Proc. Natl. Acad. Sci. USA
Vol. 81, pp. 3973–3977, July 1984
Biochemistry

Model of specific complex between catabolite gene activator protein and B-DNA suggested by electrostatic complementarity

(protein–DNA specificity/DNA bending/x-ray crystallography)

IRENE T. WEBER AND THOMAS A. STEITZ

Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT 06511

Communicated by Frederic M. Richards, February 24, 1984

ABSTRACT **Calculation of the electrostatic potential energy surfaces of *Escherichia coli* catabolite gene activator protein (CAP) dimer suggests a model for the complex between CAP and a specific DNA sequence. The positive electrostatic charge density of CAP lies on the two COOH-terminal domains and about 20–30 Å from the molecular 2-fold axis. Assuming that the 2-fold axes of the CAP dimer and the DNA to which it binds are coincident, the positions of the positive electrostatic potential surfaces strongly suggest the rotational orientation of the DNA relative to the protein. A specific complex between CAP and its DNA binding site in the *lac* operon has been built with the DNA in this orientation. The amino ends of the two protruding F α-helices interact in successive major grooves of the DNA. Four side chains emanating from each F helix can form hydrogen bonds with the exposed edges of four bases in the major groove. Electrostatic considerations as well as the necessity to make interactions between CAP and a DNA site as much as 20 base pairs long require us to bend or kink the DNA. In our model of CAP complexed with B-DNA, as with those proposed for Cro and λcI repressors, the protruding second helices of the two-helix motif from both subunits interact in successive major grooves of B-DNA. However, unlike Cro and similar to λcI, the protruding α-helices are nearly parallel to the bases rather than the groove.**

The catabolite gene activator protein (CAP), also known as the cyclic AMP receptor protein (CRP), regulates transcription from many operons in *Escherichia coli* (1–3). In the presence of the allosteric effector cAMP, CAP binds to specific DNA sites at or near the promoter. CAP stimulates transcription from some operons, such as the *lac* and *gal* operons, while it inhibits transcription from others, such as the *ompA* (4) and *crp* genes (5). CAP also binds nonspecifically to DNA in both the presence and absence of cAMP. The structure of the 45,000-Da CAP dimer complexed with cAMP has been solved at 2.9-Å resolution (6, 7), and the coordinates have been partially refined at 2.5-Å resolution (unpublished data). The CAP subunit consists of a large NH_2-terminal domain that binds cAMP, and a smaller COOH-terminal domain is implicated in DNA binding.

The crystal structures of CAP (6, 7), λ phage cro protein (8), and a proteolytic fragment of λcI repressor (9), three proteins that regulate transcription, have led to a number of general conclusions concerning the structural basis of DNA sequence recognition by proteins. The structures of CAP and cro contain an identical two α-helix structure (10) as does the NH_2-terminal proteolytic fragment of λcI repressor (11). Sequence homologies found between these proteins and other transcription regulators (12–14) suggest that this two-helix motif will be found in many repressors and activators. Mutations in surface residues that alter DNA binding lie in this two-helix motif in λcI protein (15) as do mutations that abolish its ability to activate transcription (16). Thus, it is likely that the two-helix motif is directly involved in DNA sequence recognition. A detailed model of cro interacting with its operator DNA has been presented (17), and less detailed models for CAP (18) and λcI interaction (19) have been published. We conclude here that all three proteins interact with DNA in related, though not identical, ways.

McKay and Steitz (6) noted that the shape of the α-carbon backbone structure of CAP is complementary to the structure of a left-handed B-DNA, and they suggested that CAP might bind to left-handed B-DNA with one protruding α-helix from each subunit fitting into successive major grooves. They pointed out that if specific CAP binding converts two turns of DNA from right- to left-handed, the super helix density of closed circular DNA would be altered. Kolb and Buc (20) measured the change that occurs in the super helix density on binding CAP to specific sites in closed circular DNA and found very little change, which demonstrates that CAP binds to right-handed B-DNA.

