and some will have $a < a_{OLR}$. Because particles undergoing Lindblad resonant forcing on opposite sides of a resonance would have apsidal lines differing by 180°, a low optical depth ring (in which collisions are infrequent) composed of such particles would exhibit a width that varies over 360°/42° = 8.5714° from essentially zero (at the point where the particle orbits cross) to ~60 km (twice the radial distortion arising from the Lindblad forcing) and back again to zero. As there is an unmistakable and phased 30-km distortion across the arcs, we can dismiss the possibility that $\Delta a = W_r$.

18. D. Davis and E. Ryan, *Icarus* **83**, 156 (1990).

19. V. Haemmerle is acknowledged for extensive help in image data reduction and in the preparation of Fig. 3. I am indebted to P. Goldreich, S. Tremaine, W. Owen, J. Cuzzi, L. Dones, and an anonymous referee for valuable criticisms and suggestions following their readings of a preliminary draft of this article. In particular, I thank W. Owen for many illuminating discussions on his orbit modeling and for kindly providing his results on Neptune satellite orbits prior to publication, and J. Holberg for providing information on the Voyager UVS ring arc occultation. This research was supported by NASA grants no. NAGW-960 and no. NAGW-2342.

22 May 1991; accepted 17 July 1991

Crystal Structure of a CAP-DNA Complex: The DNA Is Bent by 90°

STEVE C. SCHULTZ, GEORGE C. SHIELDS,* THOMAS A. STEITZ

The 3 angstrom resolution crystal structure of the *Escherichia coli* catabolite gene activator protein (CAP) complexed with a 30–base pair DNA sequence shows that the DNA is bent by 90°. This bend results almost entirely from two 40° kinks that occur between TG/CA base pairs at positions 5 and 6 on each side of the dyad axis of the complex. DNA sequence discrimination by CAP derives both from sequence-dependent distortion of the DNA helix and from direct hydrogen-bonding interactions between three protein side chains and the exposed edges of three base pairs in the major groove of the DNA. The structure of this transcription factor–DNA complex provides insights into possible mechanisms of transcription activation.

SEVERE PROTEIN-INDUCED BENDING OF DUPLEX DNA HAS been demonstrated by a variety of biochemical and biophysical techniques (*1–10*), including a low-resolution crystal structure of the nucleosome core particle (*10*). Current high-resolution crystal structures of sequence-specific DNA binding proteins complexed with their DNA binding sites have shown relatively modest departures from straight canonical B-DNA (*11–14*), although the lower resolution (3.9 Å) crystal structure of the λ cro repressor–DNA complex shows an overall bend of ~40° (*15*). We describe a 90° bend in DNA bound specifically to the *E. coli* catabolite gene activator protein (CAP) as observed in a 3 Å resolution crystal structure of this complex.

When CAP [or cAMP (adenosine 3′,5′-monophosphate) receptor protein (CRP)] is complexed with its allosteric effector cAMP, it activates transcription at more than 20 different promoters in *Escherichia coli* [reviewed in (*16, 17*)]. Activation occurs when CAP-cAMP interacts with a specific DNA sequence located at positions that vary from −41 to −103 relative to the transcription start site in various operons.

The crystal structure of CAP-cAMP initially solved at 2.9 Å resolution (*18*) and then refined at 2.6 Å resolution (*19*) showed that the chemically identical 209–amino acid subunits of the CAP dimer consist of a larger amino-terminal domain that binds cAMP and a smaller carboxyl-terminal domain that binds DNA. The CAP-cAMP dimer is structurally asymmetric in these crystals; one subunit has a large cleft between the two domains (the "open" subunit), whereas the other does not (the "closed" subunit). B-DNA was positioned across the helix-turn-helix motifs of CAP according to electrostatic complementarity to generate a model for the CAP-DNA complex (*20, 21*). Subsequent mutagenic experiments are consistent, in part, with specific interactions proposed by this model (*22–27*).

That CAP induces a sharp bend in its DNA binding site has been demonstrated by analyses of the mobility of CAP-DNA complexes in polyacrylamide gels (*4, 5*), electrodicroism measurements of the rotational relaxation times of CAP-DNA complexes (*3*), enhanced rates of cyclization of DNA fragments (*6, 7*), and model building (*21, 28, 29*). Warwicker *et al.* (*28*) constructed CAP-DNA models that bent the DNA by 100° to 160° in order to place the sugar-phosphate backbone in contact with large regions of positive electrostatic potential that exist on the "sides" of CAP. Whereas CAP could only interact with ~20 base pairs of straight DNA, bending the DNA allows for interactions with a 28-bp segment that Liu-Johnson *et al.* (*29*) demonstrated is required for full affinity.

The 3 Å resolution crystal structure of CAP complexed with a 30-bp DNA sequence shows an overall bend of ~90° in the DNA that results primarily from two 40° kinks, one on each side of the dyad axis of the complex. The kinks, which occur in the conserved TGTGA sequence, as well as smaller distortions in other conserved regions of the CAP binding sequence, derive from interactions between the protein and the DNA phosphates and provide, in part, for specific binding through sequence-dependent distortability of the DNA. In addition, sequence specificity is achieved through direct hydrogen-bonding interactions between three side chains emanating from the "recognition" helix of CAP and the exposed edges of three base pairs in the major groove of the DNA helix.

We believe that the bend is an integral part of the mechanism for activation of transcription and propose that in addition to properly orienting CAP for possible interaction with RNA polymerase,

The authors are at Yale University, New Haven, CT 06511. S. C. Schultz and G. C. Shields are in the Department of Molecular Biophysics and Biochemistry and the Howard Hughes Medical Institute and T. A. Steitz is in the Departments of Molecular Biophysics and Biochemistry and of Chemistry and the Howard Hughes Medical Institute.

*Present address: Department of Chemistry, Lake Forest College, Lake Forest, IL 60045.

wrapping of the DNA around CAP may result in upstream DNA contacts with RNA polymerase.

Structure determination. Crystals of CAP complexed with a 30-bp DNA duplex with protruding 5′-dG termini [31-2 sequence in (*30*)] were grown in 5 percent (w/v) polyethylene glycol, 0.2 M NaCl, 0.1 M $CaCl_2$, and 2 mM cAMP, and then stabilized as described previously (*30*). The crystals are space group $C222_1$, with unit cell dimensions of a = 138.0, b = 152.6, and c = 76.0 Å, and contain one CAP-DNA dimer per asymmetric unit. The DNA consists of a duplex half-site containing a four-nucleotide self-complementary 3′-overhang that facilitates dimerization of two half sites to form a completely symmetric 30-bp DNA segment.

