J. Mol. Biol. (1996) **260**, 395–408

JMB

Structure of the CAP-DNA Complex at 2.5 Å Resolution: A Complete Picture of the Protein-DNA Interface

Gary Parkinson, Christopher Wilson, Angelo Gunasekera Yon W. Ebright, Richard E. Ebright and Helen M. Berman*

Department of Chemistry and The Waksman Institute Rutgers University New Brunswick NJ 08855-0939, USA

**Corresponding author*

The crystallographic structure of the CAP-DNA complex at 3.0 Å resolution has been reported previously. For technical reasons, the reported structure had been determined using a gapped DNA molecule lacking two phosphates important for CAP-DNA interaction. In this work, we report the crystallographic structure of the CAP-DNA complex at 2.5 Å resolution using a DNA molecule having all phosphates important for CAP-DNA interaction. The present resolution permits unambiguous identification of amino acid-base and amino acid-phosphate hydrogen bonded contacts in the CAP-DNA complex. In addition, the present resolution permits accurate definition of the kinked DNA conformation in the CAP-DNA complex.

Keywords: catabolite activator protein (CAP); cAMP receptor protein (CRP); protein-DNA interaction; transcription; X-ray crystallography

Introduction

The *Escherichia coli* catabolite gene activator protein (CAP; also referred to as the cAMP receptor protein CRP) is a structurally, biochemically, and genetically characterized transcription activator protein (reviewed by Kolb *et al*., 1993). CAP functions by binding, in the presence of the allosteric effector cAMP, to specific DNA sites located at or near promoters. The consensus DNA site for CAP is 5'-AAATGTGATCTAGATCACATTT-3' (Berg & Von Hippel, 1988; Ebright *et al*., 1989; Gunasekera *et al*., 1992). The consensus DNA site is 22 base-pairs in length and exhibits perfect 2-fold sequence symmetry. Positions 4 to 8 and 15 to 19 of the consensus DNA site (underlined above) are the most important for CAP-DNA complex formation.†

The crystallographic structure of CAP has been determined to 2.5 Å resolution (Weber & Steitz, 1987). CAP is a dimer of two identical subunits, each of which is 209 amino acids in length and contains a helix-turn-helix DNA binding motif (reviews of the helix-turn-helix motif by Brennan, 1992; Pabo & Sauer, 1992). The crystallographic structure of the CAP-DNA complex has been determined to 3.0 Å resolution (Schultz *et al*., 1991). The CAP-DNA complex is approximately 2-fold symmetric: one subunit of CAP interacts with one half of the DNA site; the other subunit of CAP interacts in an approximately 2-fold-symmetry-related fashion with the other half of the DNA site. The majority of CAP-DNA interactions are mediated by the helix-turn-helix DNA binding motif present in each subunit of CAP. CAP sharply bends DNA in the CAP-DNA complex (Porschke *et al*., 1984; Wu & Crothers 1984; Liu-Johnson *et al*., 1986; Kotlarg *et al*., 1986; Schultz *et al*., 1991). The magnitude of the CAP-induced DNA bend is 90°; the orientation of the CAP-induced DNA bend is such that the DNA wraps towards and around the sides of CAP.

The DNA molecule used in the previous crystallographic study of the CAP-DNA complex consisted of two 15 base-pair half-site oligodeoxyribonucleotide duplexes annealed through 3'-overhanging ends (DNA molecule ''31-2''; Figure 1(b); Schultz *et al*., 1990, 1991). Using two half-site

Present addresses: C. Wilson, Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA; A. Gunasekera, Abbott Laboratories, Abbot Park, IL 60064, USA.

Abbreviation used: CAP, catabolite gene activator protein.

† In this paper, positions within the DNA site for CAP are numbered as by Ebright *et al*. (1984a,b, 1987, 1989), Zhang & Ebright (1990), Zhang *et al*. (1991) and Gunasekera *et al.* (1992) (Figure 1(a), convention 1). A different numbering convention is used by Schultz *et al*. (1991) (Figure 1(a), convention 2).

0022–2836/96/280395–14 $18.00/0

oligodeoxyribonucleotide duplexes rather than an intact, full-site oligodeoxyribonucleotide duplex avoided complications due to the hairpin formation by full-site oligodeoxyribonucleotides and facilitated the preparation of a series of different DNA molecules for crystallization trials (Schultz *et al*., 1990, 1991). However, the use of the particular half-site oligodeoxyribonucleotide duplexes of Schultz *et al*. resulted in a DNA molecule that lacked two phosphates known from ethylation-interference experiments (Majors, 1977; Shanblatt & Revzin, 1986) to make critical interactions with CAP; i.e. the phosphate 5' to the bottom-strand nucleotide at position 9', and the phosphate 5' to the top-strand nucleotide at position 14. Thus, use of the half-site oligodeoxyribonucleotide duplexes of Schultz *et al*. (1991) precluded obtaining information about known critical interactions. Consistent with the loss of known critical interactions, electrophoretic mobility shift DNA binding experiments indicate that the half-site oligodeoxyribonucleotide duplexes of Schultz *et al*. (1990) bind at least three orders of magnitude less well to CAP than does an intact, full-site oligodeoxyribonucleotide duplex (A.G., Y.W.E. & R.H.E., unpublished results).

The DNA seen in this complex has served as a model for the understanding of DNA bending and has been the subject of studies that have attempted to relate sequence and protein interactions to the DNA conformation (Zhurkin *et al*., 1991; Gorin *et al*., 1995, Ivanov *et al*., 1995). Studies of this nature depend on detailed analyses of the base morphology parameters and would benefit from a more accurate model of the DNA.

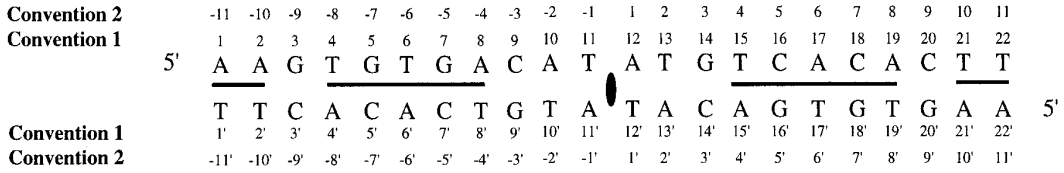
In this work, we report the crystallographic structure of the CAP-DNA complex at 2.5 Å resolution using a DNA molecule containing all phosphates important for CAP-DNA interaction (DNA molecule ''31-2E''; two 15 base-pair half-site oligodeoxyribonucleotide duplexes annealed through 5'-overhanging ends, Figure 1(c)). The higher resolution, as well as the use of an updated DNA parameter file (Parkinson *et al*., 1996), provide an improved model for the complex and each of its components.

Results

Overall structure

Figure 2(a) shows the structure of the CAP-DNA complex at 2.5 Å resolution. The CAP-DNA complex exhibits approximate 2-fold symmetry and consists of the CAP dimer (residues 9 to 207 for subunit A; residues 9 to 204 for subunit B), two

