the best opportunity for CAP placement.

The resulting loop structures are not essentially different from those previously obtained for the 1LBG structure of the LR (Balaeff et al., 1999, 2003). Two main conformations of the loop exist: one, underwound by 1.7° /bp on the average, and another, overwound by 2.6° /bp on the average. The two conformations will be referred to below as U and O loops, respectively. The U loop is straight, open, and almost planar (Figure 3A), and the O loop bends over one of its ends, exhibiting a near self-crossing (Figure 3A'). The elastic energy of the U loop amounts to 24 kT, whereas that of the O loop amounts to 31 kT.

The energies of similar two loops, obtained in the case of the 1LBG structure, amount to 23 and 26.5 kT (Balaeff

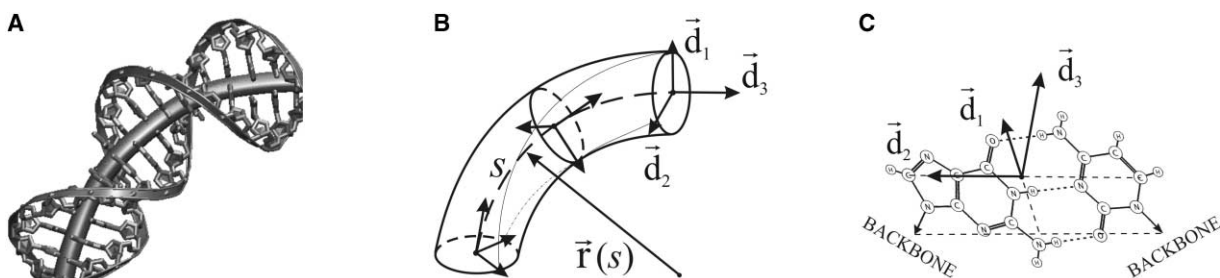


Figure 2. Elastic Rod Model of DNA

(A) The elastic rod fitted into an all-atom structure of DNA.

(B) Parameterization of the elastic rod: shown are the centerline $\vec{r}(s)$ and the local coordinate frame $(\vec{d}_1, \vec{d}_2, \vec{d}_3)$ associated with the elastic rod cross-section.

(C) A coordinate frame, associated with a DNA base pair according to Olson et al. (2001), allows one to align the base pair with an elastic rod cross-section and vice versa.

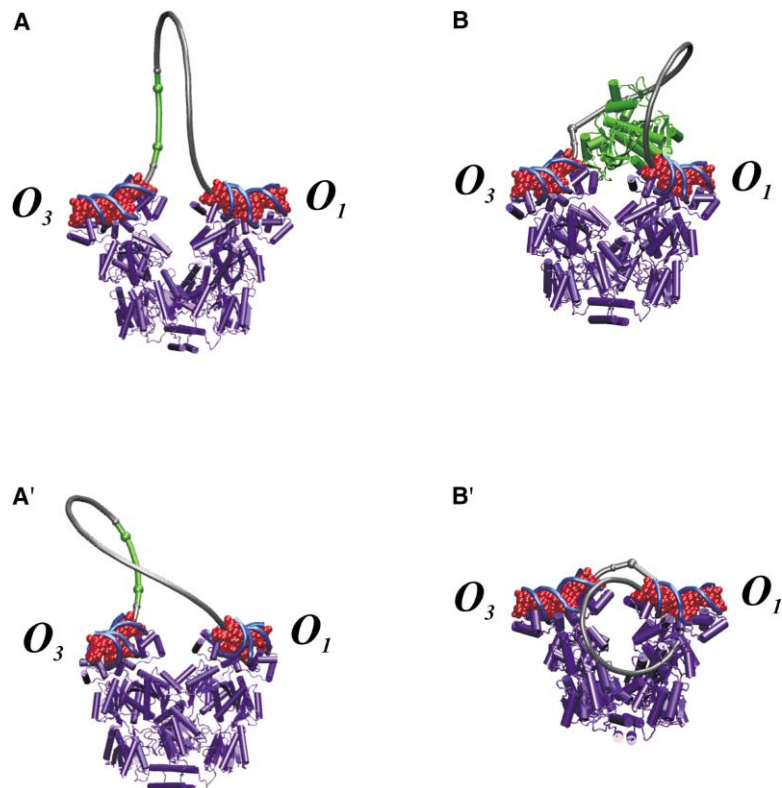


Figure 3. The Effect of Mimicking the CBS Structure within the O_1 - O_3 Loop

(A) U and (A') O elastic rod solutions for the “empty” O_1 - O_3 loop (Balaeff et al., 1999). The elastic rod centerlines are shown as gray tubes connecting the DNA segments in the hybrid structure of the LR-DNA complex (cf. Figure 1D); the CBS is shown as a green segment with two spheres indicating the kink sites. (B) U and (B') O loops with the mimicked CBS (cf. Figures 1C and 8). Crystal structure of CAP from Parkinson et al. (1996) is shown in (B) fitted within the CBS in the U loop. The structure of the O loop (B') sterically overlaps with the LR and clearly cannot accommodate CAP.

et al., 1999, 2003). For comparison, the experimental looping energy is estimated to be 20 kT from the results of kinetic and equilibrium binding experiments (Balaeff et al., 2003; Hsieh et al., 1987)—which is comparable to the computed U loop energy, but is still a few kT away. Similarly, the model produces variations in the energy on the order of several kT depending on the exact choice of elastic rod model parameters (Balaeff et al., 2003). Therefore, the energy values obtained in this work should allow for mostly qualitative rather than quantitative interpretation. However, even this level of precision of the model can furnish significant results. For example, the energy difference between the U and the O loop is sufficiently large to safely conclude that the U loop should be predominant in the conformational ensemble under conditions of thermal equilibrium, especially that the energy difference is consistent throughout a broad range of the problem parameters (Balaeff et al., 2003).

Structural Changes Due to the Mimicked CAP Binding

The structure of the CBS observed in the CAP-DNA complex (Parkinson et al., 1996) is mimicked within the O_1 - O_3 loop using special intrinsic curvature and twist terms in the equations of elasticity, as described in Modeling Procedures. The resulting structures of the U and O loops are shown in Figures 3B and 3B'. The U loop remains a wide open structure, inside which CAP could be fitted with minimal steric overlaps, as discussed below. The energy of the loop, excluding the modified CBS, increases to 26 kT. In contrast, the O loop, while preserving its energy of 31 kT, changes into a sterically

prohibitive structure. The centerline of the O loop runs through the O_1 DNA segment and the corresponding LR hand (Figure 3B'). Such structure is clearly unrealistic and is generated only because our elastic rod computations deal with a phantom LR-DNA complex, which provides the boundary conditions, but does not influence the computations in any other way. Were a proper steric repulsion term used, the resulting O loop would have a different, significantly stressed structure and a very high energy. Therefore, the U loop appears at this point to be the only candidate for the ternary complex of the CAP, LR, and the O_1 - O_3 DNA loop.