The structure of the complex was solved by molecular replacement methods with the use of x-ray intensity data between 20 and 3 Å resolution (*31*) and a newly refined structure for CAP-cAMP (*32*). The entire CAP-cAMP molecule was used as the search model for calculating rotation and translation search functions with the MERLOT computer program package (*33*). The resulting position for CAP-cAMP in the co-crystal lattice showed no overlap between adjacent protein molecules. Likewise, the protein did not overlap with the positions for four Br atoms that had been located by difference Patterson methods with the use of data collected from crystals grown with 5-bromodeoxyuridine (5-BrdU) (*34*) in place of Thy at positions 10 and 13 in each half-site of the DNA. Moreover, the Br positions near the closed subunit of CAP were within 3 and 8 Å of those predicted by the model of Warwicker *et al.* (*28*) for the 5-methyl groups of thymine 10 (T10) and T13, respectively; the DNA in this model could be easily adjusted to precisely fit these positions.

DNA associated with the open subunit in this model, however, could not be fitted to the Br positions without severe overlap with the protein. Rather, these Br positions were more consistent with a closed subunit conformation. Indeed, when a difference electron density map was calculated with diffraction amplitudes from brominated and unbrominated complexes ($F_{Br} - F_{Nati}$) and phases were calculated from the atomic positions of a symmetrized protein model that contained two closed domains, the four largest peaks at 16, 14, 10, and 8 standard deviations (SD) were located at the expected Br positions. In contrast, a difference electron density map calculated with phases obtained from the atomic positions of the CAP-cAMP structure contained only small peaks (<4 SD) corresponding to the Br atoms. Therefore, in subsequent calculations a model was used for the CAP-cAMP dimer that contained two closed subunits.

Since CAP contributes only ~60 percent of the x-ray scattering from this complex, electron density for the DNA was weak in maps phased with the protein coordinates alone. In order to improve the phases, solvent electron density was flattened outside of an envelope surface calculated (*35*) to be 3.5 Å from the protein and 5 Å from DNA derived from the model of Warwicker *et al.* (*28*) that was symmetrized from the closed subunit and adjusted to fit the Br positions previously identified. Solvent electron density outside the envelope was set to 60 percent of the initial average density inside the mask and density inside the mask was modified by truncating the lowest and highest 10 percent of the map. This density-modified map was Fourier transformed with the use of the Chicago fast Fourier transform (*36*). The resulting calculated structure factor amplitudes were modified according to the procedure of Read (*37*) and used to calculate a new electron density map. This procedure was repeated five times since additional cycles resulted in an average phase change of <1°. Difference Fourier maps calculated with these phases contained peaks at 16, 14, 11, and 10 SD that corresponded to the Br atom positions.

The solvent-flattened electron density map was sufficiently improved that 20 nucleotides (nt) (ten in each DNA half-site) could be fit into the density in a B-DNA conformation. Rigid-body refinement (*38*) of these 20 nt as individual base-sugar-phosphate units together with the four domains of CAP resulted in an R factor of 0.42. The resulting atomic positions were used to calculate (*38*) a $(2F_o - F_c)\phi_{calc}$ electron density map (where F_o and F_c are the observed and calculated structure factor amplitudes and ϕ_{calc} are the calculated phases) that was used to fit several additional nucleotides. Subsequent $2F_o - F_c$ maps were used to fit the remaining nucleotides. The R factor was 0.36 after rigid-body refinement of the individual base-sugar-phosphate units and the four protein domains.

Positional refinement was performed with the XPLOR software package (*38*) with data from 15 to 3 Å resolution, excluding ~1,000 of the smallest reflections, which provided 14,100 reflections for refining the positions of 4,411 non-hydrogen atoms. Stereochemical constraints were set such that the root-mean-square (rms) deviations from standard bond lengths was <0.016 Å and from standard bond angles was <3.6°. Initially, only one overall B factor was applied. As the refinement proceeded, group B factors were defined for each base-sugar-phosphate unit, each protein large domain, each protein small domain, and each of the two cAMP molecules. The R factor was 0.281 with the use of all data from 12 to 3 Å with the group B factors described above. Refinement of individual temperature factors yields an R factor of 0.235 with the use of only reflections from 8.0 to 3.0 Å whose $F > 2$ SD. The rms deviation from standard bond lengths is currently 0.014 Å and from standard bond angles is 3.4°.

An "annealed" electron density omit map (Fig. 1) (*39*) was calculated by omitting a region of the structure surrounding the kink that includes all of the atoms shown in Fig. 1 and refining the remaining atomic positions with the simulated annealing routine in XPLOR (*38*). This procedure produces an electron density map that is not biased by the model in the region of the structure that was omitted and therefore verifies that these atoms are correctly positioned. The correctness of the overall DNA position is independently verified by an exact correlation between Br atom positions obtained from difference Patterson and Fourier methods and posi-

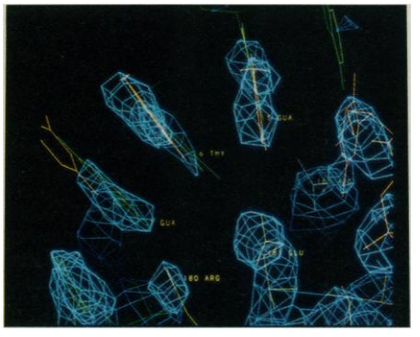

Fig. 1. "Annealed" electron density omit map (*39*) at the DNA kink. This electron density map was generated as follows: (i) all atoms of base pairs 4, 5, 6, and 7 as well as those within 8 Å of these nucleotides were omitted from all calculations, which includes all atoms shown in the figure; (ii) the positions of atoms within 3 Å of this omitted region were held constant; (iii) the remaining atomic positions were refined with the simulated annealing routine in XPLOR (*38*) (initial temperature = 3000°C); and (iv) the resulting atomic positions were used to calculate a $(2F_o - F_c)\phi_{calc}$ electron density map, where F_o and F_c are the observed and calculated structure factor amplitudes and ϕ_{calc} are the calculated phases. This procedure produces an electron density map that is not biased by the model in the region of the structure that was omitted. The contour level of this electron density map is 1.5 SD. Electron density corresponding to base pairs 5 and 6 clearly define their base planes and verify that these base pairs are ~40° from being parallel to each other. Density exists for C7 and A6 at lower contour levels (1.0 SD) than shown here. The side chains of Arg^{180}, Glu^{181}, and Arg^{185} are clearly present in the major groove, although at 3 Å resolution the positions of side chain atoms are not unambiguously defined. All non-hydrogen atoms are shown. Individual strands of the DNA are colored orange or green. This figure was generated with FRODO (*65*).

tions of the 5-methyl groups of T10 and T13 on each side of the molecule.

General features of the CAP-DNA complex. The CAP dimer with two bound cAMP molecules interacts directly with 27 of 30 base pairs of duplex DNA (Figs. 2 and 3). These interactions involve the protein small domain exclusively except for one residue (Lys^{26}) from the large domain. The amino end of the second helix of the helix-turn-helix penetrates the major groove of the DNA with its helix axis parallel to the base planes rather than the groove (Fig. 2). Amino acid side chains emanating from this helix interact directly with bases in the major groove of the DNA helix (Fig. 3).