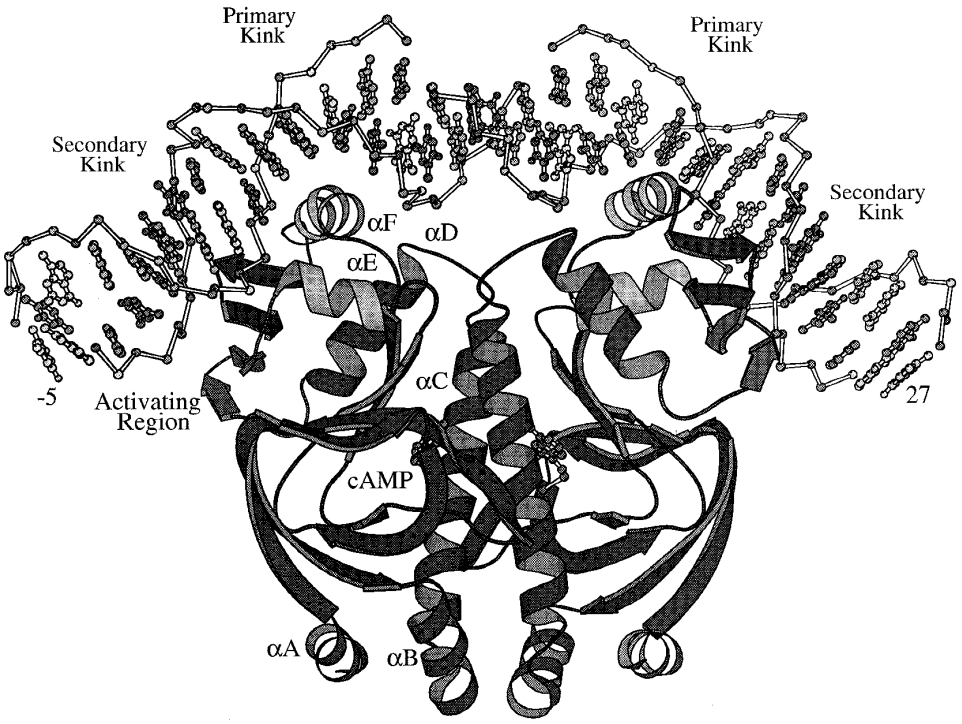
a) DNA Site for CAP



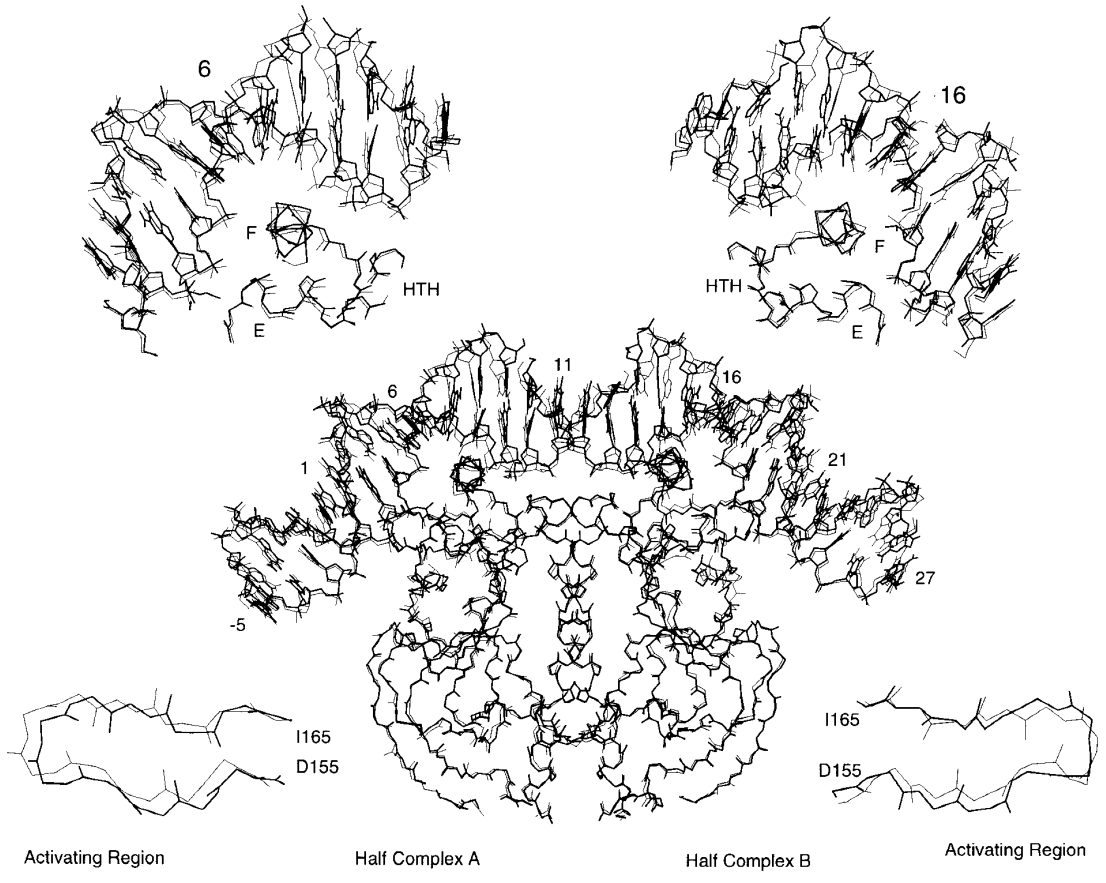
b) CAP-DNA Complex: Schultz *et al.* (1991)

c) CAP-DNA Complex: Present Work

Figure 1. (a) DNA site for CAP. There are two conventions for numbering of positions within the DNA site for CAP: convention 1 (Ebright *et al*., 1984a,b, 1987, 1989; Zhang & Ebright, 1990; Zhang *et al*., 1991; Gunasekera *et al*., 1992) and convention 2 (Schultz *et al*., 1991). In this paper we use convention 1. (b) DNA molecule ''31-2'' used in the previously reported structure of the CAP-DNA complex (Schultz *et al*., 1990, 1991). DNA molecule ''31-2'' contains symmetry-related single-phosphate gaps between positions 9' and 10' and 13 and 14 (filled bars; see the text). (c) DNA molecule ''31-2E'' used in the present structure of CAP-DNA complex. DNA molecule ''31-2E'' contains symmetry-related single-phosphate gaps between positions 9 and 10 and 13' and 14' (filled bars; see the text).



(a)



(b)

Figure 2. (a) Structure of the CAP-DNA complex at 2.5 Å resolution. For the DNA, the bases and backbone are shown. The protein is represented as a ribbon diagram. (b) Least-squares superimposition of the previously reported structure of the CAP-DNA complex (fine lines; Schultz *et al*., 1991) and the present structure (bold lines). Insets present enlarged views of the transcription activating region (residues 155 to 165; Zhou *et al*., 1993, 1994; Heyduk *et al*., 1993) and the helix-turn-helix motif (residues 169 to 191).

molecules of cAMP bound to the CAP dimer, and a sharply bent DNA duplex. The overall DNA bend angle is 87°. The present structure at 2.5 Å resolution is similar to the previously reported structure at 3.0 Å resolution (Schultz *et al*., 1991; Figure 2(b)). The average r.m.s. differences between the present structure and the previously reported structure are 0.53 Å for CAP backbone atoms, 1.2 Å for CAP side-chain atoms, and 0.87 Å for DNA atoms.

The fact that the CAP-DNA complex exhibits approximate 2-fold symmetry rather than perfect 2-fold symmetry appears to be a consequence of asymmetric packing environments in the crystalline lattice. The pseudo-2-fold axis of the CAP-DNA complex does not coincide with a crystallographic 2-fold axis, and therefore the packing environments of half-complexes A and B are substantially different. The packing environment of half-complex B tightly constrains the DNA conformation in half-complex B and is incompatible with the DNA conformation in half-complex A. In contrast, the packing environment of half-complex A does not tightly constrain the DNA conformation in half-complex A and is compatible with the DNA conformation in half-complex B. We infer that, in solution, the CAP-DNA complex is likely to be perfectly or nearly perfectly 2-fold symmetric and is likely to exhibit CAP conformation, DNA conformation, and CAP-DNA contacts corresponding to those in the A half-complex in the crystallographic structure.

CAP structure

In both CAP subunits, the domain-domain interface is in a ''closed'' conformation (see Weber & Steitz, 1987), and the bound cAMP molecule exhibits an *anti* glycosidic-bond conformation (Figure 2(a),(b)).

There are subtle differences in the structure of CAP in the present structure as compared to in the previously reported structure (Schultz *et al*., 1991; Figure 2(b)), especially on the protein surface, at turns, and at sites involved in packing interactions in the crystalline lattice. The largest difference is for residues 150 to 165 as shown in Table 1. There are several large differences in the π, φ values that result in a loss of hydrogen bond contacts in the loops. This difference is of importance in that residues 150 to 165 comprise the ''activating region'' of CAP, which makes direct protein-protein contact with RNA polymerase in transcriptional activation (Zhou *et al*., 1993, 1994a,b; Heyduk *et al*., 1993; Chen *et al*., 1994). In the present structure, residues 150 to 165 fold as a canonical β turn, with residues 159 through 162 being positions *i* through *i* + 4. Half-complex A has a type (2:4) turn while half-complex B has a type (6:6) turn (Sibanda *et al*., 1989). Residues 150 to 165 exhibit much higher than average temperature factors; 44 Å^2 in half-complex A, 47 Å^2 in half-complex B. Average temperature factor for surface amino acid residues $\approx$35 Å^2, indicating that residues 150 to 165 have high conformational flexibility.