Here we use the electrostatic complementarity between CAP and DNA to determine the orientation of B-DNA relative to the protein. The electrostatic charge distribution is calculated using the solvent accessibility modified Tanford–Kirkwood theory that was developed by Gurd and co-workers (21). The method has previously been applied to ribonucleases (22) and to a complex of flavodoxin and cytochrome *c* (23). Our calculations strongly suggest the orientation of DNA when bound to CAP and are also consistent with the possibility that CAP is bending or kinking the DNA. Specific side-chain interactions between CAP and DNA are proposed. Preliminary accounts of the electrostatic calculations and model building with B-DNA were presented at the 1982 Cold Spring Harbor Symposium (18) and at the 1983 Biostereodynamic Symposium (24).

RESULTS

Electrostatic Complementarity in CAP–DNA Model. The positive electrostatic charge density of the CAP dimer lies on the two COOH-terminal small domains and extends along the outside of the two protruding F α-helices (Fig. 1*a*). It is similar, but not identical, on the two subunits due to the nonequivalent conformation of the subunits. The net negative electrostatic charge potential lies on the cAMP-binding NH_2-terminal domains (18). The positive electrostatic charge density is concentrated away from the molecular symmetry axis and is located more on the sides of the DNA binding domains rather than on the top, as viewed in Fig. 1*a*. The location of positive electrostatic charge potential strongly suggests the orientation of B-DNA bound to the small domains of CAP. If we assume that in the CAP–DNA complex the approximate 2-fold axis of the CAP dimer is coincident with the approximate 2-fold axis of DNA, then there are only two parameters

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: bp, base pair(s); CAP, catabolite gene activator protein.