The CAP protein in this complex is a symmetric dimer of closed subunits rather than an asymmetric dimer with one open and one closed subunit as in the CAP-cAMP structure (*18*) (Fig. 4). Whether such a large change in the position of this domain has functional significance or simply results from crystal packing influences in the CAP · cAMP and CAP · DNA crystals is unknown. Not surprisingly, previously proposed models for the structure of the specific complex (*20, 21*) and for DNA bending (*28*) constructed with the structure of the asymmetric CAP-cAMP dimer were more correct for the closed subunit.

DNA bending. The overall DNA bend is ~90° in the CAP-DNA complex (Fig. 2B), which agrees well with the value of ~100° proposed by Zinkel and Crothers (*40*) and is smaller than the value of 140° proposed by Thompson and Landy (*8*). The bend results almost entirely from 40° kinks between base pairs 5 and 6 on each side of the dyad axis (Fig. 3, B and C) that are formed by roll angles (rotation around the long axis of the base pairs) of ~40°. This roll angle is sufficiently large that base pairs 5 and 6 are unstacked. Discussions on the nature of DNA bending in nucleosomes distinguish kinks, in which the bases are unstacked to provide sharp localized bends, from smooth curving, in which the base pairs are only partially unstacked so as to spread the energy of bending over several base pairs (*1, 41*). Accordingly, we are distinguishing these distortions as true DNA kinks.

In addition to these kinks, the left side of the complex (Fig. 2) is bent ~8° toward the protein at ~10 bp from the dyad axis of the DNA (Fig. 2B). This small bend allows Lys^{26} of the cAMP binding domain to interact with a DNA phosphate (Fig. 3A). The analogous region of DNA on the other side of the complex is essentially straight, and Lys^{26} of this subunit cannot interact with the DNA. As discussed below, crystal-packing forces may influence bending near the ends of the DNA and may result in loss of the Lys^{26} interaction on the right side of the complex that accompanies a lack of additional bending. Perhaps in solution the DNA would be more bent. Indeed, the extension assay of Liu-Johnson *et al.* (*29*) demonstrates that the phosphate that interacts with Lys^{26} (as observed on the left side of the complex) (Fig. 3A) is necessary for full affinity in solution.

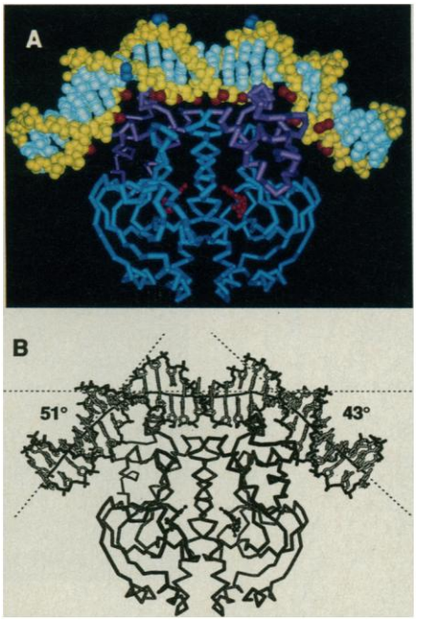

Fig. 2. (**A**) Structure of the CAP-DNA complex. The protein is represented as an α-carbon backbone trace with the cAMP binding domain in blue and the DNA binding domain in purple. The DNA is represented as a space-filling model with the bases in white and the sugar phosphate backbone in yellow. DNA phosphates whose ethylation interferes with DNA binding to CAP (*66*) are shown in red, and phosphates that are hypersensitive to deoxyribonuclease I (DNase I) are shown in blue. The cAMP molecules are shown as a ball-and-stick model in red. The phosphates whose ethylation interferes with binding to CAP all lie at the protein-DNA interface as required for direct interaction. The phosphates that are hypersensitive to DNase I bridge the kinks where the minor groove is very wide with an accessible groove width (*67*) of ~10 Å. A dramatic change occurs in minor groove width from ~10 Å at the kinks to ~3 Å 1.0 helical turn from the dyad axis of the molecule where the minor groove faces the protein. This figure was generated with Maximage (*68*) and Szazam (*69*). (**B**) Angle of DNA bending in the CAP-DNA complex. The DNA helix axis as defined by the program "Curves" (*70*) is shown as a black line running down the middle of the DNA helix. The complex is positioned with the recognition helix perpendicular to the page and lines are extended across the axis of the central ten base pairs and from the axis of the five terminal base pairs. The measured angles are as shown. The DNA on the right side of the complex (beyond the kink) is essentially straight, whereas the DNA on the left side is bent ~8° toward the protein near base pairs 10 and 11. In addition, the bend is out of plane such that the axis on the right side of the complex protrudes in front of the page and the axis on the left side extends behind the page. The angle of out of plane bending is ~35°. This figure was generated with Maximage (*68*).

Source of the DNA bend. What contributes to the energy of DNA binding and bending in the CAP-DNA complex? Binding of CAP to its specific DNA site involves (i) hydrogen-bonding interactions in the major groove of the DNA helix between three protein side chains and three base pairs (per half-site) and (ii) hydrogen-bonding and ionic interactions between 13 protein functional groups and 11 DNA phosphates (per half site) (Fig. 3).

The central 10-bp region is anchored to the protein by eight hydrogen bonds to the six DNA phosphates that face the protein at the dyad axis (Figs. 2 and 3). The protein hydrogen-bonding groups include the backbone amide of residue 139, which occurs at the amino terminus of the D-helix, and the hydroxyl groups of Thr^{140}, Ser^{179}, and Thr^{182} on both sides of the dyad axis. The central ten base pairs bend slightly toward the protein, apparently to allow for these phosphate interactions. In addition to these phosphate interactions, Glu^{181} and Arg^{185} interact directly with bases in the major groove of the DNA helix.

Interactions between CAP and the ten base pairs flanking the kink (distal to the dyad axis) require that the DNA be bent. Warwicker *et al.* (*28*) proposed that favorable electrostatic interactions between the DNA phosphates and regions of positive electrostatic potential on the "sides" of CAP could easily compensate for the energy required to bend DNA. The difference in electrostatic energy of interaction between a model-built complex similar to the one observed here and one with straight DNA interacting with CAP was calculated to be approximately −17 to −20 kcal/mole (*28*). As shown in Fig. 3A, the ten base pairs flanking the kink interact extensively with the protein to provide five ionic interactions and four hydrogen bonds to DNA phosphates as well as hydrogen bonding between Arg^{180} and G7. These interactions could not occur with straight DNA and therefore can contribute to binding only if the DNA is bent. Thus, the energy required to bend the DNA in this complex is largely supplied by extensive hydrogen-bonding and ionic interactions with the DNA phosphates of the two flanking 10-bp segments that can only occur when they are kinked ~40° relative to the central 10-bp segment.