DNA structure

The DNA contains two extremely sharp kinks between pseudo symmetry-related nucleotide pairs 6 and 7 in half-complex A and 16 and 17 in half complex B (''primary kinks''; Figures 2, 3(a)). The primary kinks exhibit roll angles of 52° and 35° and twist angles of 17° and 12° (Babcock & Olson, 1993; Figure 3(c)). The DNA deformation in each primary kink is remarkably localized, with all DNA deformation occurring at a single nucleotide pair-nucleotide pair step and with the flanking nucleotide pairs exhibiting the standard *B*-DNA geometry (Figure 3(c)). The primary kinks result in

Table 1. Distances and angles

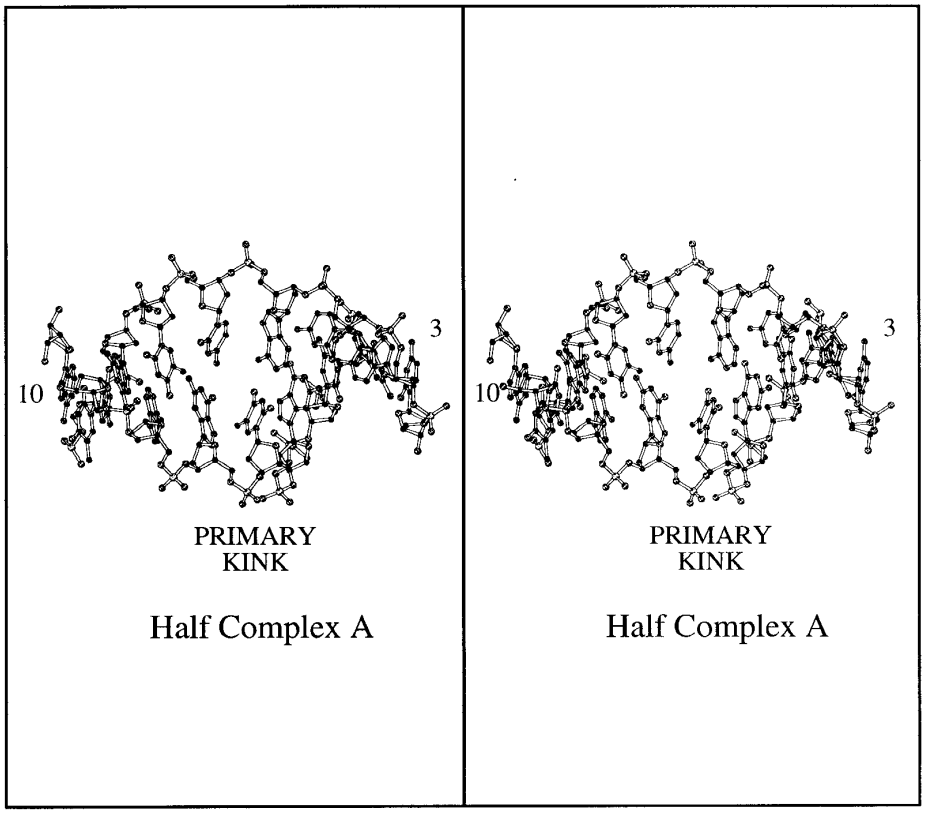
A. *Hydrogen bonding distances in the activating region*

				Distance *i-j* (Å)	
Atom *i*	Residue	Atom *j*	Residue	Half-complex A	Half-complex B
N	Met157	O	Gln164	3.2 (2.9)	3.5 (2.7)
N	His159	O	Gly162	3.4 (3.0)	4.5 (2.9)
N	Asp161	O	His159	3.1 (2.9)	3.3 (2.6)
N	Gly162	O	His159	3.9 (2.9)	4.1 (3.5)
ND1	His159	O	Asp161	2.8 (2.9)	2.9 (3.8)

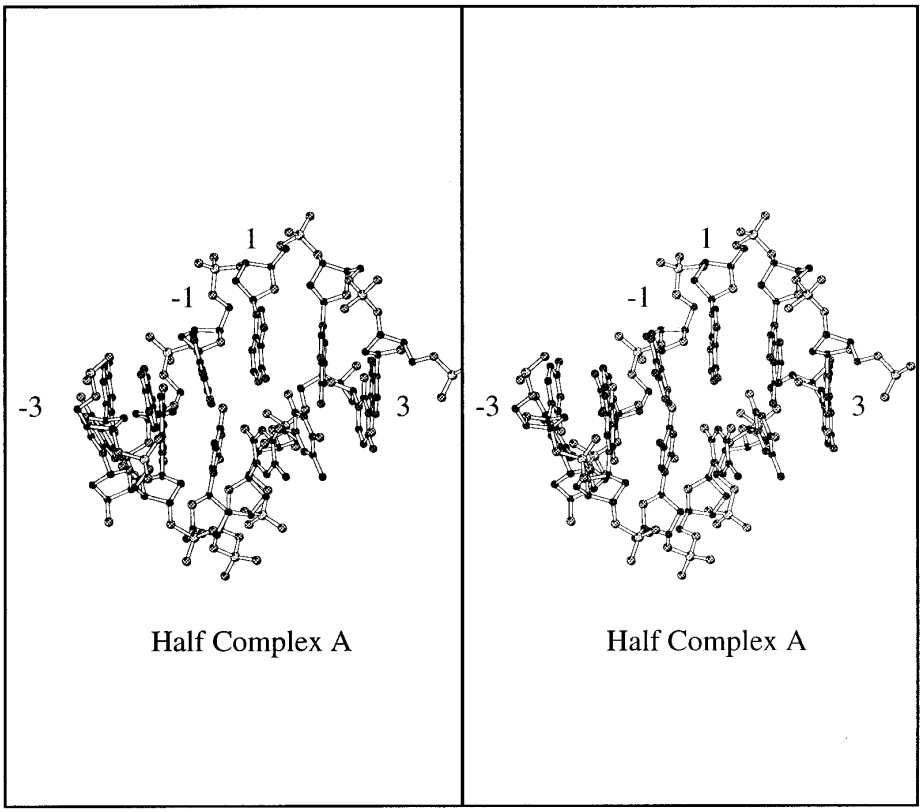
B. *Torsion angles*

	Half-complex A		Half-complex B	
Residue	φ(°)	ψ(°)	φ(°)	ψ(°)
Met157	–68 (–82)	150 (141)	–130 (–90)	128 (154)
Thr158	–82 (–52)	150 (141)	–38 (–92)	93 (150)
His159	–139 (–132)	141 (135)	–78 (–129)	150 (100)
Pro160	–40 (–53)	–59 (–36)	–43 (4)	–65 (–78)
Asp161	–112 (–129)	134 (98)	–120 (–111)	87 (109)
Gly162	68 (124)	176 (–175)	95 (123)	115 (170)
Met163	–106 (–131)	135 (136)	–48 (–100)	166 (158)
Gln164	–91 (–113)	107 (127)	–111 (–129)	–177 (124)

The numbers in parentheses are the values found in the CAP-DNA by Schultz *et al*. (1991).

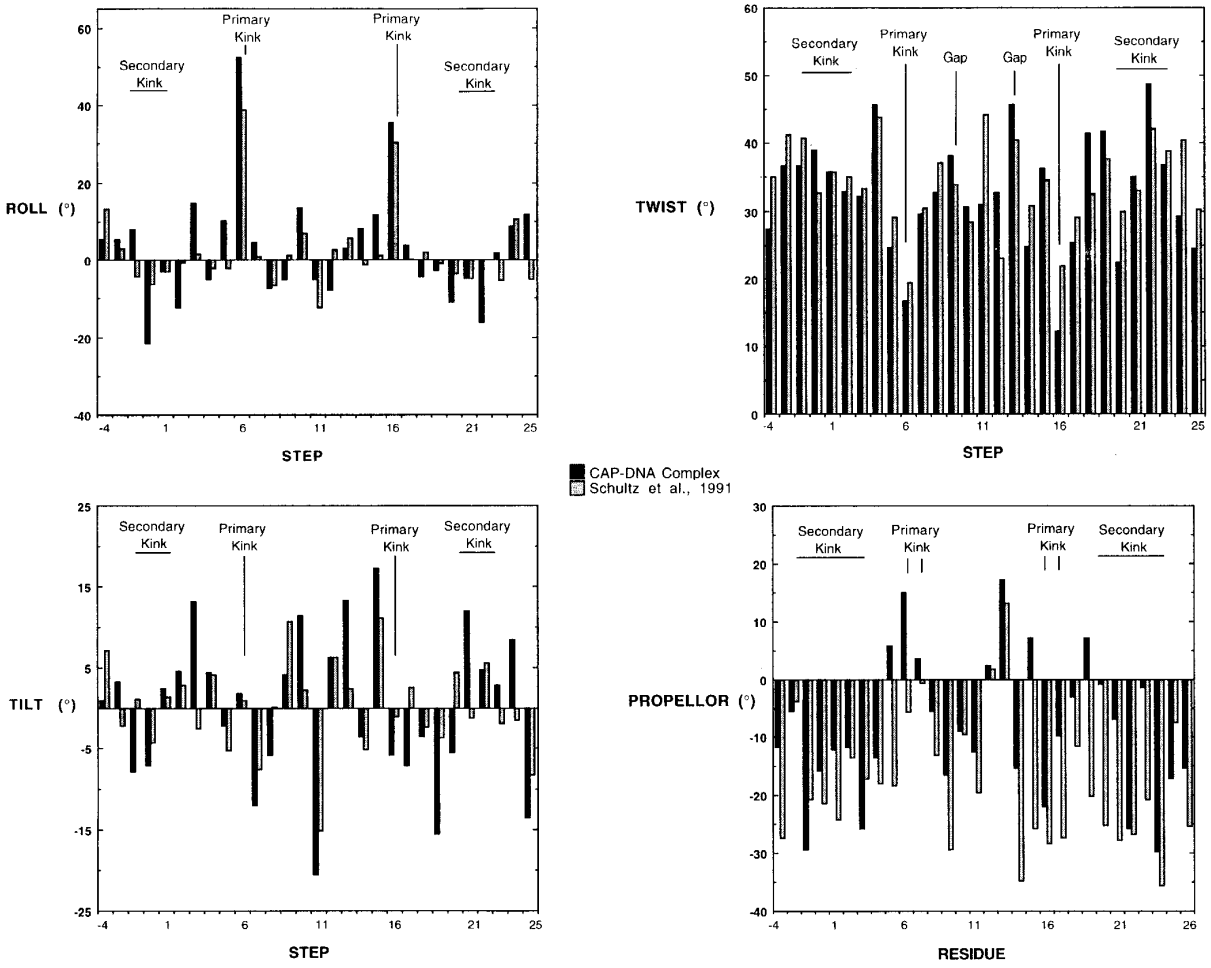


(a)