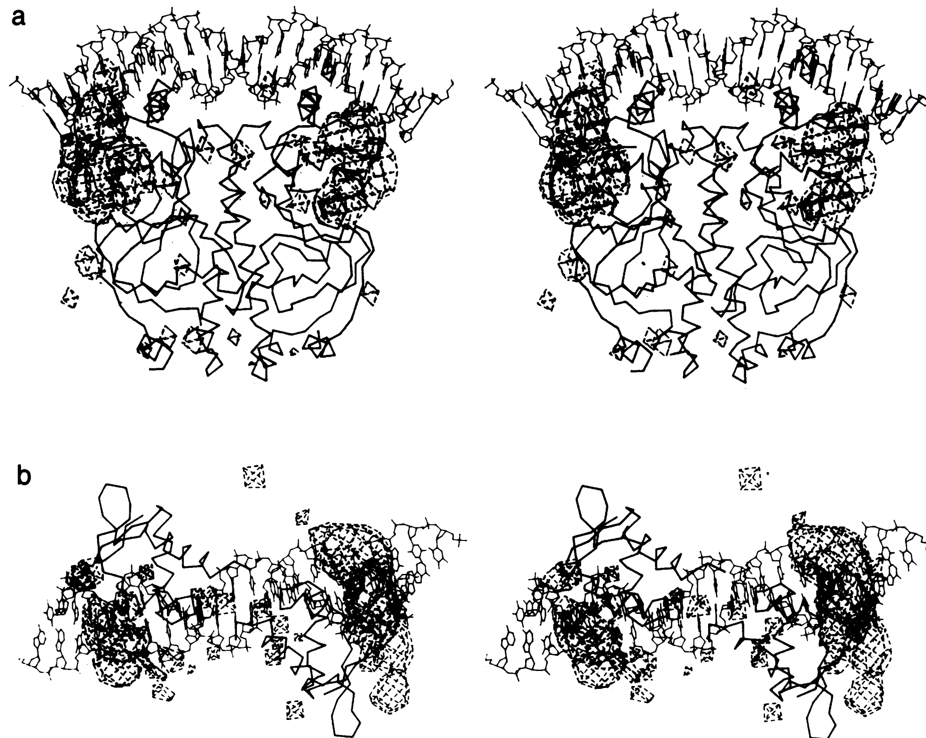

FIG. 1. (*a*) The positive (dashed lines) electrostatic potential energy surfaces of the CAP dimer with cAMP are shown together with the α-carbon backbone of CAP and 24 bp of B-DNA (25) bent to a radius of curvature of 70 Å. The electrostatic potential surfaces were calculated using a modification (21, 22) of the Tanford–Kirkwood theory (26) and are contoured at levels of 2 kT outside the van der Waals surface of the protein. The charged sites were taken from the crystallographic coordinates of the titratable amino acid side chains and the phosphates of the two bound molecules of cAMP. The electrostatic work factors were taken from ref. 27, and the calculation was performed for an ionic strength of 0.01 and at pH 7.0. (*b*) The CAP–DNA complex and the positive electrostatic potential are viewed along their mutual 2-fold axis. Only the COOH-terminal domains of CAP are shown. The DNA is oriented to best overlap the positive electrostatic potential of CAP in this projection. The model was built with an Evans and Sutherland Picture System 2 Interactive Graphics interfaced to a PDP 11/70 computer. The molecular graphics computer program FRODO was modified by Steve Anderson to allow the adjustment of the DNA relative to the protein.

left to relate the DNA and the protein—their relative rotational orientation and the distance between them. Only one relative orientation of DNA and protein strongly overlaps the negative electrostatic potential surface of the DNA with the positive electrostatic potential surfaces of the protein (18) (Fig. 1*b*), and we have used it in building the specific complex. This orientation also maximizes the interactions between the F α-helices and the major grooves of B-DNA.

A Specific CAP–DNA Complex. In the model for the complex between CAP and the specific DNA sequence to which it binds in the *lac* operon, hydrogen bonds are formed between side chains of the protein and the exposed edges of base pairs in the major groove and to the sugar–phosphate backbone. The electrostatic free energy calculated for the formation of this complex of CAP and straight DNA (28) is −10.5 kcal/mol (1 cal = 4.184 J) at pH 7.0 and 0.01 ionic strength. This was calculated by subtracting the electrostatic energy of the two separate molecules from that of the complex.

The number of interactions between CAP and DNA can be increased by bending or kinking the DNA so that it contacts more of the protein surface and also is closer to the positive electrostatic potential surface. The DNA may be smoothly bent to some appropriate radius of curvature, as we have done. However, it must be recognized that it could be singly or multiply kinked at specific locations to result in a bend of somewhat different conformation. Bending the DNA lowers the electrostatic free energy of complex formation by ≈1 kcal/mol, to −11.4 kcal/mol. Furthermore, a smooth bend to a radius of 70 Å allows CAP to make additional interactions over 20–21 base pairs (bp) of DNA. Two arginine-180 side chains, one from each F helix, now make hydrogen bond interactions with two guanine bases instead of interacting mainly with the phosphates. Bending the DNA also allows additional interactions to the sugar phosphate backbone at the extreme ends of the site by lysine-201 and glutamine-170. Although some or all of the increased contact between CAP and DNA could result from a change in the CAP structure, the number of possible conformational changes is too large to be explored usefully by model building.

In this model with bent DNA (Figs. 2–5) there are 14 hydrogen bonds that have been made between 8 protein side chains and the exposed edges of 8 bp in the major grooves (Fig. 5). These are arginine-180, glutamate-181, arginine-185, and lysine-188 from the F α-helix of each subunit. The position of glutamate-181, which makes hydrogen bond interactions with two adjacent base pairs, may be stabilized by salt bridges to arginine-180 or lysine-188. Furthermore, 9 hydrogen bonds or salt links can be made to the phosphates of the DNA backbone. Since the DNA to which CAP binds may not have precisely the regular structure that has been assumed and the protein conformation could also change, there may be more interactions in the complex than we have described. CAP binds most tightly to the *lac* operon; 0.5, 1.5, and 2.4 kcal/mol less tightly to the *Mal T*, *gal*, and *lac* 2 operons; and 2.7 kcal/mol less tightly to the L8 mutant in the *lac* operon (32). Models with these DNAs appear to have one fewer hydrogen bond in the case of *Mal T* and two fewer hydrogen bonds for the others. Model building predicts that CAP may bind in a similar manner to other known sites.