Crystal packing influence on DNA bending. Of the nine high-resolution crystal structures of proteins complexed with duplex DNA, eight contain individual DNA segments stacked end-to-end

to form an extended DNA helix (*11–15, 42, 43*). In the CAP-DNA crystals, end stacking is also observed despite a 90° bend in the DNA (Fig. 5A). The DNA in adjacent asymmetric units does not stack to form a continuous helix, however. Instead the 3′-dG from the terminal GC base pair stacks against the same base of the adjacent molecule. The protruding 5′-dG nucleotide is positioned to form a triple-base structure with the terminal GC base pair of the adjacent molecule (Fig. 5B).

The importance of crystal packing and its effect on the observed conformations of DNA in crystals has been clearly demonstrated (*44*). DNA structure in protein-DNA crystals may be less influenced by crystal packing effects than it is in crystals of DNA alone because of the large energy of interaction with the protein. Nevertheless, favorable stacking could select for a conformation that might be somewhat less favored in solution, particularly for the ends of the DNA. Crystal packing considerations are unlikely to affect the two kinks observed in this structure since the interactions between CAP and DNA in this region are extensive. However, the DNA termini are not as tightly tethered to the protein and, indeed, differ slightly in their structures, presumably because of crystal packing forces.

The influence of crystal packing and possible additional protein contacts on DNA conformation can only be assessed by determining structures of complexes that contain different DNA lengths and packing arrangements. The structure of the 434 repressor fragment

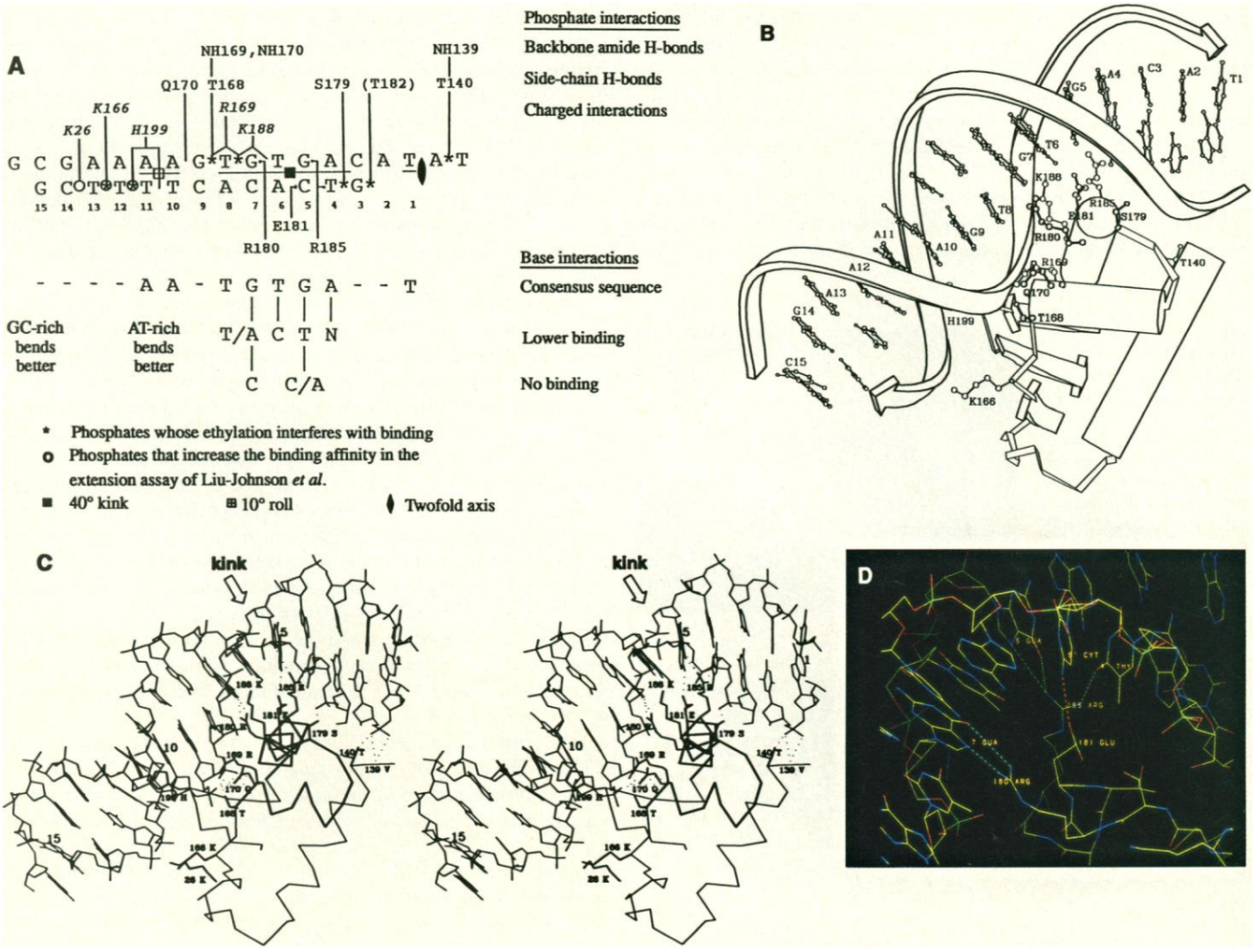

Fig. 3. (**A**) Nucleotide sequence of the half-site DNA fragments contained in the CAP-DNA complex described here and assignments for interactions with the protein. Criterion for assigning hydrogen-bonding and ionic interaction are given in (*71*). Note that at 3 Å resolution, these assignments should be considered as probable rather than certain since the average error in the positions of side chain atoms is ~0.5 Å. Numbering of the base pairs is from the dyad axis as described in Steitz (*47*). The missing phosphate that results from the use of half-site sequences occurs between base pairs 2 and 3. One of the interactions (Thr^{182} with the phosphate of T2) is not observed directly since this phosphate is absent due to the half-site sequences used to grow the crystals (*30*). However, a phosphate can easily be placed into the gap without altering the DNA backbone conformation; this phosphate is within hydrogen-bonding distance of Thr^{182}. Also noted are phosphates whose ethylation interferes with binding (*66*); phosphates that increase binding affinity in the extension assay of Liu-Johnson *et al.* (*29*); the consensus CAP binding site (*16*); and base pair changes that alter binding affinity (*23–25*). (**B**) Schematic drawing of potential interactions between one DNA half-site and one small domain of the CAP dimer. Protein helices are shown as tubes and β strands as arrows. Side chains of the small domain that are close enough to interact with the DNA are included. The DNA sugar-phosphate backbone is represented as an arrow pointing 5′ to 3′. This figure was generated with Arplot (*72*). (**C**) Stereo view of one small domain of CAP (shown as an α-carbon backbone) bound to one half of the CAP binding site. The helix-turn-helix motif is highlighted in bold. All protein side chains close enough to interact with the DNA (*71*) are shown, including Lys^{26} in the large domain. Potential hydrogen bonds between the protein and DNA are indicated as dotted lines. This figure was generated with FRODO (*65*) and Plot (*73*). (**D**) Apparent hydrogen-bonding interactions between protein side chains and nucleotide functional groups in the major groove of DNA (*71*). The atoms are colored according to type: O, red; N, blue; and C, green. Arg^{180} interacts with the O-6 and N-7 of G7, Glu^{181} interacts with the N-4 of C5, and Arg^{185} interacts with the O-6 or N-7 of G5 and the O-4 of T6. This figure was generated with FRODO (*65*).