(b)

Figure 3(a–b) *legend overleaf*



(c)

Figure 3. (a) Stereo diagram of the primary DNA kink (half-complex A, positions 3 to 10). (b) Stereo diagram of the secondary DNA kink (half-complex A, positions –3 to 3). (c) DNA helical parameters: roll, twist, tilt and propeller twist (calculated as described by Babcock & Olson, 1993). Illustrations in this and subsequent Figures were prepared using MOLSCRIPT (Kraulis, 1991).

compression of the DNA major groove from 17.5 Å in standard *B*-DNA to 14.4 Å and 16.8 Å in half-complexes A and B, respectively, and expansion of the DNA minor groove from 17.5 Å in standard *B*-DNA to 19.4 Å and 20.0 Å in half-complexes A and B, respectively (distances measured from phosphate 7 to 7' and 16' to 16, respectively). Most of the DNA bend angle of 87° in the CAP-DNA complex is accounted for by the primary kinks.

The DNA contains two additional and less sharp kinks between symmetry-related nucleotide pairs –1 and 1 in half-complex A, and 22 and 23 in half-complex B (''secondary kinks''; Figures 2, 3(b)). The secondary kinks exhibit roll angles of approximately –22° and –16° (Figure 3(c)). The secondary kinks are less highly localized than the primary kinks, with DNA deformation distributed over 1-2 adjacent DNA base-pairs. The secondary kinks result in compression of the DNA minor groove from 11.5 Å in standard *B*-DNA to 6.7 Å and 8.3 Å in half-complexes A and B, respectively (distances measured from phosphate –3' to 2 and 24 to 21', respectively). The secondary kinks are located within A/T-rich DNA segments. The A/T-rich DNA segments exhibit unusually large nucleotide-pair propeller twists, and at their junctions with non-A/T-rich DNA segments, low twist angles. The A/T-rich character of these DNA segments probably facilitates CAP-induced kinking (see Gartenberg & Crothers, 1988; Dalma-Weiszhaus *et al*., 1990).

There are relatively large differences in DNA conformation between our structure and the structure previously reported (Schultz *et al*., 1991). These occur at the positions flanking the phosphates that are present in the previously reported structure but missing in our structure: nucleotides 9, 10, 13', and 14'. R.m.s. differences for these nucleotides are 0.88 Å. In terms of the base morphology, the propeller twist of the base-pairs show dramatic differences between the two

Table 2. CAP-DNA contacts

A. *Half-complex A*

Base contacts

Hydrogen bonds

Atom *i*	Residue	Chain	Atom *j*	Residue	Distance *i-j* (Å)
N7	G5	C	NH2	Arg180	3.3
O6	G5	C	NH2	Arg180	2.5
N7	G7	C	NH1	Arg185	2.8
O6	G7	C	NH1	Arg185	3.5
N4	C7'	F	OE2	Glu181	2.7

Thymine methyl interactions (<4.2 Å)

C5A	T4	C	NE	Arg180	3.8
C5A	T4	C	OE1	Gln170	3.9
C5A	T6	C	CE	Glu181	3.5
C5A	T6	C	NZ	Arg185	3.8
C5A	T8'	C	CD	Glu181	3.1

Phosphate contacts

Hydrogen bonds

Atom *i*	Residue	Chain	Atom *j*	Residue	Distance *i-j* (Å)
O2P	T10'	F	N	Val139	3.3
O2P	G9'	F	OG	Ser179	3.0
O2P	G9'	F	N	Ser179	3.3
O2P	G9'	F	CB	Thr182	3.3
O2P	G9'	F	OG1	Thr182	2.4
O2P	T-1'	F	O	His199	3.3
O1P	T-2'	F	NZ	Lys166	3.3
O1P	C-3'	F	NZ	Lys26	3.4
O2P	C-3'	F	NZ	Lys26	3.2
O1P	G3	C	NE2	Gln170	3.1
O1P	T4	C	N	Gln170	3.5
O2P	T4	C	NH1	Arg169	2.9
O2P	T4	C	N	Arg169	3.0
O1P	T6	C	NZ	Lys188	3.0

B. *Half-complex B*

Base contacts

Hydrogen bonds

Atom *i*	Residue	Chain	Atom *j*	Residue	Distance *i-j* (Å)
N7	G18'	E	NH1	Arg180	3.5
O6	G18'	E	NH1	Arg180	2.6
N7	G14	D	NH2	Arg185	2.7
O4	T15	D	NH1	Arg185	3.5
N4	C16	D	OE1	Glu181	3.4

Thymine methyl interactions (<4.2 Å)

C5A	T19'	E	NE2	Gln170	4.1
C5A	T15	E	CG	Glu181	3.9

Phosphate contacts

Hydrogens bonds

Atom *i*	Residue	Chain	Atom *j*	Residue	Distance *i-j* (Å)
O2P	T13	D	N	Val139	3.1
O1P	G14	D	N	Ser179	3.4
O2P	G14	D	OG1	Thr182	2.7
O2P	T22	D	NE2	His199	3.4
O1P	T19'	E	N	Gln170	2.9
O1P	T19'	E	N	Arg169	3.2
O2P	T19'	E	N	Arg169	2.5

structures most noticeably in the region of the primary kink.

CAP-DNA interactions: bases

Three amino acids of CAP make hydrogen bonds with DNA base edges: i.e. Arg180, Glu181, and Arg185 (Table 2; Figures 4(b), 5(a)).

The contacts by Arg180 and Glu181 in the present structure are essentially as in the previous structure (Schultz *et al*., 1991; Figure 4). In half-complex A, the guanidinium side-chain of Arg180 forms hydrogen bonds with the O6 and N7 atoms of G5, and the carboxylate side-chain of Glu181 forms a hydrogen bond with the N4 atom of C7'. In half-complex B, Arg180 and Glu181 form symmetry-related hydrogen bonds with G18' and C16.

The contacts by Arg185 differ from those in the previous structure (Schultz *et al*., 1991; Figure 4). In half-complex A, the guanidinium side-chain of Arg185 forms hydrogen bonds with the O6 and N7 atoms of G7, one water-mediated hydrogen bond with the N7 atom of G9', and another water-mediated hydrogen bond involving a different water

molecule with the phosphate of T6. In half-complex B, the guanidinium side-chain of Arg185 donates hydrogen bonds to the O4 atom of T15 and the N7 atom of G14.

CAP-DNA interactions: phosphates

As described in the Introduction, the DNA molecule in the previously reported structure (Schultz *et al*., 1991) was missing the DNA phosphates of symmetry-related nucleotides 9' and 14 (Figures 1(b), 4(a)). The DNA molecule in the present structure contains these phosphates (Figure 1(c), 4(b)) and thus permits for the first time definition of interactions involving these phosphates. The present structure shows that CAP makes extensive hydrogen-bonded interactions with these phosphates (Table 3; Figures 4(b), 5(b)). In each half-complex, the backbone NH of Ser179 forms one hydrogen bond with the phosphate in question, and the hydroxyl side-chains of Ser179 and Thr182 both form hydrogen bonds with the phosphate in question.

Discussion

Structure

The improved resolution of the present structure permitted unambiguous interpretation of the DNA geometry and of the hydrogen-bonded contacts between CAP and DNA bases and DNA phosphates. It also permitted the inclusion of water molecules in the CAP-DNA interface. The improved resolution of the present structure is attributable to: (1) use of a DNA fragment with all phosphates important for CAP-DNA interaction, (2) use of synchrotron radiation in data collection, and (3) use of an improved DNA-structure dictionary in refinement.