In order to position CAP within the constructed U loop, we build an idealized all-atom DNA structure on top of the elastic rod structure of the loop, as described in Modeling Procedures. Then the crystal structure from Parkinson et al. is fitted inside the mimicked CBS by aligning the DNA kinks in that structure with their counterparts in the all-atom U loop, after which the crystallographic DNA is discarded. The resulting ternary complex of CAP, LR, and DNA is shown in Figure 3B.

The ternary structure accommodates CAP with only a small steric overlap (to the depth of a few angstroms) between the N-terminal part of the protein and the O_1 DNA segment. This is a very good result, considering that CAP was not really included in the computations, except through mimicking its binding site. A perfect fit with no overlap can be achieved by either slightly modifying the CBS parameters, or by adding a force term accounting for the CAP- O_1 steric repulsion to the equations of elasticity. Either way, the structure of the loop should not change much. For example, the ternary complex, shown in Figure 4, is built for the values of CBS

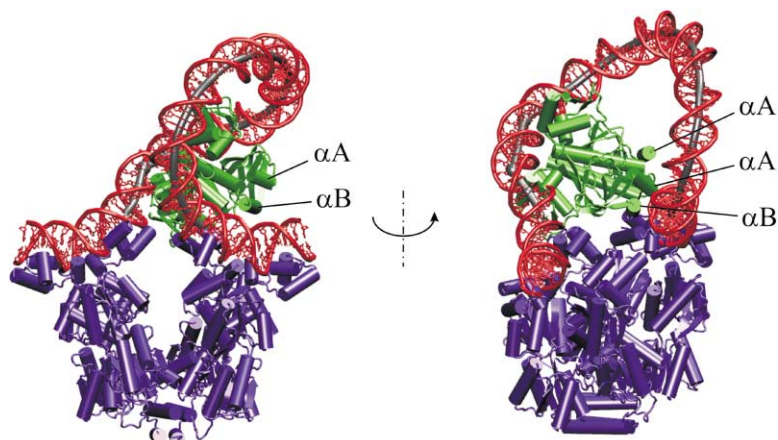


Figure 4. All-Atom Structure of the Ternary Complex of CAP and LR with the U Loop

The idealized all-atom DNA structure is built on top of the U loop; the crystal structure of CAP from Parkinson et al. (1996) is fitted inside the mimicked CBS (cf. Figure 3B). The right view is rotated by 90° around the vertical axis with respect to the left view. The better (compared to that in Figure 3B), sterically unimpeded fit of CAP inside the U loop results from using the CBS parameters averaged between the two kinks in the structure of Parkinson et al. (1996). Two α helices αA (residues 9–17) and αB (residues 99–107), located near the N terminus of the CAP, are seen in close proximity to DNA and therefore provide a potential target site for mutagenesis experiments. Positively charged residues, introduced into those helices, are likely to stabilize the predicted U conformation of the ternary complex.

kinks and unwinding angles averaged between the two kinks in the structure of Parkinson et al. The resulting ternary complex is very similar to the one originally built (cf. Figure 3B), yet accommodates CAP without a steric overlap.

CAP-LR Cooperation in the Ternary Complex with the U Loop

In order to assess the plausibility of the constructed ternary complex, we estimate the free energy of cooperation ΔG_{coop} between CAP and LR, experimentally measured to be within 1–2 kT (Hudson and Fried, 1990). Once again, the uncertainties of our model do not allow us to compute the energies with the precision of a single unit of kT, so a negative ΔG_{coop} within a few kT from the experimental value would be considered a good match.

One can express

$$\Delta G_{coop} = \Delta G_{Lac+CAP} - (\Delta G_{Lac} + \Delta G_{CAP}), \quad (1)$$

where $\Delta G_{Lac+CAP}$ is the free energy of the ternary complex and ΔG_{Lac} and ΔG_{CAP} are the free energies of individual LR and CAP complexes with DNA. ΔG_{Lac} is decomposed as $\Delta G_{Lac} = \Delta G_{O_1} + \Delta G_{O_3} + U_{loop}$, where ΔG_{O_1} and ΔG_{O_3} are the free energies of interaction between the LR hands and O_1 and O_3 , respectively, and U_{loop} is the free energy of the DNA loop. The free energy of the ternary complex is similarly decomposed as $\Delta G_{Lac+CAP} = \Delta G'_{CAP} + \Delta G'_{O_1} + \Delta G'_{O_3} + U'_{loop}$, where the prime indicates the energies of the protein-DNA interactions within the ternary complex as opposed to those in the individual complexes. The term U'_{loop} does not include the energy of the CBS kinks, because it is already included in $\Delta G'_{CAP}$.

Therefore, the free energy of cooperation between the two proteins is

$$\Delta G_{coop} = \Delta \Delta G_{CAP} + \Delta \Delta G_{O_1} + \Delta \Delta G_{O_3} + \Delta U_{loop}, \quad (2)$$

where $\Delta \Delta G_i = \Delta G'_i - \Delta G_i$ and $\Delta U_{loop} = U'_{loop} - U_{loop}$.

Let us estimate the terms in equation 2 in the case of the U loop. The elastic rod calculations yield $\Delta U_{loop} = 26\text{kT} - 24\text{kT} = 2\text{kT}$. One can assume $\Delta \Delta G_{O_1} \approx 0$, since the O_1 -bound hand of the LR is little disrupted by CAP (cf. Figures 3B and 4). The values of $\Delta \Delta G_{O_3}$ and $\Delta \Delta G_{CAP}$ are less certain, but can be expected to drive ΔG_{coop} up

by several kT, because the adjacent ends of CBS and O_3 should be stressed due to the abrupt change in DNA geometry between them, thereby disrupting the protein-DNA interactions at both sites.

On the other hand, the N-terminal part of CAP carries an excess of positively charged residues and should interact favorably with the nearby O_1 DNA (Figure 4). We estimate the resulting contribution to $\Delta G'_{CAP}$ and $\Delta \Delta G_{CAP}$ to be between -2 and -4 kT, assuming an ionic strength between 0 and 100 mM, all the residues of CAP protonated as under $\text{pH} = 7$, and the DNA charge reduced to 0.25 e per phosphate due to Manning counterion condensation (Manning, 1978). The new contribution partially offsets the other terms in equation 2, and the resulting sign of ΔG_{coop} is unclear. Therefore, since the present model can not estimate ΔG_{coop} with a sufficient precision, the validity of the constructed U conformation of the ternary complex is uncertain.