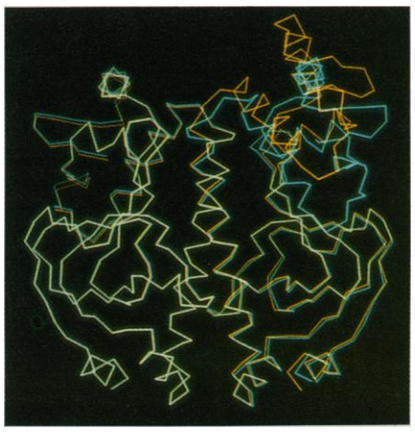

Fig. 4. Comparison of α-carbon backbone traces of CAP-cAMP (orange) and CAP-cAMP-DNA (blue). Whereas the crystal structure of the CAP-cAMP dimer contains an open and a closed subunit, the CAP-DNA complex shows a symmetric dimer of closed subunits. The rms difference in α-carbon positions of the two protein subunits in CAP-DNA are 0.61 Å and 0.71 Å with respect to the closed subunit of CAP-cAMP. This figure was generated with FRODO (*65*).

(residues 1 to 69) complexed with a 20-nt DNA exhibits slightly greater DNA bending (*12*) than with a 14-nt fragment (*42*). Additional crystal forms of CAP-DNA complexes have been obtained with different DNA sequences and lengths (*30, 45*) whose structures could be used to show the effect of crystal packing on the DNA conformation.

Sequence-specific DNA binding. Two important sources of DNA sequence specificity in protein-DNA complexes are: (i) direct hydrogen-bonding and van der Waals interactions between protein side chains and the exposed edges of base pairs in the major groove of B-DNA (*11–13, 46, 47*); and (ii) sequence-dependent bendability or deformability of duplex DNA (*14, 28, 47–49*). The former results from structural complementarity that allows better interactions with correct than with incorrect sequences, while the latter results from the ability of some nucleic acid sequences to adopt a particular structure required for binding to a protein at lower free energy cost than other sequences (*47*).

Specific base contacts. Three side chains emanating from the "recognition helix" of the helix-turn-helix motif in CAP appear to hydrogen bond directly to three base pairs in the major groove of the DNA (Fig. 3). The guanidinium group of Arg^{180} hydrogen bonds to the O-6 and N-7 of G7, as proposed in earlier models (*20, 21, 23*), which is consistent with the altered DNA affinity exhibited by CAP mutated at residue 180 (*22, 27*). The carboxylate of Glu^{181} interacts with the N-4 of C5, again as expected from model building (*20, 21*) and mutagenic studies (*23, 24*). Reduction of CAP affinity due to a G to A base change at position 5 is suppressed by Glu^{181} to Val or Leu mutations in CAP (*23*). This suppression results not because the Glu^{181} interaction is lost and replaced with another favorable interaction of Val or Leu with thymine at position 5, but because the interaction between Glu^{181} and other nucleotides at position 5 is unfavorable such that nonspecific DNA binding affinity is enhanced for the mutant protein (*24*). This result demonstrates the importance of negative as well as positive complementarity for sequence discrimination.

Although Arg^{185} interacts with either the O-6 or the N-7 of G5 and the O-4 of T6, its role in sequence discrimination is less clear since the mutation Arg^{185} to Leu has little effect on specific DNA binding in vitro or activation of transcription in vivo (*22*). The Arg side chain may be flexible enough to accommodate a number of base sequences in the major groove and therefore may not be effective in sequence discrimination. Nevertheless, these results do not eliminate the possibility of a role for Arg^{185} in sequence discrimination since the affinity of the Arg^{185} to Leu mutant protein for DNA mutated at base pairs 5 and 6 has not been determined.

DNA bendability and specificity. The kink between base pairs 5 and 6 in the CAP-DNA complex occurs at a TG step in the sequence GTG. These TG steps are highly conserved on both sides of the dyad axis in different CAP binding sites (*16*). Lu *et al.* (*50*) have observed that the rate of imino proton exchange in duplex DNA segments shows a local maximum at GTG sequences. This result might be explained by more frequent transient kinking of the DNA at these sequences such that the imino proton would be more exposed to solvent. Conservation of the TG sequence at positions 5 and 6 in CAP binding sites might be due in part to the ease with which it can be kinked, as is required for CAP binding.

Do the binding constants of CAP to altered DNA binding sites support the proposal that TG steps are more easily kinked and therefore contribute to specific binding? When the T:A base pair at position 6, which is not contacted directly by CAP, is changed to a C:G base pair, the dissociation constant increases ~6.7 fold (*24*). Likewise, altering the G:C base pair at position 5 dramatically reduces binding (*5, 23–25, 51*), which presumably results primarily from interaction of this base pair with Glu^{181} (*23, 24*). However, mutation of Glu^{181} to Val or Leu removes this interaction. These mutant proteins discriminate very poorly between DNA binding sites containing TG, TC, TA, and TT base steps at position 5 and 6 (*23, 24*). This result suggests that either the differences in stacking energies of these base sequences are smaller than expected from various estimates (*52*) and therefore such forces are not effective in sequence discrimination, or that base pairs 5 and 6 are not unstacked upon binding to the mutant protein. Perhaps Glu^{181} positions base pair 5 in a way that requires a kink, and a different mode of bending occurs when this interaction is removed. Indeed, the 6.7-fold reduction in wild-type CAP binding to DNA molecules that contain a C:G rather than a T:A base pair at position 6 becomes a 1.6-fold reduction in binding to the mutant proteins (*24*).