As described in Results, the structure of the CAP-DNA complex is not perfectly 2-fold symmetric, and the differences from perfect 2-fold symmetry appear to be due to asymmetric packing environments in the crystalline lattice. Since the conformation of half-complex A, but not half-complex B, appears unconstrained by packing environment in the crystalline lattice and thus is

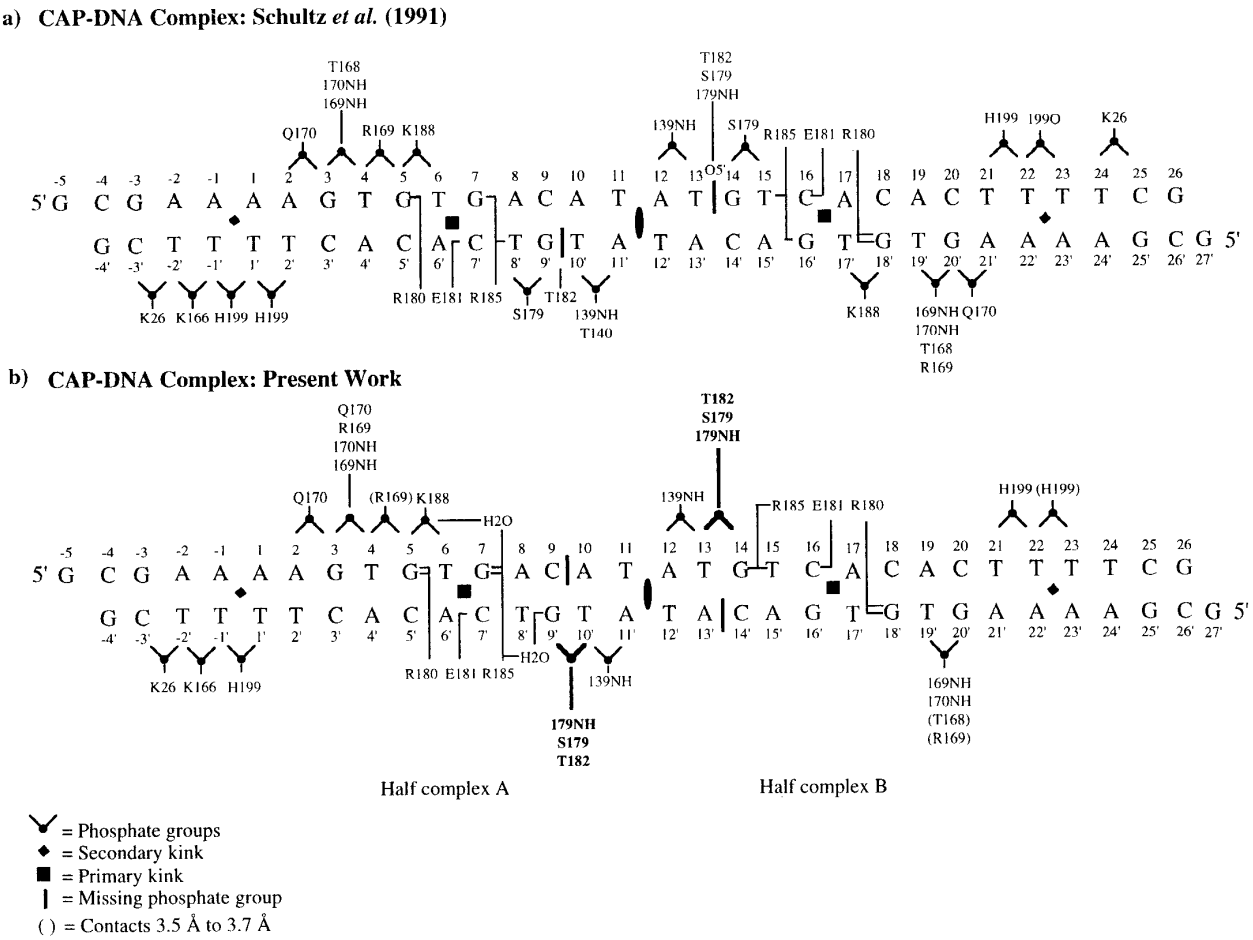


Figure 4. Summary of CAP-DNA contacts in the structure of Schultz *et al*. (1991) (see footnote to p. 000). Contacts in the half-complex A are from Schultz *et al*. (1991); contacts in the half-complex B are from the inspection of coordinates (Brookhaven Protein Data Bank accession code 1cgp). (b) Summary of the CAP-DNA contacts in the present structure. Contacts to the DNA phosphates missing in the structure of Schultz *et al* (1991) are in boldface.

likely to correspond more closely to the conformation in solution, the following discussion focuses exclusively on interactions as observed in half-complex A.

Implications for specificity

Otwinowski *et al*. (1988) have introduced a useful distinction between ''direct readout'' and ''indirect readout''. Direct readout refers to DNA-sequence recognition mediated by direct hydrogen-bonded or van der Waals interactions with DNA base-pairs. Indirect readout refers to DNA-sequence recognition mediated by the sensing of DNA-sequence-dependent effects on DNA-phosphate position, DNA-phosphate solvation, or susceptibility to DNA deformation (such as protein-induced DNA bending and protein-induced DNA twisting). In principle, base-pairs can be recognized exclusively through direct readout, exclusively through indirect readout, or by a combination of direct and indirect readout.

Direct readout

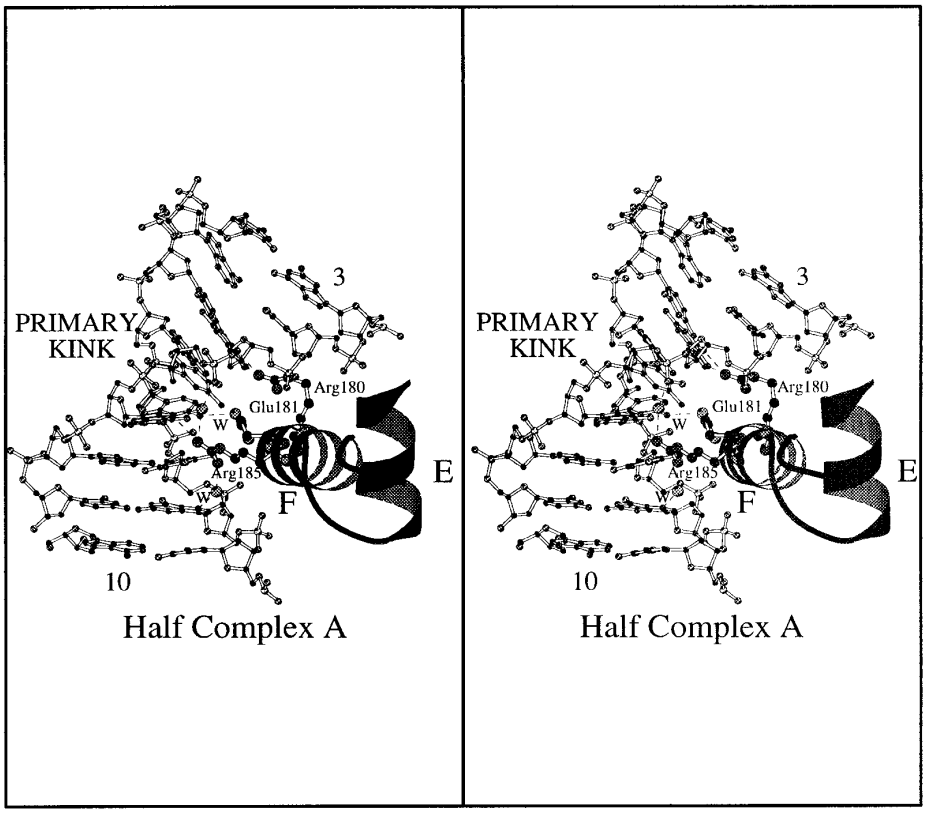
Of the four most highly conserved base-pairs in the DNA half-site, three are clear candidates for recognition through direct readout: i.e., G·C base-pair 5, G·C base-pair 7, and A·T base-pair 8.

At G·C base-pair 5, Arg180 of CAP forms H-bonds with the guanine O6 and N7 atoms (Figure 5; Table 2). The experimentally observed rank order of base-pair preferences at position 5 (G·C ≫ T·A > A·T > C·G; Gunasekera *et al*., 1992) is in agreement with rank order predicted by simple enumeration of possible hydrogen bonds by Arg180 (respectively, two hydrogen bonds, one sub-optimal hydrogen bond, one sub-optimal hydrogen bond, and zero hydrogen bonds; see Seeman *et al*., 1976). More important, removal of the side-chain of Arg180 eliminates base-pair preferences at position 5, in compelling agreement with direct readout by Arg180 (Zhang & Ebright, 1990).

At G·C base-pair 7, Glu181 of CAP forms an H-bond with the cytosine N4 atom, and Arg185 forms hydrogen bonds with the guanine O6 and N7 atom (Figure 5; Table 3). The experimentally observed rank order of base-pair preferences at position 7 (G·C ≫ T·A > A·T C·G; Gunasekera *et al*., 1992) is in agreement with the rank order predicted by enumeration of hydrogen bonds by Glu181 and Arg185 (respectively, three hydrogen bonds, two sub-optimal hydrogen bonds, one sub-optimal hydrogen bond, and no hydrogen bonds). Elimination of the side-chain of Glu181 eliminates base-pair preferences at position 7, in compelling agreement with direct readout by Glu181 (Ebright *et al*., 1984a,b, 1987; Ebright, 1991).