DISCUSSION

Guided by the location of the positive electrostatic charge potential in CAP, we have oriented DNA on CAP and constructed a detailed model of a complex between CAP and the

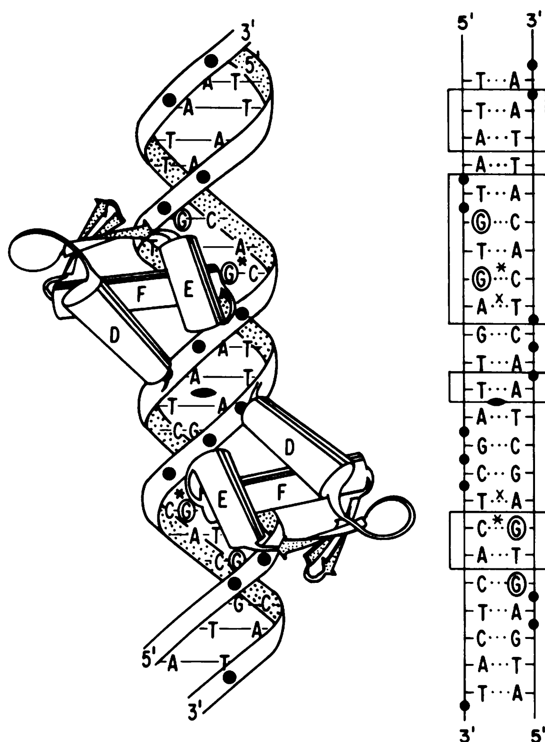

FIG. 2. A schematic drawing of the two CAP DNA binding domains interacting with the CAP binding site in the *lac* operon. On the right is a summary of results from chemical (29, 30) and enzymatic (31) protection by CAP binding in the *lac* operon. These data suggest that CAP is interacting with 18–24 bp. Dots mark phosphates in which ethylation prevents CAP binding; circled Gs are protected from methylation when CAP binds; * indicates the *lac* L8 and L29 mutations that decrease CAP affinity. Boxed sequences are >75% conserved in eight sequenced CAP sites (18), and the × indicates a T that can be crosslinked to CAP (29).

DNA sequence to which it binds in the *lac* operon (Figs. 2–5). This complex maximally overlaps the positive electrostatic potential of CAP with the negative potential of DNA and is calculated to have an electrostatic contribution to its stability of approximately −11 kcal/mol. Various experimental data on the expected nature of the CAP–DNA complex appear to be in broad agreement with this model (Fig. 2).

DNA sequences of CAP binding sites in various operons suggest that CAP is recognizing a sequence spanning 17 to 19 bp and making some interactions over a region spanning as much as 20 bp. The chemical protection data (29) are consistent with the model of CAP binding to bent DNA as follows:

(*i*) All of the phosphates in which ethylation decreases binding (30) are either in contact with the protein or within 5 Å of it (Fig. 2), except for two positions at the extreme end of the site.

(*ii*) Guanines in which methylation by dimethyl sulfate is prevented by CAP binding (29, 30) are hydrogen-bonded to arginine-180 and lysine-188.

(*iii*) The thymine that can be crosslinked to CAP by UV irradiation (29) is in contact with threonine-182 and is 8 Å from the thiol of cysteine-178.

(*iv*) The G-C base pair that is changed to A-T in the *lac* L8 and L29 mutations (33) is interacting with lysine-188 and glutamate-181 in our model. While this manuscript was in preparation, R. Ebright, J. Beckwith, P. Cossart, and B. Gicquel-Sanzey (personal communication) selected a CAP revertant to the L8 or L29 mutations. They found that a change from glutamate-181 to leucine, valine, or lysine increased CAP affinity for *lac* L8 or L29 DNA. These results (obtained after the construction of our model) provide support for our specific model. Ebright *et al.* have proposed a similar model based on different criteria. Since glutamate-181 does not interact with DNA in the model complex with left-handed DNA, these genetic data also support the conclusion that CAP does not bind to left-handed DNA.