Ternary Complex with the O Loop of Variable Length

The unrealistic ternary complex structure obtained for the O loop indicates that such a complex cannot exist. However, varying the length L of the loop can have a dramatic effect on the loop energy and structure. Physically, the increase in L corresponds to sliding of the O_3 binding hand of LR upstream the DNA, as if being pushed by the incoming CAP.

The effect of increasing L from the original 76 to 86 bp is presented in Figure 5. The increase in L is accompanied by a significant drop in the elastic energy, which reaches the minimum of 10 kT at $L = 83$ bp (Figure 5A). Structurally, the loop flips upwards and becomes wide open, so that CAP can be easily accommodated (Figures 5B and 6).

Such a picture perfectly conforms with the results of DNase I footprinting experiments (Perros et al., 1996). It has been observed that after the ternary complex is formed, the O_3 binding hand of LR gets relocated upstream the DNA, apparently, due to the disturbance from the nearby-bound CAP. The estimated relocation distance (and the resulting increase in L) is 6 bp, in good agreement with our computations.

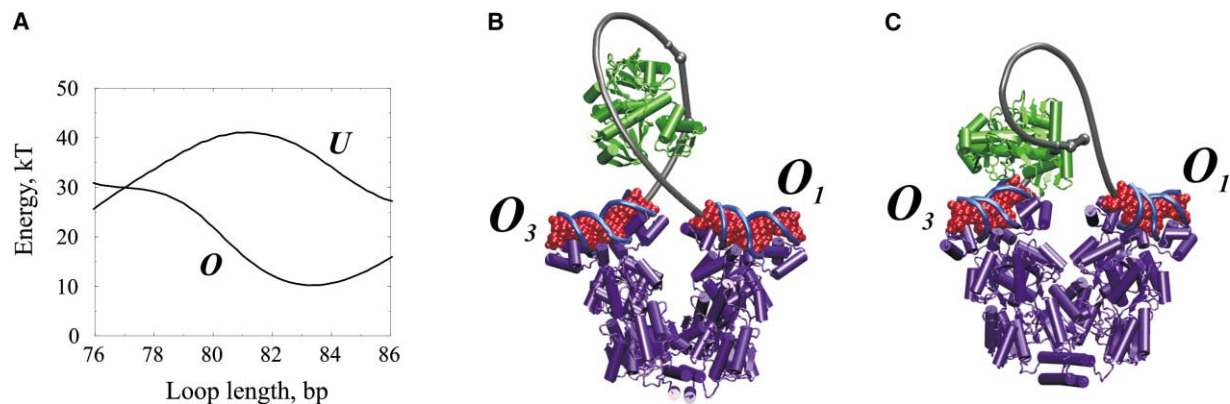


Figure 5. Changes in the Structure and Energy of the O and U Loops, Containing the Mimicked CBS, Due to Varying the Length L of the Loop (A) Elastic energies of the O and U loops (as indicated), plotted versus L in the range of 76–86 bp, reveal that increasing L to 82–84 bp is energetically favorable for the O loop, but not for the U loop. (B) The structure of the 83 bp long O loop is wide open and can easily accommodate CAP. (C) The structure of the 83 bp long U loop accommodates CAP with steric conflicts and displays a prohibitive near self-crossing.

The significant decrease of U_{loop} results in a much better cooperation free energy than that estimated for the U loop, even though the O_3 binding hand of LR is now relocated from its preferred binding site. The corresponding increase in $\Delta\Delta G_{O_3}$ is estimated from the experimental data (Barkley and Bourgeois, 1980) as $\Delta\Delta G_{O_3} = kT \log K_{n/s} - kT \log K_{O_3} = 8$ to 14 kT, where $K_{O_3} = 10^{-11}M$ and $K_{n/s} \approx 2 \cdot 10^{-6}$ to $1 \cdot 10^{-3}M$ are the equilibrium constants of the LR binding to O_3 and to nonspecific DNA, respectively. That is a significant increase; however, the decrease in U_{loop} amounts to $\Delta U_{loop} = -20$ to -21 kT for $L = 82$ to 84 bp (cf. Figure 5A). Finally, one can expect $\Delta\Delta G_{O_1} \approx 0$ and $\Delta\Delta G_{CAP} \approx 0$; the latter because CAP, fitted inside the mimicked CBS, is now well removed from the LR and shows no unusual contacts with DNA (Figure 5B).

The resulting energy balance (2) is then

$$\begin{aligned} \Delta G_{coop} &\approx 0 + 0 + (8 \text{ to } 14 \text{ kT}) - (20 \text{ to } 21 \text{ kT}) \\ &= -6 \text{ to } -13 \text{ kT}, \end{aligned} \quad (3)$$

clearly suggesting that the formation of the ternary com-

plex with the lengthened O loop is energetically more favorable than the formation of the complex with the 76 bp U loop. An all-atom structure built for the ternary complex with the 83 bp long O loop is shown in Figure 6.

In contrast, the ternary complex with the U loop is not improved by a similar increase in L . When the U loop becomes longer, its elastic energy goes up rather than down, reaching a maximum of about 40 kT at $L = 81$ bp (Figure 5A). The structure of the loop becomes conformationally prohibitive, showing near overlaps first with itself, as seen in Figure 5C, and then with the O_1 and the LR hand, similarly to the 76 bp long O loop. Neither can CAP be fitted inside most of the lengthened U loops without severe steric overlaps. The ternary U complex with the 76 bp long loop is therefore the most likely among all possible U complexes, and yet that structure is clearly inferior to the ternary complex with the lengthened O loop.

Conformational Switch

In conclusion, we predict that the CAP-LR-DNA complex with the O loop, lengthened to 82–84 bp, is the most

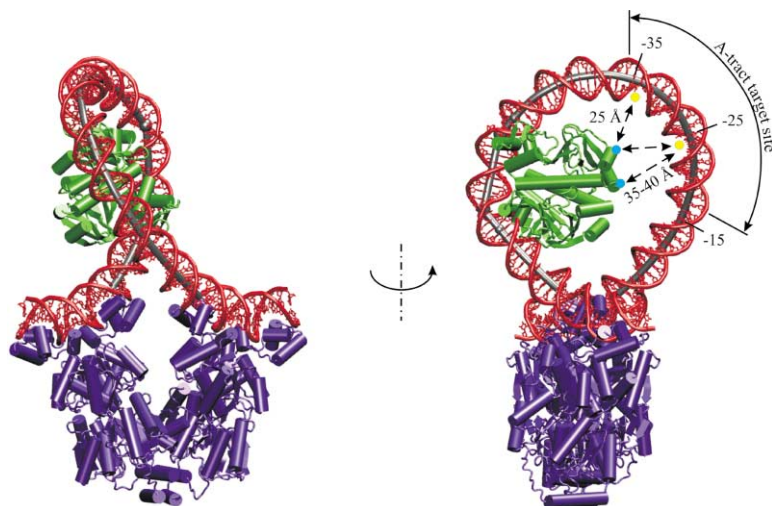


Figure 6. All-Atom Structure of the Ternary Complex of CAP and LR with the 83 bp Long O Loop