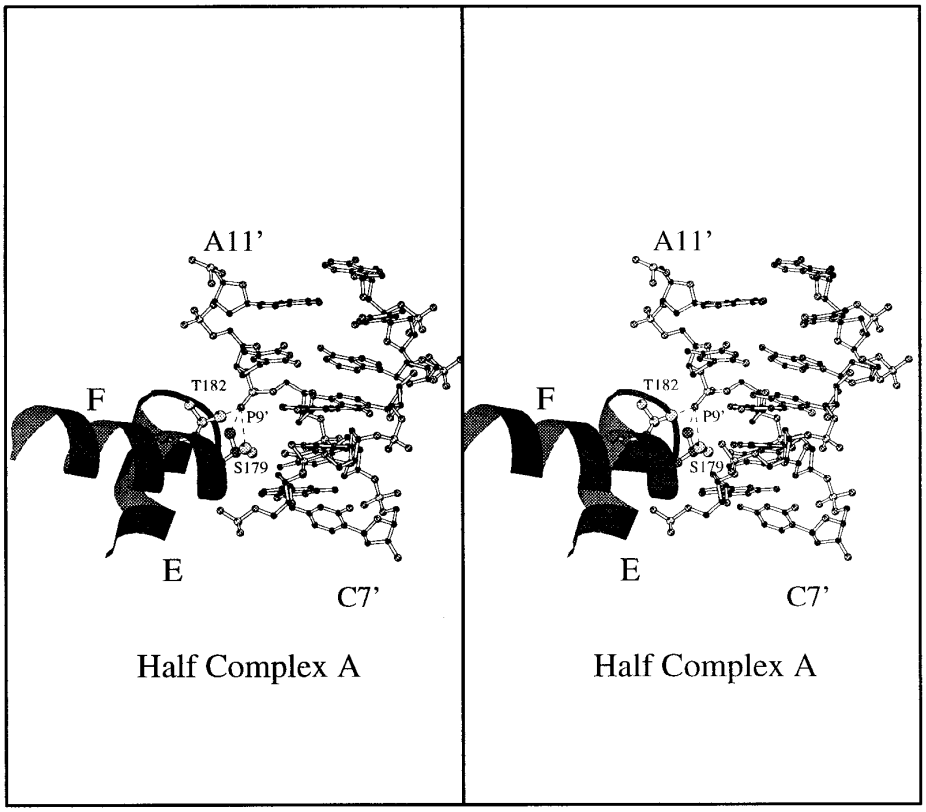
Indirect readout

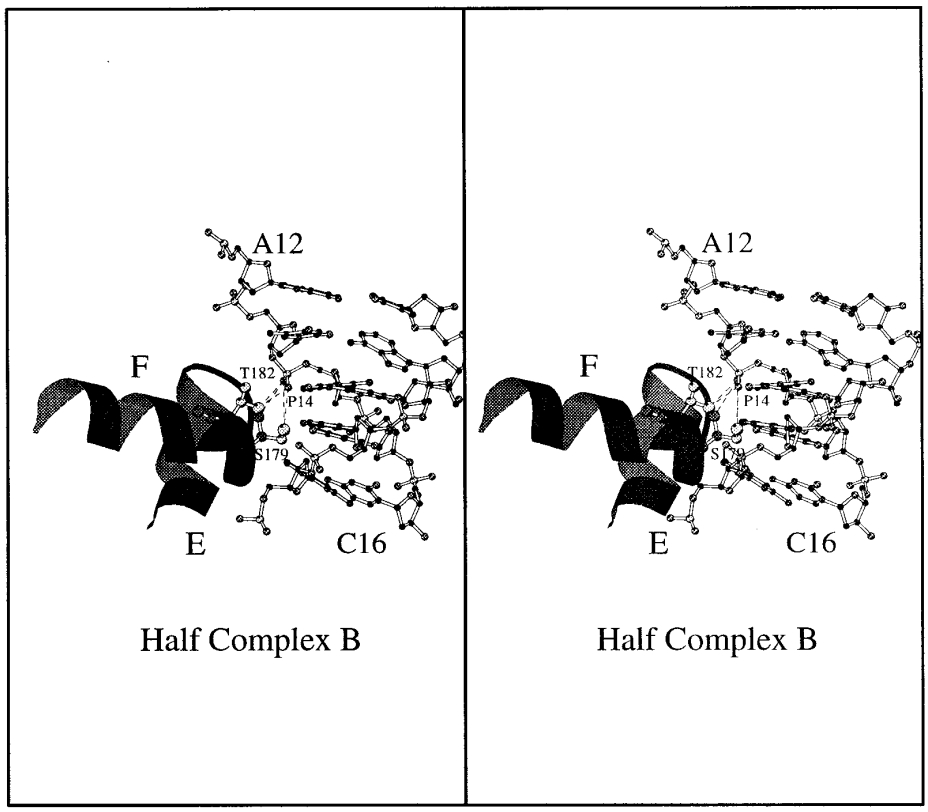
Of the four most highly conserved base-pairs in the DNA half-site of the CAP-DNA complex, T·A



(a)

Figure 5(a) *legend overleaf*





(b)

Figure 5. (a) Stereo diagram of contacts by CAP with DNA bases (half-complex A). (b) Stereo diagrams of contacts by CAP and with DNA phosphates at symmetry-related positions 9' (half-complex A) and 14 (half-complex B). These DNA phosphates were missing in the previously reported structure (Schultz *et al*., 1991).

Table 3. Crystallographic and data collection parameters

Contents of asymmetric unit	
Molecules of CAP	1 (dimer)
Molecules of DNA	1 (duplex)
V_m (Å^3)	3.0
Space group	$C222_1$
Average unit cell	
a (Å)	136.99
b (Å)	152.80
c (Å)	76.06
$\alpha = \beta = \gamma$ (°)	90
Data collection parameters	
Temperature (°C)	–15
Crystal mounting	capillary
Data collection source	CHESS
Radiation (λ) (Å)	0.908
Number of crystals used	9
Total reflections collected	165,000
Unique reflections collected	23,876
R_{sym} %[a]	11.47
Percentage of total to 2.5 Å, $\|F\| > 2\sigma$	84.9

[a] $R_{sym} = \Sigma_h \Sigma_k | < I_k - I_{ki} > | / \Sigma_h \Sigma_i I_k \times 100$.

base-pair 6 is a clear candidate for recognition exclusively through indirect readout.

At T·A base-pair 6, CAP does not make any hydrogen bonds with base-pair functional groups (Figure 5). T·A base-pair 6, together with G·C base-pair 7, form the primary CAP-induced DNA kink. We propose that specificity for T·A at position 6 results from either: (1) effects of T·A at position 6 on the orientation of G·C base-pair 7 and thus on the ability to make, or the free energy contribution of, the contacts to G·C base-pair 7 by Glu181 and Arg185; and/or (2) effects of T·A at position 6 on susceptibility to DNA kinking. With respect to mechanism (1), it is noteworthy that elimination of the side-chain of Glu181 reduces or eliminates specificity at position 6 (Ebright *et al*., 1987; Gunasekera & Ebright, unpublished data). With respect to mechanism (2), it is noteworthy that T·A base-pair 6 is part of a T·A-G·C dinucleotide-pair step, and is one that has been proposed to exhibit higher-than-average susceptibility to DNA kinking (McNamara, *et al*., 1990; Zhurkin *et al*., 1991; Beutel & Gold, 1992; Lyubchenko *et al*., 1993).

Prospect

The crystallographic structure of the CAP-DNA complex at 2.5 Å resolution provides baseline information for further analysis of DNA binding, DNA bending, and transcription activation by CAP. Determination of the crystallographic structure of mutant CAP-DNA complexes containing mutant CAP derivatives and/or mutant DNA-site derivatives is in progress.

Materials and Methods

CAP

CAP was purified by cAMP-affinity chromatography as described by Zhang *et al*. (1991) except that the third column wash was with 500 mM potassium phosphate (pH 7.2), 0.1 mM EDTA, 2 mM dithiothreitol, 0.02% (w/v) NaN_3, and the column elution was with the same buffer plus 5 mM cAMP. CAP was stored in elution buffer in aliquots at –70°C. The typical yield was 10 mg per liter of bacterial culture (protein estimated as described by Bradford, 1976), and the typical purity was greater than 98%.

Immediately before use in crystallization, 0.7 ml samples of CAP were further purified by gel-filtration chromatography on two consecutive 2.5 ml columns of Bio-Gel P-6DG (Bio-Rad, Inc.) in 5 mM Tris-HCl (pH 7.5), 200 mM NaCl, 0.1 mM EDTA, 2 mM dithiothreitol, 0.02% (w/v) NaN_3, and concentrated to 0.2 mM by centrifugal ultrafiltration (Centricon 30 filter unit, Amicon, Inc.; 5000 ***g***; 80 minutes at 4°C).