Comparison of the sequences of the numerous sites to which CAP binds specifically suggests that CAP shows preference for a sequence 5′ A-A-N-T-G-T-G-A-N-N-T 3′ on one side of the 2-fold axis in the site (18) (Fig. 2). In our model of the complex, specific interactions between protein side chains can only be made with the exposed edges of 4 bp, 5′ G-T-G-A 3′ (Figs. 4 and 5). The apparent preference of CAP for the preceding A-A-N-T sequence may arise from CAP recognizing a sequence-dependent variation in the sugar–phosphate backbone structure. Bending the DNA in our model of a CAP complex increases both the favorable electrostatic free energy of interaction and the number of contacts between CAP and DNA. As pointed out earlier (18), straight DNA can only interact with CAP over a 14-bp

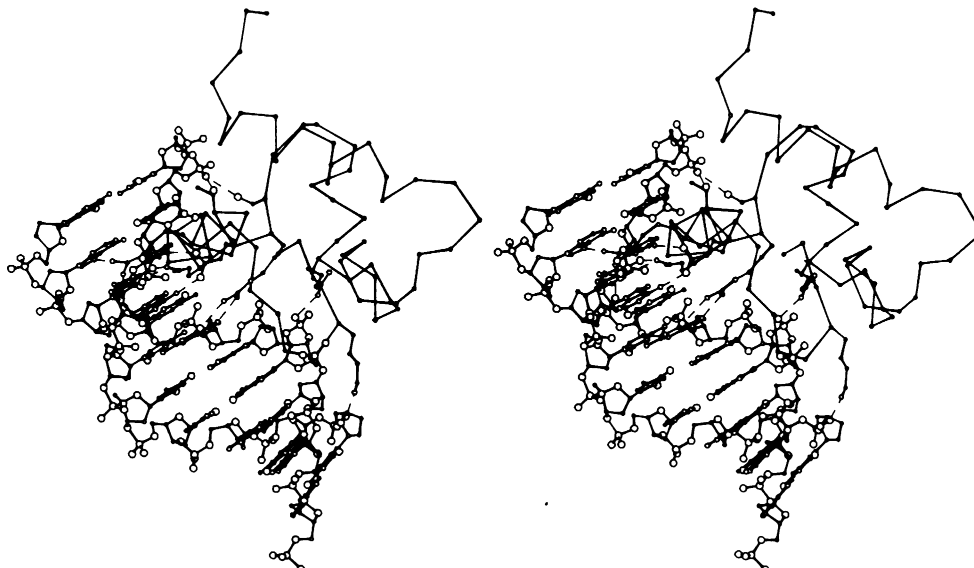

FIG. 3. A stereo drawing of one COOH-terminal domain of CAP interacting with one-half of the DNA binding site. Only those protein side chains that are proposed to interact with DNA are shown, together with the α-carbon backbone.

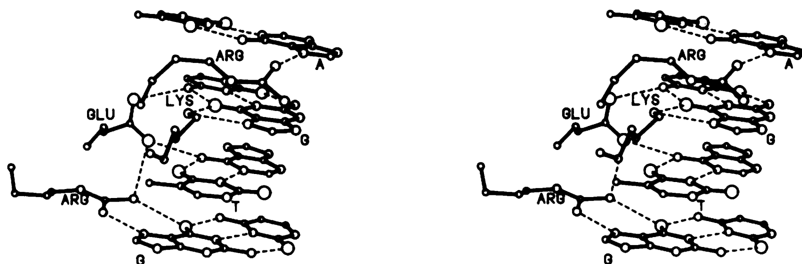

FIG. 4. A stereo drawing of a portion of the CAP–DNA complex showing four amino acid side chains from CAP hydrogen-bonded to the exposed edges of 4 bp in the major groove.

stretch. Experimental evidence suggesting bending has been obtained by Wu and Crothers (34), who conclude from the anomalously low gel mobility of a CAP complex with a 203-bp *lac* operator fragment that CAP bends DNA by some unknown amount.

Since the two DNA binding domains of the CAP dimer are not related by a precise 2-fold axis in the crystal (6), the interactions between CAP and DNA are also not precisely 2-fold symmetric (Fig. 5). In our current model, some of the interactions made by the two subunits with sugar–phosphate backbone differ as shown by comparison of Fig. 5 *a* and *b*. Several observations suggest that the CAP structure in the CAP–DNA complex may be asymmetrical (35, 36). However, we cannot exclude the possibility that the asymmetric structure of CAP in the crystal is a result of the crystal environment and not a property of the dimer in solution.