The right view is rotated by 90° around the vertical axis with respect to the left view. The N-terminal helix bundle $\alpha A/\alpha B$ (cf. Figure 4) comes within 25–40 Å of the normally remote DNA segments, as indicated. This change in distance upon CAP binding to the LR-DNA complex is detectable by FRET; yellow and blue dots indicate representative sites for fluorophore labels attachment. A semicircular arrow indicates the loop section with approximately uniform curvature. Introducing properly phased A-tracts into that section of the loop should increase the stability of the predicted ternary O complex.

energetically favorable conformation of the ternary complex, whereas the U loop is the preferred loop conformation in the absence of CAP. Thus, it is possible that CAP binds to the U loop and forces the conformational switch to the O loop. In that case one would expect $\Delta U_{loop} = U'_{O\ loop} - U_{U\ loop} = 10\text{ kT} - 24\text{ kT} = -14\text{ kT}$ and (cf. equation 3)

$$\begin{aligned}\Delta G_{coop} &= 0 + 0 + (8\text{ to }14\text{ kT}) - 14\text{ kT} \\ &= 0\text{ to }-6\text{ kT}.\end{aligned}\quad (4)$$

Even considering the inaccuracies of the present model, the obtained value of ΔG_{coop} clearly shows that the conformational switch is energetically favorable. The experimental estimate of ΔG_{coop} (-1 to -2 kT) falls within the predicted range. Thus, we can conclude that the cooperativity in DNA binding between CAP and LR results from the significant drop in the elastic energy of the DNA loop which occurs due to the conformational switch coupled with CAP binding.

Outlook

The constructed structures of the ternary complex of CAP, LR, and DNA explain the cooperative binding by CAP and LR on the structural level and predict the changes in the CAP-DNA and LR-DNA complexes upon CAP binding. These predictions can be tested in a number of possible experiments. For example, measurements of the efficiency of fluorescence resonance energy transfer (FRET) between two fluorophore labels attached near the N terminus of CAP and at a certain point on the DNA loop can be conducted (Edelman et al., 2003; and references therein). One possible target for the label could be the section of the loop between base pairs -15 and -25 , which is presumably distant from the N-terminal part of CAP by more than 100 \AA in the binary CAP-DNA complex, where the DNA is not looped, but comes within $35\text{--}40\text{ \AA}$ in the extended O complex (Figure 6). The same distance in the ternary U complex equals $50\text{--}70\text{ \AA}$. Therefore, FRET experiments may help to verify whether the preferred conformation of the ternary complex is indeed the O conformation predicted here.

In another experiment, the DNA loop in the ternary complex can be stabilized or destabilized by replacing predicted curved sections of it with intrinsically bent A-tracts, either in phase with or oppositely to the predicted bend. A similar study recently addressed the DNA loops clamped by the LR alone (Mehta and Kahn, 1999). Comparing the observed degree of stabilization by the A-tracts to the expected effect may serve to confirm or disprove the predicted model. Mutations may also be introduced into CAP in order to manipulate the relative stability of the U and O complexes. For instance, replacing a neutral or a polar amino acid residue in the N-terminal α helix 9–17, or the parallel α helix 99–107 (Figure 4) with a positively charged residue should increase the stability of the U complex. Finally, the distribution of protected and hypersensitive DNase I cleavage sites in the $O_1\text{--}O_3$ loop, available from the footprinting experiments (Hudson and Fried, 1990; Perros et al., 1996), may be compared with that computed using the predicted geometry of the loop.

All our conclusions, however, are subject to the valid-

ity of the assumption that the structure of the LR in the ternary complex is not significantly different from the V shape observed in the crystal by Lewis et al. (1996) (Figure 1D). The protein-DNA interfaces, both CAP-CBS and LR- $O_{1/3}$, can also be seriously disrupted by the ternary complex formation, enough to alter the predicted structure of the system. Such structural changes are beyond the reach of the present model; however, they can be readily addressed by extending our multiresolution approach. For example, molecular dynamics simulations of the all-atom structure of the LR complex with O_1 and O_3 can be conducted using the forces and torques obtained from the elastic rod model, akin to how it is done in steered molecular dynamics simulations (Isralewitz et al., 2001). The forces and torques would be iteratively updated during the course of the simulations in response to the changes of the LR-DNA complex and, consequently, in the boundary values for the elastic rod problem. A similar multiscale simulation of the CAP-CBS complex, or of any other interesting part of the $O_1\text{--}O_3$ loop, can be conducted, resulting in a more consistent description of the structure and dynamics of the CAP-LR-DNA complex. With the advent of massively parallel computers, even the molecular dynamics simulations of the all-atom structures of the whole ternary complex, such as those shown in Figures 4 and 6, can eventually be performed.

Such advanced simulations, supported by data from and providing ideas for multiple experimental studies, such as those described above, have a great potential to describe in unprecedented detail the properties of large protein-DNA complexes, such as the ternary complex of CAP, LR, and DNA.

In summary, a multiscale structure of the ternary complex of CAP, LR, and a DNA loop was built on the basis of the elastic rod model of the loop. The energetics of the loop, altered by the bound CAP, favors the relocation of one of the DNA binding hands of the LR by 6–8 bp upstream, in agreement with footprinting experiments. It is predicted that the protein hand relocation, in turn, results in switching of the preferred loop conformation from the underwound to the overwound state; the resulting gain in elastic energy appears to be the driving force behind the experimentally observed cooperation in DNA binding between CAP and LR. The predicted structure, together with multiscale simulations to which the structure opens the path, provides a fertile ground for numerous interesting experiments that can test the predictions and further advance our understanding of the CAP-LR-DNA complex.

Modeling Procedures

LR and CAP Structures

The Protein Data Bank (<http://www.rcsb.org/pdb>) contains a number of all-atom structures of both the whole LR and its smaller parts, including several complexes with DNA. The structure from Lewis et al. (1996) (PDB code 1LBG) includes the whole LR binding two operator DNA segments; that structure has been used in our previous work (Balaeff et al., 1999, 2003). However, the protein in 1LBG contains only $^{\circ}\text{C}$ atoms; the side chains are not resolved. Therefore, in view of future multiscale and all-atom simulations of the complex (as discussed above) we constructed an all-atom structure of the LR-DNA complex using other relevant X-ray and NMR structures from the PDB.