DNA

The 17 nucleotide oligodeoxyribonucleotide 5'-ATAT-GTCACACTTTTCG-3'and the 14 nucleotide oligodeoxy-ribonucleotide 5'-GCGAAAAGTGTGAC-3' were synthesized using solid-phase β-cyanoethylphosphoramidite chemistry on an AB380A automated synthesizer (Applied Biosystems Inc.). Oligodeoxyribonucleotides

Table 4. Refinement statistics

Atoms included in refinement	6120		
Solvent molecules included in refinement	286		
Reflections with $\|F\| > 2\sigma$ between 10.0 Å and 2.5 Å	23,644		
Reflections possible between 10.0 Å and 2.5 Å	27,527		
Deviations from ideal geometry	Combined	Protein	DNA
Bond lengths (Å)	0.017	0.018	0.015
Bond angles (°)	2.27	2.31	2.11
Torsion angles (°)	22.9	22.9	23.0
Improper torsion angles (°)	1.4	1.3	1.5
Average overall isotropic B factor (Å^2)	30.9		
Average B factor for solvent molecules (Å^2)	40.3		

Statistics	Resolution range (Å)	Percent observed $\|F\| > 2\sigma$	R-factor[a] (%)
	10.0–3.91	89	15.8
	3.91–3.13	91	20.8
	3.13–2.75	86	25.0
	2.75–2.50	77	29.5
Cumulative		86	19.9

[a] $R = \Sigma[|F_o| - |F_c|] / \Sigma|F_o|$.

were purified by trityl-on C18 reversed-phase HPLC followed by trityl-off C18 reversed-phase HPLC. Oligodeoxyribonucleotide concentrations were determined from absorbance at 260 nm using calculated extinction coefficients (Fasman, 1977; 160,000 M^{-1} cm^{-1} for the 17 nucleotide oligodeoxyribonucleotide; 140,000 M^{-1} cm^{-1} for the 14 nucleotide oligodeoxyribonucleotide). Oligodeoxyribonucleotides were annealed to yield DNA molecule ''31-2E'' by combining 0.125 μmol 17 nucleotide oligodeoxyribonucleotide and 0.125 μmol 14 nucleotide oligodeoxyribonucleotide in 65 μl solution containing 5 mM sodium cacodylate (pH 7.4), 200 mM NaCl, 5 mM EDTA, heating for five minutes at 90°C, and then allowing it to cool over 12 to 15 hours to 22°C.

CAP-DNA crystals

Crystals were grown by vapor diffusion as sitting or hanging drops using conditions similar to those of Schultz *et al*. (1990, 1991). Wells (1 ml) contained 50 mM Mes-NaOH (pH 6.5), 200 mM NaCl, 100 mM $MgCl_2$, 100 mM $CaCl_2$, 0.02% (w/v) NaN_3, 2 mM dithiothreitol, 2 mM spermine, 2 mM cAMP, 0.3% (w/v) *n*-octyl-β-D-glucopyranoside, and 6.5% (w/v) polyethylene glycol 3350. Drops (3 to 20 μl) contained 0.1 to 0.2 mM CAP and a 1.2- to 1.5-fold molar excess of DNA molecule 31-2E in 50% well buffer. Crystals grew to approximately 0.4 mm × 0.3 mm × 0.25 mm in 30 to 72 hours at 19°C. Crystals were stabilized by successive transfers into stabilization solutions as described by Schultz *et al*. (1990, 1991) except that stabilization was carried out at 4°C over a period of one hour. Immediately before data collection, crystal temperature was reduced from 4°C to –15°C.

Data collection

X-ray crystallographic data were collected on the F1 beam line at the Cornell High Energy Synchrotron Source (CHESS), using a fixed monochromatic X-ray source of wavelength 0.91 Å. Data were collected on Fuji image plates or film using a rotation camera (Arndt, 1968) with oscillation steps of 1.5° with a 0.2° overlap, and collimation of 0.1 mm. The film-to-detector distance was 240 mm, and the crystal temperature was –15°C. One to nine image plate exposures were recorded at each crystal position, with exposure times of 6 to 12 seconds. Typically, crystals decayed from 2.4 Å to 2.7 Å resolution during each set of exposures. Nine crystals were used and 71 image plates and films were generated. A summary of crystallographic data and collection parameters is presented in Table 3.

Processing, indexing and refinement of image plates and films were performed as described (Rossmann, 1979, 1985; Rossmann *et al*., 1979). Merging, scaling, and post-refinement of data were done using programs developed specifically for synchrotron sources (Vriend *et al*., 1986).

Data analysis

Initial phases were derived from the previously reported structure at 3 Å (Schultz *et al*., 1991). The model included residues 9 through 205 of CAP and all residues of DNA; the initial *R*-factor was 36%. Rigid-body refinement, including data with $F > 2\sigma$ to 3.5 Å resolution, was followed by a combination of Powell positional refinement and molecular dynamics using the X-PLOR slow cool protocol (Brünger, 1992). The initial annealing temperature of 2000 K was reduced during subsequent cycles of refinement. Individual weights were applied to both the protein and DNA constraints and were monitored by checking the r.m.s. deviations of bond distances, bond angles and dihedral angles. Rescaling of constraints was applied to ensure balanced refinement between the protein and the DNA. (A scale of approximately 0.65 was found appropriate in most cases.) Manual rebuilding, using CHAIN (Sack, 1988), was used to improve the DNA backbone and protein side-chain geometries. Omit maps, in which specific areas of the protein and all DNA atoms were systematically removed, were calculated as a method of checking each section of the structure. After several cycles of refinement, the *R*-factor was 22.5% for data between 10 to 2.5 Å before the addition of water molecules.

Water molecules were included when three criteria were met: (1) the spherical difference density map peaks ($F_o - F_c$) were greater than 3σ, (2) the peaks overlapped observed density in the $2F_o - F_c$ maps, and (3) the modeled waters were within hydrogen bonding distance of acceptor or donor atoms on DNA or protein. Individual water temperature factors were refined and water molecules removed when the temperature factor was greater than 55 $Å^2$.

In the final stages of refinement, a new dictionary for DNA parameters was used (Parkinson *et al*., 1996). This resulted in lower r.m.s. deviations and improved electron density maps. The final refinement statistics are presented in Table 4. The final *R*-factor was 19.9% for 23,644 reflections, with $F > 2\sigma$ between 10 and 2.5 Å resolution. Ramachandran analysis using the program PROCHECK (Laskowski *et al*., 1992) indicated that 81% of the residues were in the most favored region and 2% within the generously allowed region. Only one non-glycine residue, Glu78 of half-complex A, was in a disallowed region. Analysis of the DNA geometry, using programs developed by the Nucleic Acid Database (Berman *et al*., 1992), indicates that all parameters were within acceptable ranges.

Acknowledgements

We thank Drs D. Oren and E. Arnold for assistance in the initial phase of the work, and Dr J. Vojtechovsky for assistance in the completion of the overall project. This work was supported by National Institutes of Health grants GM41376 to R.H.E. and GM21589 to H.M.B.

References

Arndt, U. W. (1968). The optimum strategy in measuring structure factors. *Acta Crystallog. sect. B,* **24**, 1355–1357.

Babcock, M. S. & Olson, W. K. (1993). A new program for the analysis of nucleic acid structure: implications for nucleic acid structure interpretation. In *Computation of Biomolecular Structures: Achievements, Problems, and Perspectives* (Soumpasis, D. M. & Jovin, T. M., eds), pp. 65–85, Springer Verlag, Heidelberg.

Berg, O. & von Hippel, P. (1988). Selection of DNA binding sites by regulatory proteins. The binding specificity of cyclic AMP receptor protein to recognition sites. *J. Mol. Biol.* **200**, 709–723.

Berman, H. M., Olson, W. K., Beveridge, D. L., Westbrook, J., Gelbin, A., Demeny, T., Hsieh, S.-H.,

Srinivasan, A. R. & Schneider, B. (1992). The Nucleic Acid Database—a comprehensive relational database of three-dimensional structures of nucleic acids. *Biophys. J.* **63**, 751–759.

Beutel, B. A. & Gold, L. (1992). In vitro evolution of intrinsically bent DNA. *J. Mol. Biol.* **228**, 803–812.

Bradford, M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye-binding. *Anal. Biochem.* **72**, 248–254.

Brennan, R. G. (1992). DNA recognition by the helix-turn-helix motif. *Curr. Opin. Struct. Biol.* **2**, 100–108.

Brünger, A. T. (1992). *X-PLOR Manual, Version 3.1*. Yale University Press, New Haven, CT.

Chen, Y., Ebright, Y. W. & Ebright, R. H. (1994). Identification of the target of a transcription activator protein by protein-protein photocrosslinking. *Science*, **265**, 90–92.

Dalma-Weiszhausz, D. D., Gartenberg, M. R. & Crothers, D. M. (1990). Sequence-dependent contribution of distal binding domains to CAP protein-DNA binding affinity. *Nucl. Acids Res.* **19**, 611–616.

Ebright, R. H. (1991). Identification of amino acid-base pair contacts by genetic methods. In *Methods in Enzymology* (Sauer, R. T., ed.), pp. 620–640, Academic Press, Inc., San Diego.