Comparison with Models of cro and λcI Complexes. There are several general conclusions that can be made about the mechanisms of DNA sequence recognition by regulatory proteins from the structures and proposed DNA complexes of λ phage cro repressor (17), *c*I repressor fragment (19), and CAP. It appears that much of the specific sequence recognition is achieved by a two α-helix structural motif that is common to these three proteins (10, 11) and probably to many sequence-specific DNA binding proteins (12–14). The second helix of the two-helix structure (*i*) protrudes from the surface of the protein, (*ii*) is separated by 34 Å across a molecular diad axis from a dimer related mate, and (*iii*) is proposed to interact in the successive major grooves of B-DNA. Some of the specific DNA sequence recognition is achieved by hydrogen bonds formed between side chains from the protruding helix and the edges of base pairs exposed in the major groove. Furthermore, the same residue positions in the two-helix motif appear to be making the DNA interactions in all three cases (Fig. 6). Additional specific interactions, with both the bases and the backbone, are likely to be made by either the extreme NH_2 or extreme COOH terminus of cro and λ*c*I (17, 37) and may also occur with CAP (6).

There are, however, differences among the three known regulatory protein structures and their proposed complexes with DNA. Although each protein has the two-helix tertiary structure motif, the quaternary structure of these proteins differs in a manner that makes the tilt of the two dimer-related protruding helices (F in CAP and α_3 in cro and λ*c*I) rela-