Our hybrid structure is based on the X-ray structure from Bell and

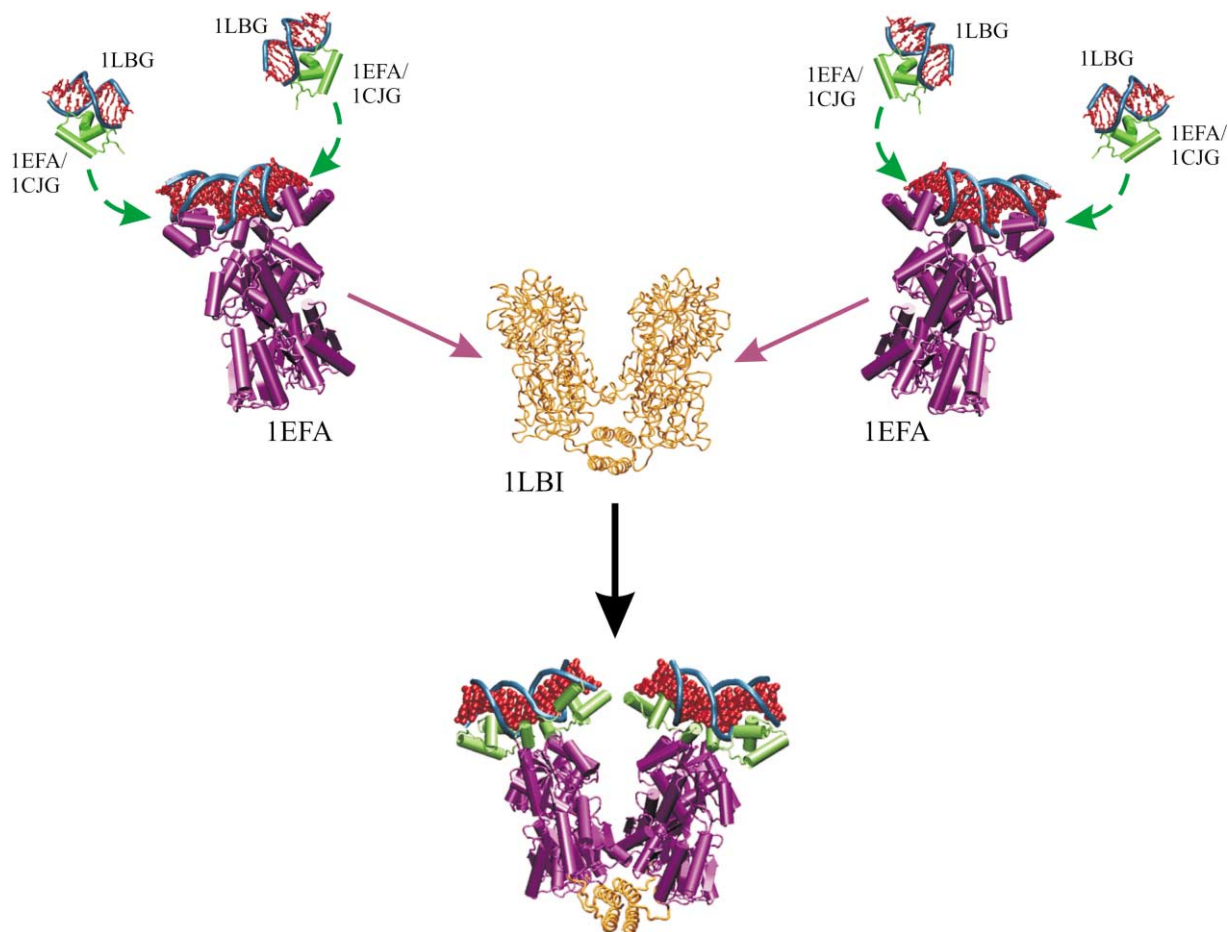


Figure 7. Construction of the Hybrid All-Atom Structure of LR from the Available PDB Structures

Two patched 1EFA structures (Bell and Lewis, 2000) (drawn as purple protein cartoon), are aligned with two halves of the 1LBI structure (Lewis et al., 1996) (drawn as orange tubes), and linked via the α -helical bundle taken from 1LBI. The LR headgroups (drawn in green) are patched with residues from the NMR structure 1CJG (Spronk et al., 1999) and combined with the DNA segments from the 1LBI structure (Lewis et al., 1996) (drawn as blue tubes and red spheres).

Lewis (2000) (PDB code 1EFA) of a complex between a single LR “hand” (residues 3–329) and an operator DNA. The construction of the hybrid structure is illustrated in Figure 7. First, several residues (1, 2, 31, 36, 37, and 44) missing in the LR headgroups in 1EFA are patched using the NMR structure from Spronk et al. (1999) (PDB code 1CJG, model #3) which fits the 1EFA headgroup better than all the other available headgroup structures. Then, the DNA half-segment (base pairs 4–11) contacting the headgroup is replaced by the half-segment from the 1LBI structure, since that segment fits the protein headgroup more tightly and better preserves the Watson-Crick structure of the base pairs. In the next step, two thus constructed headgroup-DNA complexes are aligned with the headgroups of the 1EFA structure and connected to the 1EFA protein core domains (residues 63–329). The protein backbone coordinates used for the alignment are averaged between the two halves of the protein dimer in 1EFA; in this way, the headgroups, tilted to one side in the 1EFA structure, become symmetrically oriented with respect to the core domains. The resulting mended 1EFA structure is duplicated and aligned with the two halves of the LR tetramer from Lewis et al. (1996) (PDB code 1LBI). Next, the two 1EFA halves are connected by the α -helical bundle obtained from the 1LBI tetramer (residues 330–357). The coordinates of the hydrogen atoms, missing in the X-ray structures, are built using X-PLOR (Brünger, 1992), and the completed all-atom hybrid structure is energy minimized using the CHARMM22 force field (MacKerell, 1998).

The CAP-DNA complex is also represented in the PDB by several structures. The calculations in this work are based on the structure

from Parkinson et al. (1996) (PDB code 1J59, Figure 1C). The protein in that structure binds the 30 bp consensus DNA and causes two primary kinks of 52° and 35° at two DNA steps located 5 bp away on each side from the center of the CBS. The kinked DNA steps are unwound by 17° and 22° , respectively. Two secondary kinks of -22° and -16° occur 6 bp upstream and downstream from the primary kinks. Those smaller kinks seem, however, to occur solely due to DNA sequence effects as protein-DNA contacts are virtually absent at those DNA steps. Therefore, we account only for the primary kinks in the present work.

Our calculations were also repeated for two alternative choices of CBS kink and unwinding angles. First, we employed the average values of the angles from the 1J59 structure; then, we used the angles corresponding to the structure of the CAP-DNA complex from Schultz et al. (1991) (PDB code 1CBG). The results of these tests were never essentially different from those of the original calculation, yielding only slightly different geometries of the ternary complex as mentioned above, and therefore are not described in detail in this manuscript.