Ebright, R. H., Cossart, P., Gicquel-Sanzey, B. & Beckwith, J. (1984a). Molecular basis of DNA sequence recognition by the catabolite gene activator protein: detailed inferences from three mutations that alter DNA sequence specificity. *Proc. Natl Acad. Sci. USA*, **81**, 7274–7278.

Ebright, R. H., Cossart, P., Gicquel-Sanzey, B. & Beckwith, J. (1984b). Mutations that alter the DNA sequence specificity of the catabolite gene activator protein of *E. coli*. *Nature*, **311**, 232–235.

Ebright, R. H., Kolb, A., Buc, H., Kunkel, T. A., Krakow, J. S. & Beckwith, J. (1987). Role of glutamic acid-181 in DNA-sequence recognition by the catabolite gene activator protein (CAP) of *Escherichia coli*: altered DNA-sequence-recognition properties of [Val^{181}]CAP and [Leu^{181}]CAP. *Proc. Natl Acad. Sci. USA*, **84**, 6083–6087.

Ebright, R. H., Ebright, Y. W. & Gunasekera, A. (1989). Consensus DNA site for the *Escherichia coli* catabolite gene activator protein (CAP): CAP exhibits a 450-fold higher affinity for the consensus than for the *E. coli lac* DNA site. *Nucl Acids Res.* **17**, 10295–10305.

Fasman, G. D. (1977). *CRC Handbook of Biochemistry and Molecular Biology vol. 1*, CRC Press, Cleveland, Ohio.

Gartenberg, M. R. & Crothers, D. M. (1988). DNA sequence determinants of CAP-induced bending and protein binding affinity. *Nature*, **333**, 824–829.

Gorin, A. A., Zhurkin, V. B. & Olson, W. K. (1995). *B*-DNA twisting correlates with base-pair morphology. *J. Mol. Biol.* **247**, 34–48.

Gunasekera, A., Ebright, Y. W. & Ebright, R. H. (1992). DNA sequence determinants for binding of the *Escherichia coli* catabolite gene activator protein. *J. Biol. Chem.* **267**, 14713–14720.

Heyduk, T., Lee, J. C., Ebright, Y. W., Blatter, E. E., Zhou, Y. & Ebright, R. H. (1993). CAP interacts with RNA polymerase in solution in the absence of promoter DNA. *Nature*, **364**, 548–549.

Ivanov, V. I., Minchenkova, L. E., Chernov, B. K., McPhie, P., Ryu, S., Garges, S., Barber, A. M., Zhurkin, B. V. & Adhya, S. (1995). CRP-DNA complexes: inducing the A-like form in the binding sites with an extended central spacer. *J. Mol. Biol.* **245**, 228–240.

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

Kotlarg, D., Fritsch, A. & Buc, H. (1986). Variations of intramolecular ligation rates allow the detection of protein-induced bends in DNA. *EMBO J.* **5**, 799–803.

Kraulis, P. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallog.* **24**, 946–950.

Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1992). *Procheck v.2.0 – Programs to Check the Stereochemical Quality of Protein Structures*, University College, London.

Liu-Johnson, H.-N., Gartenberg, M. R. & Crothers, D. M. (1986). The DNA binding domain and bending angle of *E. coli* CAP protein. *Cell*, **47**, 995–1005.

Lyubchenko, Y., Shlyatenko, L., Appela, E. & Harrington, R. (1993). CA runs increase DNA flexibility in the complex of 1 cro protein with the OR_3 site. *Biochemistry*, **32**, 4121–4127.

Majors, J. (1977). Control of the *E. coli* lac operon at the molecular level. Doctoral Dissertation, Harvard University.

McNamara, P. T., Bolshow, A., Tritinov, E. N. & Harrington, R. E. (1990). Sequence-dependent kinks induced in curved DNA. *J. Biomol. Struct. Dynam.* **8**, 529–538.

Otwinowski, Z., Schevitz, R. W., Zhang, R.-G., Lawson, C. L., Joachimiak, A., Marmorstein, R. Q., Luisi, B. F. & Sigler, P. B. (1988). Crystal structure of *trp* repressor/operator complex at atomic resolution. *Nature*, **335**, 321–329.

Pabo, C. O. & Sauer, R. T. (1992). Transcription factors: Structural families and principles of DNA recognition. *Annu. Rev. Biochem.* **61**, 1053–1095.

Parkinson, G., Vojtechovsky, J., Clowney, L., Brünger, A. T. & Berman, H. M. (1996). New parameters for the refinement of nucleic acid containing structures. *Acta Crystallog. sect. D*, **52**, 57–64.

Porschke, D., Hillen, W. & Takahashi, M. (1984). The change of DNA structure by specific binding of the cAMP receptor protein from rotation diffusion and dichroism measurements. *EMBO J.* **3**, 2873–2878.

Rossmann, M. G. (1979). Processing oscillation diffraction data for very large unit cells with an automatic convolution technique and profile fitting. *J. Appl. Crystallog.* **12**, 225–238.

Rossmann, M. G. (1985). Determining the intensity of Bragg reflections from oscillation photographs. In *Methods in Enzymology* (Wyckhoff, H. W., Hirs, C. H. W. & Timassheff, S. N. eds), pp. 237–280, Academic Press, Inc., New York.

Rossmann, M. G., Leslie, A. G. W., Abdel-Meguid, S. S. & Tsukihara, T. (1979). Processing and post-refinement of oscillation camera data. *J. Appl. Crystallog.* **12**, 570–581.

Sack, J. S. (1988). CHAIN – a crystallographic modeling program. *J. Mol. Graphics*, **6**, 224–225.

Schultz, S. C., Shields, G. C. & Steitz, T. A. (1990). Crystallization of *Escherichia coli* catabolite gene activator protein with its DNA binding site. *J. Mol. Biol.* **213**, 159–166.

Schultz, S. C., Shields, G. C. & Steitz, T. A. (1991). Crystal structure of a CAP-DNA complex: The DNA is bent by 90°. *Science*, **253**, 1001–1007.

Seeman, N. C., Rosenberg, J. M. & Rich, A. (1976). Sequence specific recognition of double helical

nucleic acids by proteins. *Proc. Natl Acad. Sci. USA*, **73**, 804–808.

Shanblatt, S. H. & Revzin, A. (1986). The binding of catabolite activator protein and RNA polymerase to the *Escherichia coli* galactose and lactose promoters probed by alkylation interference studies. *J. Biol. Chem*. **261**, 10885–10890.

Sibanda, L. B., Blundell, L. T. & Thornton, J. M. (1989). Confomation of B-hairpins in protein structures. A systematic classification with applications to modeling by homology, electron density fitting and protein engineering. *J. Mol. Biol.* **206**, 759–777.

Vriend, G., Rossmann, M. G., Arnold, E., Luo, M. & Griffith, J. (1986). Post-refinement of oscillation diffraction data collected at a synchrotron radiation source. *J. Appl. Crystallog.* **19**, 134–139.

Weber, I. T. & Steitz, T. A. (1987). Structure of a complex of catabolite gene activator protein and cyclic AMP refined to 2.5 Å resolultion. *J. Mol. Biol.* **198**, 311–326.

Wu, H.-M. & Crothers, D. M. (1984). The locus of sequence-directed and protein-induced DNA bending. *Nature*, **308**, 509–513.

Zhang, X. & Ebright, R. H. (1990). Identification of a contact between arginine-180 of the catabolite gene activator protein (CAP) and base-pair 5 of the DNA site in the CAP-DNA complex. *Proc. Natl Acad. Sci. USA*, **87**, 4717–4721.

Zhang, X. P., Gunasekera, A., Ebright, Y. W. & Ebright, R. H. (1991). Derivatives of CAP having no solvent-accessible cysteine residues, or having a unique solvent-accessible cysteine residue at amino acid 2 of the helix-turn-helix motit. *J. Biomol. Struct. Dynam.* **9**, 463–473.

Zhou, Y., Zhang, X. & Ebright, R. E. (1993), Identification of the activating region of catabolite activator protein (CAP): isolation and characterzation of mutants of CAP specifically defective in transcription activation. *Proc. Natl Acad. Sci. USA*, **90**, 6081–6085.

Zhou, Y., Merkel, T. J. & Ebright, R. H. (1994a). Characterization of the activating region of *Escherichia coli* catabolite gene activator protein (CAP) II. Role at class I and class II CAP-dependent promoters. *J. Mol. Biol.* **243**, 603–610.

Zhou, Y., Pendergrast, P. S., Bell, A., Williams, R., Busby, S. & Ebright, R. H. (1994b). The functional subunit of a dimeric transcription activator protein depends on promoter architecture. *EMBO J.* **13**, 4549–4557.

Zhurkin, V. B., Gorin, A. A., Ulyanov, N. B. & Jernigan, R. L. (1991). Static and statistical bending of DNA evaluated by Monte Carlo simulations. *Proc. Natl Acad. Sci. USA*, **88**, 7046–7050.

Edited by F. E. Cohen

(Received 25 January 1996; received in revised form 25 April 1996; accepted 1 May 1996)