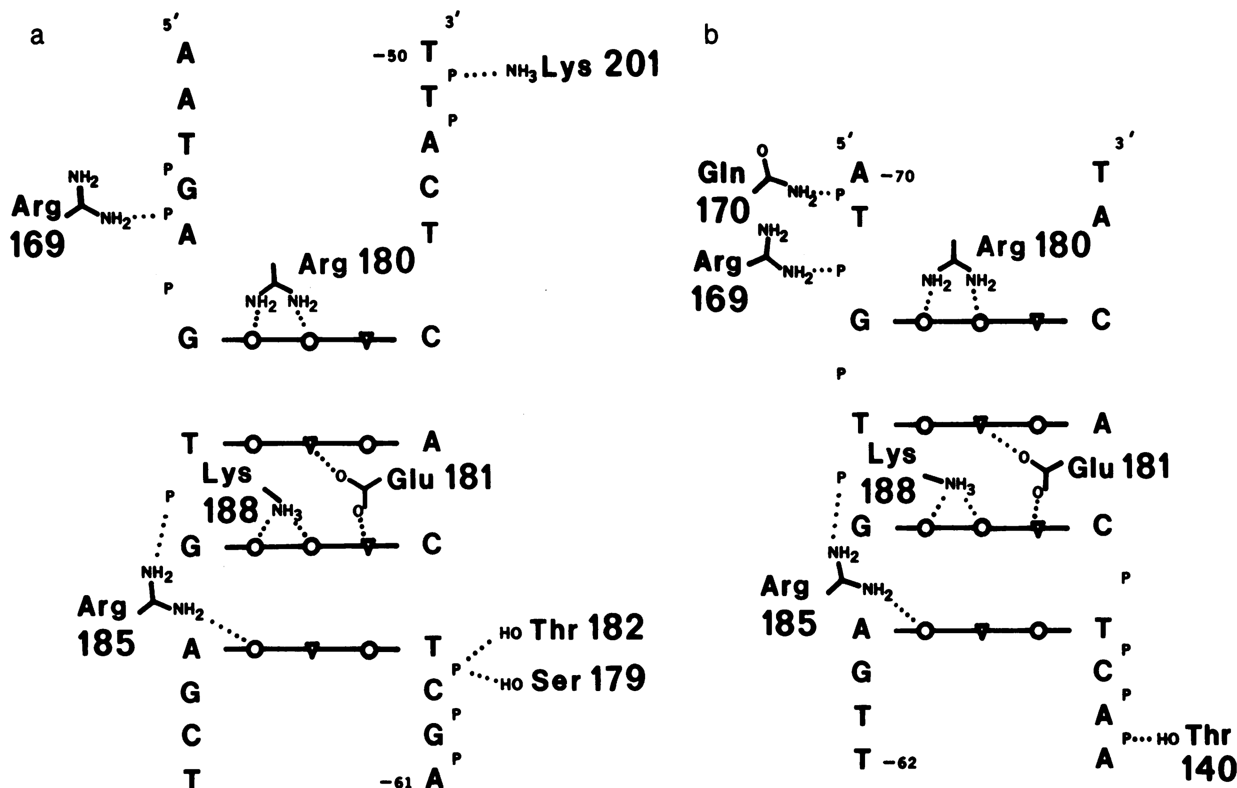

FIG. 5. A schematic diagram of some of the interactions proposed between the two small domains of a CAP dimer and its DNA binding site in the *lac* operon. Interactions made by each subunit with one-half of the DNA site are shown separately in *a* and *b*. Hydrogen bond donors on the bases are indicated by ▽ and acceptors are shown by ○. Phosphates within 5 Å of protein atoms are indicated by P.

```
            α HELIX E              α HELIX F

           * *                 * * * *     *     *
CAP    I-T-R-Q-E-I-G-Q-I-V-G-C-S-R-E-T-V-G-R-I-L-K-
           * *       *         * * *     * *     *
CRO    F-G-Q-T-K-T-A-K-D-L-G-V-Y-Q-S-A-I-N-K-A-I-H
           *                     * * *     *     *
λcI    L-S-Q-E-S-V-A-D-K-M-G-M-G-Q-S-G-V-G-A-L-P-N-
```

FIG. 6. Comparison of the homologous amino acid sequences of CAP, cro, and λ*c*I in the region of the CAP E and F helices. Residues proposed to interact with DNA in each case (17, 19) are shown by an asterisk. Amino acids are represented by the standard one-letter abbreviations.

tive to the line connecting their centers different, which affects the ways that it is possible to fit the protruding α-helices into the major groove. In the cro-DNA model the α_3-helix axis is nearly parallel to the major groove, which is inclined at about 32° to the plane of the bases, whereas in both the λ*c*I and CAP models with DNA the F helix is nearly parallel to the bases and therefore does not extend as far into the groove, so that fewer interactions are possible. It appears that, although the two-helix DNA binding motif has been conserved, the exact way in which it interacts with DNA has not.

A second difference between cro and CAP is the location of the positive electrostatic potential. The positive electrostatic charge potential of cro (38) lies entirely between the two α_3-helices in cro, whereas it lies outside the F helices in CAP. This may suggest that the extent and manner in which CAP is bending or kinking the DNA is both greater and different from the effect of cro on DNA.

Although the overall orientation of DNA on CAP and many of the specific interactions have probably been correctly identified, the extent and nature of protein-induced changes in the DNA structure and vice versa cannot be guessed by model building.

We thank J. Matthew and F. M. Richards for the computer program and information on electrostatic calculations on DNA prior to publication. This work was supported by National Science Foundation Grant PCM-81-10880 and U.S. Public Health Service Grant GM-22778.