The smaller bend that occurs about ten base pairs from the dyad axis on one side of the complex also contributes to specific DNA binding through sequence-dependent distortion of DNA. Gartenberg and Crothers (*49*) found that CAP binding sites containing AT

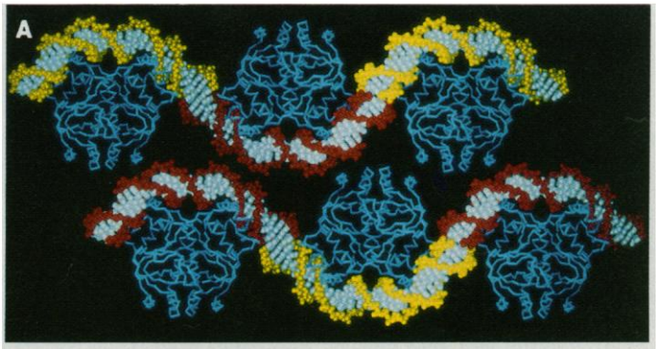

Fig. 5. End-to-end stacking of DNA in crystals of the CAP-DNA complex. (**A**) DNA molecules from six adjacent asymmetric units are colored alternately yellow and red. The protein is represented as an α-carbon backbone trace in blue. Notice that the DNA stacks 3′ to 3′ such that it does not form a continuous DNA helix. This figure was generated with Szazam (*69*). (**B**) Triple base pairing structure in which the 5′-dG terminus (G16) of one asymmetric unit hydrogen bonds to the terminal GC base pair (C15 and G15) of the adjacent asymmetric unit at each of the three functional groups in the major groove of the DNA helix. This figure was generated with Loliplot (*74*).

bases at nucleotides 10 and 11 bend more than those containing GC bases when bound to CAP (as assessed by polyacrylamide gel electrophoresis) and also bind 14-fold more tightly, thereby correlating sequence-dependent distortion with binding affinity. The structural distortions in the DNA that facilitate binding to the protein are a narrower minor groove and an 8° bend near base pair 10 that allows Lys^{26} to interact with a DNA phosphate (on one side in the crystal). The narrow minor groove allows better interaction between the sugar-phosphate backbone and Lys^{166} and His^{199}. The reduced minor groove width could result in part from the AT base pairs present in this region (*53*). Furthermore, AT base pairs at positions 10 and 11 allow DNA to be more easily bent in the direction of the minor groove, which faces the protein at this position. The nucleosome binds to DNA with a preference for AT base pairs when the minor groove faces the protein and GC base pairs when the major groove faces the protein (*54*). Accordingly, Drew and Travers (*54*) have proposed that adjacent AT base pairs favor bending into the minor groove and GC base pairs favor bending into the major groove. The 7 Å resolution crystal structure of the nucleosome core particle (*10*) shows that the major and minor grooves are compressed when facing the protein and bend toward the protein in these regions.

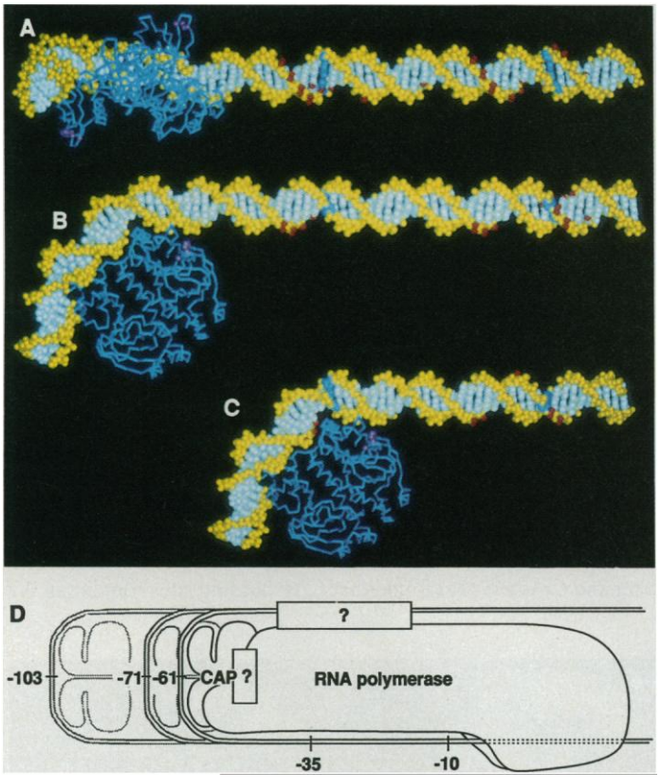

Fig. 6. (**A**) Position of the CAP-DNA complex relative to the RNA polymerase binding site in the *lac* operon. The DNA is extended on one side of the CAP-DNA structure such that CAP is positioned at −61.5. The rightmost base is −1 and the −10 and −35 base pairs are shown in blue. Phosphates whose ethylation interferes with RNA polymerase binding in the *lac* UV5 promoter (*75*) are shown in red, as are base functional groups protected by RNA polymerase binding (*75*). The side chains in lavender are those that when mutated reduce transcriptional activation by CAP but do not affect DNA binding (*57*). Note that these side chains sit on the surface of CAP such that they could easily be contacted by RNA polymerase. (**B**) Same as (A) except rotated 90° around the long axis of the DNA. Note that the DNA functional groups that interact with RNA polymerase (*75*) are on the same face of the DNA helix as CAP. (**C**) Arrangement of CAP relative to the RNA polymerase binding site in the *gal* operon. Of particular importance is that CAP interacts with the major groove at −35 such that the CAP and polymerase binding sites overlap. Thus, it is not possible that polymerase binds to the *gal* promoter in the same way as the *lac* promoter. Either CAP binds elsewhere or RNA polymerase does not interact at the −35 site. (A), (B), and (C) were generated with Szazam (*69*). (**D**) Schematic of DNA wrapping model for activation of transcription by CAP. Protein-protein interactions between CAP and RNA polymerase that may occur when CAP is bound at −61 could not occur when CAP is bound at −71 or −103. However, interactions between RNA polymerase and DNA upstream from the CAP binding site would be unaffected by the position of CAP binding, assuming that the DNA bend is in plane or that nonplanarity could compensate for the relative "phase" of the CAP and RNA polymerase binding sites.

Gartenberg and Crothers also found that GC-rich sequences around nucleotide 16 also favor bending in the CAP · DNA complex (*49*). The structure described here does not address the possibility of an additional bend near base pair 16 since our crystals contain DNA that extends only 15 base pairs from the dyad axis. However, we note that an additional bend near base pair 16 would allow phosphates of a longer DNA segment to contact regions of positive potential further down on the large domain of CAP and produce a bend larger than 100°.

Transcription activation by CAP. DNA bending by CAP could contribute to activation of transcription in two ways: (i) by properly orienting CAP and RNA polymerase for direct (protein-protein) interactions or (ii) by wrapping the DNA around CAP to provide for contacts between RNA polymerase and DNA upstream of the CAP binding site.

The severe DNA bend results in a very different relative orientation of CAP and RNA polymerase than would occur with straight DNA such that a rather unexpected surface of CAP is adjacent to the −35 region of the RNA polymerase binding site. Indeed, because of the bend, we see that the helix-turn-helix of CAP is essentially inaccessible to RNA polymerase (Fig. 6).