Elastic Rod Model of DNA

The O_1 - O_3 DNA loop is approximated in this work by a flexible elastic rod, following the approach adopted in many theoretical studies (reviewed by Olson and Zhurkin, 2000; Olson, 1996; Schlick, 1995; Vologodskii and Cozzarelli, 1994). The parameters of the model that determine the mechanical properties of such a rod are inferred from the results of numerous experiments on DNA, such as electropho-

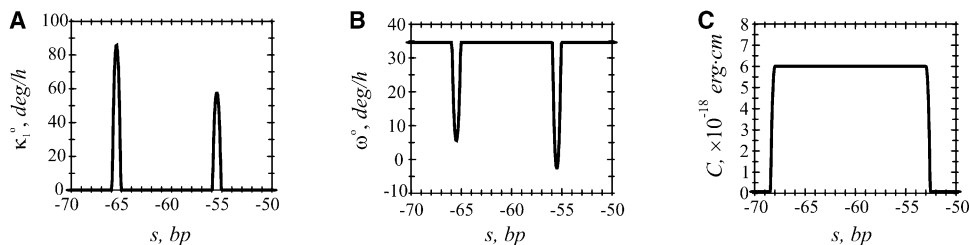


Figure 8. Parameters of the Elastic Rod Problem, Modified so as to Mimic the CBS Structure from Parkinson et al.

(A) Intrinsic curvature κ_1° . (B) Intrinsic twist ω° . (C) Twisting modulus C ; the profiles for the bending moduli A_1 and A_2 look similar and therefore are not shown. The parameters are plotted against the arclength s in the section $-75h$ to $-50h$ of the loop ($h = 3.4 \text{ \AA}$) which includes the mimicked CBS; outside the CBS, the parameters are constant.

retic mobility assays, light scattering, and cyclization kinetics (Olson, 1996; Schlick, 1995; Hagerman, 1988; and references therein), X-ray crystallography (Olson et al., 1998), and single molecule micromanipulation (Strick et al., 2000). The model was shown to correctly reproduce such DNA properties as, for example, the sedimentation coefficient (Rybenkov et al., 1997; Vologodskii and Cozzarelli, 1994), the equilibrium distribution of topoisomers (Katrich and Vologodskii, 1997; Vologodskii and Cozzarelli, 1994), and force-extension curves (Vologodskii and Marko, 1997). Our present implementation of the elastic rod model is outlined below and has been described in more detail elsewhere (Balaeff et al., 1999, 2003).

The elastic rod model of DNA is illustrated in Figure 2. The axis of the DNA helix corresponds to the rod centerline, a three-dimensional curve parameterized by the arc length s ; the Watson-Crick base pairs correspond to the cross-sections of the elastic rod. The geometry of the elastic rod is described in terms of its twist $\omega(s)$ and curvatures $\kappa_1(s)$, $\kappa_2(s)$ at each point s . The two curvatures describe the rod bending around the two principal axes of its cross-section and, in the case of DNA, correspond to the deformations of roll and tilt (Olson et al., 2001). The *intrinsic* components $\kappa_1^\circ(s)$, $\kappa_2^\circ(s)$, and $\omega^\circ(s)$ are separated from the curvatures and the twist: these parameters determine the structure of the relaxed elastic rod/DNA (Westcott et al., 1997; Yang et al., 1995; Olson, 1996). Our basic elastic rod model employs $\kappa_1^\circ = \kappa_2^\circ = 0^\circ$ and $\omega^\circ = 34.6^\circ$ per base pair.

The elastic response of the rod to any changes from its relaxed geometry is expressed via forces $\vec{N}(s)$ and torques $\vec{M}(s)$. The main parameters of the model, the elastic moduli of the rod bending (A_1 , A_2) and twisting (C), linearly relate the elastic torques to the changes in the curvatures and the twist: $M_{i=1,2} = A_i(\kappa_i - \kappa_i^\circ)$, $M_3 = C(\omega - \omega^\circ)$. Accordingly, the elastic energy U of the rod is a quadratic form in the curvatures and twist: $dU/ds = A_1(\kappa_1 - \kappa_1^\circ)^2/2 + A_2(\kappa_2 - \kappa_2^\circ)^2/2 + C(\omega - \omega^\circ)^2/2$. In this work, the values of $A_1 = 0.8 \cdot 10^{-19} \text{ erg} \cdot \text{cm}$, $A_2 = 2.4 \cdot 10^{-19} \text{ erg} \cdot \text{cm}$, and $C = 3 \cdot 10^{-19} \text{ erg} \cdot \text{cm}$ are used, except inside the mimicked CBS (cf. Figure 8C). These moduli correspond to the experimentally measured persistence lengths of 500 Å for DNA bending and 750 Å for DNA twisting (Hagerman, 1988; Strick et al., 1996; see Balaeff et al., 2003 for detailed discussion).

The model of the ternary complex is built in this work on the basis of equilibrium structures of the O_1 - O_3 loop, i.e., elastic loops with such geometries that the elastic forces and torques are balanced at each point s (Balaeff et al., 1999; Mahadevan and Keller, 1993). Such structures essentially correspond to zero temperature, i.e., entropic effects are neglected. The resulting error is presumably small because the DNA loops, studied here, are shorter than a single persistence length. The equilibrium loop structures are obtained by solving Kirchhoff equations of elasticity, a 13th order system of ordinary differential equations resulting from the linear approximation for the torques, the equations for force and torque balance, and the condition of inextensibility of the rod (Balaeff et al., 2003; Westcott et al., 1997; Mahadevan and Keller, 1993). The extensibility/deformability of the elastic rod is neglected here but, in principle, can be included in the model, resulting in a higher order system of Kirchhoff equations (Westcott et al., 1997; Shi et al., 1995).

The boundary conditions for the problem, obtained from the constructed hybrid structure of the LR-operator complex, consist of the coordinates of each end of the centerline and the orientation of

the rod cross-sections at those ends. The boundary value problem is solved by an iterative algorithm, whereby an initial simplified solution is gradually modified in order to achieve the desired solution (Balaeff et al., 1999, 2003; Mahadevan and Keller, 1993).

The elastic rod conformations obtained upon solving the Kirchhoff equations result in direct estimates of the structure and energy of the DNA loop, as well as forces and torques at each point of the loop. Simple geometric transformations of the loop ends allow one to generate the ensemble of possible loop conformations and select the ones with the lowest energy (Balaeff et al., 1999, 2003). The elastic rod structures are used as scaffolds for idealized all-atom models of the DNA loop; such models are built by aligning all-atom Watson-Crick base pairs with the cross-sections of the elastic rod at the appropriate points along the centerline (cf. Figure 2) and then energy minimizing the resulting structure with an all-atom force field, e.g., CHARMM22/27 (MacKerell, 1998).