Two lines of evidence have suggested that direct contact between CAP and RNA polymerase might be important for transcriptional activation. First, CAP-cAMP and RNA polymerase holoenzyme (including σ-70) interact in solution with a dissociation constant of ~1 μM (*55*). Second, four examples of single-site mutations in CAP are reported to affect activation of transcription without altering DNA binding (*56, 57*). Two of the residues implicated (Glu^{171} and Glu^{72}) (*56, 57*) are not exposed to the surface and therefore cannot contact RNA polymerase directly. However, the other two residues (His^{159} and Gly^{162}) (*57*) are positioned on the outer surface of the small domain of CAP more than 14 Å from the DNA. These residues are very accessible and could be contacted by RNA polymerase (Fig. 6).

However, a simple model of transcription activation that postulates only a common interaction between these two proteins is difficult to reconcile with the ability of CAP to activate transcription when bound to positions that vary from −41 to −103 nt from the transcription start in different operons (*16, 17*) and when the binding sites are moved by intervals of ten base pairs in the *lac* (*58*) and *gal* (*59*) operons. A model for activation of transcription that invokes DNA bending might explain these data more readily.

DNA wrapping model for activation of transcription. CAP induces a hairpin bend in the DNA that might facilitate contacts between RNA polymerase and DNA upstream from the CAP binding site (Fig. 6D). Such a model could explain how CAP can activate transcription from such a variety of positions since the size of the loop could vary. Busby and Buc (*60*) have shown that for the *gal* operon DNA upstream of the CAP binding site is protected from deoxyribonuclease I digestion in the presence of RNA polymerase and have suggested that this might be important for activation. However, as shown in Fig. 6C, the *gal* operon is unusual in that the binding sites for CAP and RNA polymerase actually overlap in the −35 region such that these proteins cannot simultaneously occupy

their binding sites in the same manner as for the *lac* operon. The 90° DNA bend induced by CAP would not be sufficient to provide upstream contacts as shown in Fig. 6D, but as discussed previously, additional bending might occur 10 and 16 base pairs from the dyad axis. A larger DNA bend may be transient with CAP alone but could be stabilized by upstream DNA interactions with RNA polymerase.

The observed synergistic interaction (*58, 61*) in the binding of CAP and RNA polymerase to promoter DNA could be produced by upstream contacts between polymerase and DNA that are facilitated by CAP binding as well as by direct contact between the two proteins.

Two lines of evidence suggest that DNA bending alone can activate transcription. First, CAP binding sites can be replaced by A-tract sequences that are expected to intrinsically bend the DNA such that activation of transcription occurs in a phase-dependent manner both in vitro (*62*) and in vivo (*63*); A-tract activation in vitro requires supercoiled template DNA (*62*). Second, replacing the λ repressor sites with a site for the integration host factor, which is known to produce a severe DNA bend, results in IHF-dependent activation of λ-P_L (*64*). DNA wrapping provides a simple model that might explain in part how CAP-induced DNA bending could activate transcription at a variety of distances from the RNA polymerase binding site.