The elastic rod model employed in the present study is rather basic. More advanced models could include elastic moduli and/or intrinsic curvature and twist terms varying along the loop according to the local DNA sequence (Olson et al., 1996, 1998; Hogan and Austin, 1987), DNA deformability terms (Westcott et al., 1997; Shi et al., 1995), and electrostatic and van der Waals repulsion (Balaeff et al., 2003; Coleman et al., 2000; Westcott et al., 1997). We do not believe that using such more realistic yet significantly more complex models is justified here. On the one hand, there are more serious unknown factors in the problem, such as the possible difference of the LR structure in the ternary complex from its crystal structure by Lewis et al. (1996) on which our model is based (Edelman et al., 2003). On the other hand, the more complex models would likely change some minor features of the ternary complex, such as the exact orientation and positioning of CAP, but not the more general results, such as which topology is preferred by the loop in the binary and ternary complexes with LR and LR CAP, respectively.

Mimicking the CAP Binding Site

In order to mimic the structure of the CBS within the O_1 - O_3 loop, we modify the intrinsic curvature and twist parameters $\kappa_{i=1,2}^\circ(s)$ and $\omega^\circ(s)$ and the elastic moduli $A_1(s)$, $A_2(s)$, and $C(s)$ in the section of the loop corresponding to the CBS (Figures 1A, 1B, and 3). The intrinsic curvature and twist parameters set up the desired kinking and unwinding angles, observed in the CAP-DNA crystal structure (Parkinson et al., 1996), while the elastic moduli are increased along the CBS in order to render the desired intrinsic structure “frozen,” effectively unchangeable by the elastic forces.

The profiles of the modified parameters along the O_1 - O_3 loop are shown in Figure 8. The intrinsic curvature κ_1° is zero everywhere except for the DNA steps $-67/-66$ and $-57/-56$ that are kinked by the bound CAP (Figure 8A, cf. Figures 1B and 1C). Over those steps, the intrinsic curvature is raised as a smooth bell-shaped function which results in the observed kink angles

$$\kappa_1^\circ(s) = K_0 \exp\left(\frac{(s - s_0)^2}{(s - s_0)^2 - (d/2)^2}\right), \text{ if } |s - s_0| \leq d/2. \quad (5)$$

Here d is the width of the kinks, set equal to one DNA helical step $h = 3.4 \text{ \AA}$. The constant K_0 is chosen such that

$$\int_{s_0-d/2}^{s_0+d/2} \kappa_2^{\circ}(s) ds = \phi_0 = 52^{\circ} \text{ and } 35^{\circ}$$

the kink angles observed by Parkinson et al. (1996). The two kinks are centered at the points $s_0 = -66.5h$ and $s_0 = -56.5h$, respectively.

The intrinsic curvature in the second principal direction, κ_2° , is set to zero, because the kinks caused by CAP exhibit themselves predominantly as the roll angles (Parkinson et al., 1996). The intrinsic twist ω° is smoothly decreased over the mimicked kinks, similarly to the increase in κ_1° (Figure 8B), in order to enforce the unwinding by 17° and 22° observed at the kinked DNA steps (Parkinson et al., 1996).

The "stiff" elastic moduli are set up in the loop section from $-69h$ to $-52h$ (Figure 8C), which approximately coincides with that part of the CBS that exhibits direct protein-DNA contacts (Parkinson et al., 1996) (cf. Figures 1B and 1C). A 20-fold increase in the moduli to $A_{1\text{ stiff}} = A_{2\text{ stiff}} = C_{\text{stiff}} = 6 \cdot 10^{-18} \text{ erg} \cdot \text{cm}$ proved to be sufficient to keep the desired geometry of the CBS intact. The transition between the stiff and the regular elastic moduli zones is accomplished by a smooth connecting function similar to the left or the right part of the bell-shaped function (equation 5).

Kirchhoff equations with the modified parameters are solved using two additional steps of the iterative procedure starting with the solutions for the "empty" U and O loops (Figures 3A and 3A'). In the first step, the elastic moduli over the CBS are gradually increased to the chosen stiff values. In the second step, the intrinsic curvature and twist over the mimicked kinks are gradually changed until they adapt the designed profiles.

Acknowledgments

This work was supported by the grants from Roy J. Carver Charitable Trust, NIH (PHS 5 P41 RR05969), and NSF (BIR 94-23827EQ). The figures in this paper were prepared using the molecular visualization program VMD (Humphrey et al., 1996).

Received: May 29, 2003

Revised: September 9, 2003

Accepted: September 16, 2003

Published: January 13, 2004

Bibliography

- Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., and Walter, P. (2002). *The Cell*, Fourth Edition (New York & London: Garland Science).
- Balaeff, A., Mahadevan, L., and Schulten, K. (1999). Elastic rod model of a DNA loop in the *lac* operon. *Phys. Rev. Lett.* **83**, 4900-4903.
- Balaeff, A., Mahadevan, L., and Schulten, K. (2003). Modeling DNA loops using the theory of elasticity. E-print archive arXiv.org (<http://arxiv.org/abs/physics/0301006>), pub. no. 0301006.
- Barkley, M.D., and Bourgeois, S. (1980). Repressor recognition of operator and effectors. In *The Operon*, J.H. Miller and W. S. Reznikoff, eds. (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press), pp. 177-220.
- Bell, C.E., and Lewis, M. (2000). A closer view of the conformation of the Lac repressor bound to operator. *Nat. Struct. Biol.* **7**, 209-214.
- Berg, J.M., Tymoczko, J.L., and Stryer, L. (2002). *Biochemistry*, Fifth Edition (New York: W.H. Freeman and Co.).
- Brünger, A.T. (1992). X-PLOR, Version 3.1: A System for X-ray Crystallography and NMR (New Haven, CT: Yale University Press).
- Busby, S., and Ebricht, R.H. (1999). Transcription activation by catabolite activator protein (CAP). *J. Mol. Biol.* **293**, 199-213.
- Coleman, B.D., Swigon, D., and Tobias, I. (2000). Elastic stability of DNA configurations. II. Supercoiled plasmids with self-contact. *Phys. Rev. E* **61**, 759-770.
- Edelman, L.M., Cheong, R., and Kahn, J.D. (2003). Fluorescence resonance energy transfer over ≈ 130 basepairs in hyperstable lac repressor-DNA loops. *Biophys. J.* **84**, 1131-1145.
- Friedman, A.M., Fischmann, T.O., and Steitz, T.A. (1995). Crystal structure of *lac* repressor core tetramer and its implications for DNA looping. *Science* **268**, 1721-1727.

- Hagerman, P.J. (1988). Flexibility of DNA. *Annu. Rev. Biophys. Biochem. Chem.* **17**, 265-286.
- Hogan, M.E., and Austin, R.H. (1987). Importance of DNA stiffness in protein-DNA binding specificity. *Nature* **329**, 263-266.
- Hsieh, W., Whitson, P.A., Mathews, K.S., and Wells, R.D. (1987). Influence of sequence and distance between two operators on interaction with the *lac* repressor. *J. Biol. Chem.* **262**, 14583-14591.
- Hudson, J.M., and Fried, M.G. (1990). Co-operative interactions between the catabolite gene activator protein and the *lac* repressor at the lactose promoter. *J. Mol. Biol.* **214**, 381-396.
- Humphrey, W., Dalke, A., and Schulten, K. (1996). VMD: visual molecular dynamics. *J. Mol. Graph.* **14**, 33-38.
- Israelowitz, B., Gao, M., and Schulten, K. (2001). Steered molecular dynamics and mechanical functions of proteins. *Curr. Opin. Struct. Biol.* **11**, 224-230.
- Kapanidis, A.N., Ebricht, Y.W., Ludescher, R.D., Chan, S., and Ebricht, R.H. (2001). Mean DNA bending angle and distribution of DNA bending angles in the CAP-DNA complex in solution. *J. Mol. Biol.* **312**, 453-468.
- Katrich, V., and Vologodskii, A. (1997). The effect of intrinsic curvature on conformational properties of circular DNA. *Biophys. J.* **72**, 1070-1079.
- Lewis, M., Chang, G., Horton, N.C., Kercher, M.A., Pace, H.C., Schumacher, M.A., Brennan, R.G., and Lu, P. (1996). Crystal structure of the lactose operon repressor and its complexes with DNA and inducer. *Science* **271**, 1247-1254.
- MacKerell, A.D., Jr., Brooks, B., Brooks, C.L., III, Nilsson, L., Roux, B., Won, Y., and Karplus, M. (1998). CHARMM: the energy function and its parameterization with an overview of the program. In *The Encyclopedia of Computational Chemistry*, P. Schleyer, ed. (Chichester, UK: John Wiley & Sons), pp. 271-277.
- Mahadevan, L., and Keller, J.B. (1993). The shape of a Möbius band. *Proc. R. Soc. Lond. A.* **440**, 149-162.
- Manning, G.S. (1978). The molecular theory of polyelectrolyte solutions with applications to the electrostatic properties of polynucleotides. *Q. Rev. Biophys.* **2**, 179-246.
- Mathews, K.S. (1992). DNA looping. *Microbiol. Rev.* **56**, 123-136.
- Mehta, R.A., and Kahn, J.D. (1999). Designed hyperstable lac repressor-DNA loop topologies suggest alternative loop geometries. *J. Mol. Biol.* **294**, 67-77.
- Oehler, S., Eismann, E.R., Krämer, H., and Müller-Hill, B. (1990). The three operators of the *lac* operon cooperate in repression. *EMBO J.* **9**, 973-979.
- Olson, W.K. (1996). Simulating DNA at low resolution. *Curr. Opin. Struct. Biol.* **6**, 242-256.
- Olson, W.K., Gorin, A.A., Lu, X.-J., Hock, L.M., and Zhurkin, V.B. (1998). DNA sequence-dependent deformability deduced from protein-DNA crystal complexes. *Proc. Natl. Acad. Sci. USA* **95**, 11163-11168.
- Olson, W.K., and Zhurkin, V.B. (2000). Modeling DNA deformations. *Curr. Opin. Struct. Biol.* **10**, 286-297.
- Olson, W.K., Bansal, M., Burley, S.K., Dickerson, R.E., Gerstein, M., Harvey, S.C., Heinemann, U., Lu, X.-J., Neidle, S., Shakked, Z., et al. (2001). A standard reference frame for the description of nucleic acid base-pair geometry. *J. Mol. Biol.* **313**, 229-237.
- Parkinson, G., Wilson, C., Gunasekera, A., Ebricht, Y.W., Ebricht, R.E., and Berman, H.M. (1996). Structure of the CAP-DNA complex at 2.5Å resolution: a complete picture of the protein-DNA interface. *J. Mol. Biol.* **260**, 395-408.
- Perros, M., Steitz, T.A., Fried, M.G., and Hudson, J.M. (1996). DNA looping and *Lac* repressor-CAP interaction. *Science* **274**, 1929-1931.
- Rybenkov, V.V., Vologodskii, A.V., and Cozzarelli, N.R. (1997). The effect of ionic conditions on the conformations of supercoiled DNA. I. Sedimentation analysis. *J. Mol. Biol.* **267**, 299-311.
- Schlick, T. (1995). Modeling superhelical DNA: recent analytical and dynamic approaches. *Curr. Opin. Struct. Biol.* **5**, 245-262.
- Schultz, S.C., Schields, G.C., and Steitz, T.A. (1991). Crystal struc-

ture of a CAP-DNA complex: the DNA is bent by 90° . *Science* **253**, 1001–1007.

Shi, Y., Borovik, A.E., and Hearst, J.E. (1995). Elastic rod model incorporating shear and extension, generalized nonlinear Schrödinger equations, and novel closed-form solutions for supercoiled DNA. *J. Chem. Phys.* **103**, 3166–3183.

Spronk, C.A.E.M., Bovin, A.M.J.J., Radha, P.K., Melacini, G., Boelens, R., and Kaptein, R. (1999). The solution structure of Lac repressor headpiece 62 complexed to a symmetrical lac operator. *Struct. Fold. Des.* **7**, 1483–1492.

Strick, T.R., Allemand, J.-F., Bensimon, D., Bensimon, A., and Croquette, V. (1996). The elasticity of a single supercoiled DNA molecule. *Science* **271**, 1835–1837.

Strick, T.R., Allemand, J.-F., Bensimon, D., and Croquette, V. (2000). Stress-induced structural transitions in DNA and proteins. *Annu. Rev. Biophys. Biomol. Struct.* **29**, 523–543.

Vologodskii, A.V., and Cozzarelli, N.R. (1994). Conformational and thermodynamic properties of supercoiled DNA. *Annu. Rev. Biophys. Biomol. Struct.* **23**, 609–643.

Vologodskii, A.V., and Marko, J.F. (1997). Extension of torsionally stressed DNA by external force. *Biophys. J.* **73**, 123–132.

Westcott, T.P., Tobias, I., and Olson, W.K. (1997). Modeling self-contact forces in the elastic theory of DNA supercoiling. *J. Chem. Phys.* **107**, 3967–3980.

Yang, Y., Westcott, T.P., Pedersen, S.C., Tobias, I., and Olson, W.K. (1995). Effects of localized bending on DNA supercoiling. *Trends Biochem. Sci.* **20**, 313–319.