REFERENCES AND NOTES

1. J. D. McGhee and G. Felsenfeld, *Annu. Rev. Biochem.* **49**, 1115 (1980).
2. M. Better *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **79**, 5837 (1982).
3. D. Porschke, W. Hillen, M. Takahashi, *EMBO J.* **3**, 2873 (1984).
4. H.-M. Wu and D. M. Crothers, *Nature* **308**, 509 (1984).
5. A. Kolb *et al.*, *Nucleic Acids Res.* **11**, 7833 (1983).
6. D. Kotlarz, A. Fritsch, H. Buc, *EMBO J.* **5**, 799 (1986).
7. D. Dripps and R. M. Wartell, *J. Biomol. Struct. Dyn.* **5**, 1 (1987).
8. J. F. Thompson and A. Landy, *Nucleic Acids Res.* **16**, 9687 (1988).
9. L. D. Kosturko, E. Daub, H. Murialdo, *ibid.* 17, 317 (1989).
10. T. J. Richmond *et al.*, *Nature* **311**, 532 (1984).
11. C. A. Frederick *et al.*, *ibid.* **309**, 327 (1984).
12. A. K. Aggarwal *et al.*, *Science* **242**, 899 (1988).
13. S. R. Jordan and C. O. Pabo, *ibid.*, p. 893.
14. Z. Otwinowski *et al.*, *Nature* **335**, 321 (1988).
15. R. G. Brennan, S. L. Roderick, Y. Takeda, B. W. Matthews, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 8165 (1990).
16. B. de Crombrugghe, S. Busby, H. Buc, *Science* **224**, 831 (1984).
17. A. Ullman and A. Danchin, *Adv. Cyclic Nucleotide Res.* **15**, 1 (1983).
18. D. B. McKay and T. A. Steitz, *Nature* **290**, 744 (1981).
19. I. T. Weber and T. A. Steitz, *J. Mol. Biol.* **198**, 311 (1987).
20. T. A. Steitz *et al.*, *J. Biomol. Struct. Dyn.* **1**, 1023 (1983).
21. I. T. Weber and T. A. Steitz, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 3973 (1984).
22. M. E. Gent *et al.*, *Biochem. J.* **242**, 645 (1987).
23. R. H. Ebright *et al.*, *Nature* **311**, 232 (1984).
24. R. H. Ebright *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 6083 (1987).
25. C. Jansen, A. M. Gronenborn, G. M. Clore, *Biochem. J.* **246**, 227 (1987).
26. R. H. Ebright *et al.*, *Nucleic Acids Res.* **18**, 1457 (1990).
27. X. Zhang and R. Ebright, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 4717 (1990).
28. J. Warwicker, B. P. Engelman, T. A. Steitz, *Proteins* **2**, 283 (1987).
29. H.-N. Liu-Johnson, M. R. Gartenberg, D. M. Crothers, *Cell* **47**, 995 (1986).
30. S. C. Schultz, G. C. Shields, T. A. Steitz, *J. Mol. Biol.* **213**, 159 (1990).
31. X-ray intensity data were measured by using a Xuong-Hamlin multiwire area detector system with a Rigaku RU200 rotating anode generator as the x-ray source. The data were reduced to integrated intensities by using University of California, San Diego, software and processed with either the UCSD software or Scalepack (Z. Otwinowski, personal communication). The native data were derived from 88,548 observations of 15,377 reflections, which is 99 percent of the possible 15,529 unique reflections from 20 to 3 Å resolution. The average intensity divided by the average error in intensity was 2.1 for the data between 3.2 and 3.0 Å resolution. The merging *R* factor for independent observations of the reflection intensities from 20 to 3 Å was 4.8 percent.
32. G. C. Shields and T. A. Steitz, unpublished results.
33. P. M. D. Fitzgerald, *J. Appl. Crystallogr.* **21**, 273 (1988).
34. DNA molecules containing 5-bromodeoxyuridine at positions 10 and 13 were synthesized by using 5-BrdU β-cyanoethyl phosphoramidite (Pharmacia) on an Applied Biosystem (Foster City, CA) model 380 B DNA synthesizer and purified as described previously (*30*).
35. S. Katti, personal communication.
36. D. A. Bantz and M. Zwick, *Acta Crystallogr.* **A30**, 257 (1974).
37. R. J. Read, *ibid.* **A42**, 140 (1986).
38. A. T. Brünger, *J. Mol. Biol.* **203**, 803 (1988); Program XPLOR, Yale University (1989).
39. _____ and L. Beese, personal communication.
40. S. S. Zinkel and D. M. Crothers, *Biopolymers* **29**, 29 (1990).
41. J. L. Sussman and E. N. Trifonov, *Proc. Natl. Acad. Sci. U.S.A.* **75**, 103 (1978); F. H. C. Crick and A. Klug, *Nature* **255**, 530 (1975); H. M. Sobell, C.-C. Tsai, S. G. Gilbert, S. C. Jain, T. D. Sakore, *Proc. Natl. Acad. Sci. U.S.A.* **73**, 3068 (1976); V. B. Zhurkin *et al.*, *Nucleic Acids Res.* **6**, 1081 (1979).
42. J. E. Anderson, M. Ptashne, S. C. Harrison, *Nature* **326**, 846 (1987).
43. C. Wolberger, Y. Dong, M. Ptashne, S. C. Harrison, *ibid.* **335**, 789 (1988); C. R. Kissinger, B. Liu, E. Martin-Blanco, T. B. Kornberg, C. O. Pabo, *Cell* **63**, 579 (1990); N. P. Pavletich and C. O. Pabo, *Science* **252**, 809 (1991).
44. A. D. DiGabriele *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 1816 (1989); R. E. Dickerson *et al.*, *J. Biomol. Struct. Dyn.* **5**, 557 (1987).
45. J. Passner, S. Schultz, T. A. Steitz, unpublished results.
46. C. O. Pabo and M. Lewis, *Nature* **298**, 443 (1982); S. R. Jordan, T. V. Whitcombe, J. M. Berg, C. O. Pabo, *Science* **230**, 1383 (1985).
47. T. A. Steitz, *Q. Rev. Biophys.* **23**, 205 (1990).
48. J. Anderson *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 1307 (1984); G. B. Koudelka, S. C. Harrison, M. Ptashne, *Nature* **326**, 886 (1987); G. B. Koudelka *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 4633 (1988).
49. M. R. Gartenberg and D. M. Crothers, *Nature* **333**, 824 (1988).
50. P. Lu, S. Cheung, M. Donlan, *Adv. Biophys.* **20**, 153 (1985).
51. J. Beckwith, T. Grodzicker, R. Arditti, *J. Mol. Biol.* **69**, 155 (1972); R. C. Dickson, J. Abelson, P. Johnson, W. Reznikoff, W. Barnes, *ibid.* **111**, 65 (1977).
52. R. L. Ornstein, R. Rein, D. L. Breen, R. D. MacElroy, *Biopolymers* **17**, 2341 (1978); O. Gotoh and Y. Tagashira, *ibid.* **20**, 1033 (1981); K. J. Breslauer, R. Frank, H. Blöcker, L. A. Marky, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 3746 (1986).
53. C. Yoon *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 6332 (1988).
54. H. R. Drew and A. A. Travers, *J. Mol. Biol.* **186**, 773 (1985); S. C. Satchwell, H. R. Drew, A. A. Travers, *ibid.* **191**, 659 (1986).
55. B. Blazy, M. Takahashi, A. Baudras, *Mol. Biol. Rep.* **6**, 39 (1980); M. Pinkney and J. G. Hoggett, *Biochem. J.* **250**, 897 (1988).
56. N. Irwin and M. Ptashne, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 8315 (1987).
57. A. Bell *et al.*, *Nucleic Acids Res.* **18**, 7243 (1990); A. C. Eschenlauer and W. S. Reznikoff, *J. Bacteriol.*, in press.
58. D. C. Straney, S. B. Straney, D. M. Crothers, *J. Mol. Biol.* **206**, 41 (1989).
59. K. Gaston, A. Bell, A. Kolb, H. Buc, S. Busby, *Cell* **62**, 733 (1990).
60. S. Busby and H. Buc, *Microbiol. Sci.* **4**, 371 (1987).
61. T. P. Malan, A. Kolb, H. Buc, W. R. McClure, *J. Mol. Biol.* **180**, 881 (1984); A. Spassky, S. Busby, H. Buc, *EMBO J.* **3**, 43 (1984).
62. M. R. Gartenberg and D. M. Crothers, *J. Mol. Biol.* **219**, 217 (1991).
63. L. Bracco, D. Kotlarz, A. Kolb, S. Diekmann, H. Buc, *EMBO J.* **8**, 4289 (1989).
64. H. Giladi, M. Gottesman, A. B. Oppenheim, *J. Mol. Biol.* **213**, 109 (1990).
65. Program FRODO; T. A. Jones, in *Computational Crystallography* D. Sayre, Ed. (Clarendon, Oxford, 1982), pp. 303–317.
66. J. E. Majors, thesis, Harvard University, Cambridge, MA (1977).
67. Accessible groove width is defined as the minimum phosphorus-phosphorus distance across the minor groove minus 5.8 Å as in the program Newhelix [R. Dickerson, personal communication; A. V. Fratini, M. L. Kopka, H. R. Drew, R. E. Dickerson, *J. Biol. Chem.* **257**, 14686 (1982)].
68. Program Maximage, M. R. Rould, Yale University, New Haven, CT (1990).
69. Program Szazam (version 6.2), A. Perlo, Yale University.
70. R. Lavery and H. Sklenar, *J. Biomol. Struct. Dyn.* **6**, 655 (1989); G. Ravishanker, S. Swaminathan, D. L. Beveridge, R. Lavery, H. Sklenar, *ibid.*, p. 669.
71. Functional groups are considered to be hydrogen bonding if the appropriate N or O atoms are <3.5 Å apart. The ionic interactions listed are those whose functional groups are <4.5 Å apart. Abbreviations for the amino acid residues are: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.
72. Program Arplot, A. M. Lesk and K. D. Hardman, *Science* **216**, 539 (1982).
73. Program Plot, G. Furnish, Yale University (1989).
74. Program Loliplot, J. Bienkowska, modified by C. Foreman and P. Vogt, Yale University (1989).
75. U. Siebenlist, R. B. Simpson, W. Gilbert, *Cell* **20**, 269 (1980).

76. We thank J. Friedman, S. Katti, A. Goldman, and the members of the Steitz lab for their assistance throughout this work, D. M. Crothers for many helpful discussions, and A. C. Eschenlauer and W. S. Reznikoff for viewing their manuscript prior to publication. Supported in part by NIH grant GM-22778 to T.A.S. and in part by an American Cancer Society Postdoctoral Fellowship to S.C.S.

15 March 1991; accepted 22 July 1991