1. Zubay, G., Schwartz, D. & Beckwith, D. (1970) *Proc. Natl. Acad. Sci. USA* **66**, 104–110.
2. Epstein, W., Rothman-Denes, L. B. & Hesse, J. (1975) *Proc. Natl. Acad. Sci. USA* **72**, 2300–2304.
3. de Crombrugghe, B. & Pastan, I. (1978) in *The Operon*, eds. Miller, J. H. & Reznikoff, W. S. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY), pp. 303–324.
4. Movva, R. N., Green, P., Nakamura, K. & Inouye, M. (1981) *FEBS Lett.* **128**, 186–190.
5. Aiba, M. (1983) *Cell* **32**, 141–149.
6. McKay, D. B. & Steitz, T. A. (1981) *Nature (London)* **290**, 744–749.
7. McKay, D. B., Weber, I. T. & Steitz, T. A. (1982) *J. Biol. Chem.* **257**, 9418–9524.
8. Anderson, W. F., Ohlendorf, D. H., Takeda, Y. & Matthews, B. W. (1981) *Nature (London)* **290**, 754–758.
9. Pabo, C. O. & Lewis, M. (1982) *Nature (London)* **298**, 443–447.
10. Steitz, T. A., Ohlendorf, D. H., McKay, D. B., Anderson, W. F. & Matthews, B. W. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 3097–3100.
11. Ohlendorf, D. H., Anderson, W. F., Lewis, M., Pabo, C. O. & Matthews, B. W. (1983) *J. Mol. Biol.* **169**, 757–769.
12. Anderson, W. F., Takeda, Y., Ohlendorf, D. H. & Matthews, B. W. (1982) *J. Mol. Biol.* **159**, 745–755.
13. Sauer, R., Yocum, R., Doolittle, R., Lewis, M. & Pabo, C. (1982) *Nature (London)* **298**, 447–451.
14. Weber, I. T., McKay, D. B. & Steitz, T. A. (1982) *Nucleic Acids Res.* **10**, 5085–5102.
15. Hecht, M., Nelson, H. & Sauer, R. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 2676–2680.
16. Hochschild, H., Irwin, N. & Ptashne, M. (1983) *Cell* **32**, 319–325.
17. Ohlendorf, D. H., Anderson, W. F., Fisher, R. G., Takeda, Y. & Matthew, B. W. (1982) *Nature (London)* **298**, 718–723.
18. Steitz, T. A., Weber, I. T. & Matthew, J. B. (1983) *Cold Spring Harbor Symp. Quant. Biol.* **47**, 419–426.
19. Lewis, M., Jeffrey, A., Wang, J., Ladner, R., Ptashne, M. & Pabo, C. (1983) *Cold Spring Harbor Symp. Quant. Biol.* **47**, 435–440.
20. Kolb, A. & Buc, H. (1982) *Nucleic Acids Res.* **10**, 473–485.
21. Shire, S. J., Hanania, G. & Gurd, F. (1974) *Biochemistry* **13**, 2967–2974.
22. Matthew, J. B. & Richards, F. M. (1982) *Biochemistry* **21**, 4989–4999.
23. Matthew, J. B., Weber, P. C., Salemme, F. R. & Richards, F. M. (1983) *Nature (London)* **301**, 169–171.
24. Steitz, T. A., Weber, I. T., Ollis, D. & Brick, P. (1983) *J. Biomol. Struct. Dyn.* **1**, 1023–1037.
25. Arnott, S. & Hukins, D. W. L. (1972) *Biochem. Biophys. Res. Commun.* **47**, 1504–1509.
26. Tanford, C. & Kirkwood, J. G. (1957) *J. Am. Chem. Soc.* **79**, 5333–5339.
27. Matthew, J., Hanania, G. & Gurd, F. (1979) *Biochemistry* **18**, 1919–1928.
28. Matthew, J. B. & Richards, F. M. (1984) *Biopolymers*, in press.
29. Simpson, R. B. (1980) *Nucleic Acids Res.* **8**, 759–766.
30. Majors, J. (1977) Dissertation (Harvard Univ., Cambridge, MA).
31. Schmitz, A. (1981) *Nucleic Acids Res.* **9**, 272–292.
32. Kolb, A., Spassky, A., Chapon, C., Blazy, B. & Buc, H. (1983) *Nucleic Acids Res.* **11**, 7833–7852.
33. Dickson, R. C., Abelson, J., Johnson, P., Reznikoff, W. S. & Barnes, W. M. (1977) *J. Mol. Biol.* **111**, 65–75.
34. Wu, H.-M. & Crothers, D. M. (1984) *Nature (London)* **308**, 509–512.
35. Takahashi, M., Blazy, B. & Baudras, A. (1980) *Biochemistry* **19**, 5124–5130.
36. Garner, M. M. & Revzin, A. (1982) *Biochemistry* **21**, 6032–6036.
37. Pabo, C. O., Krovatin, W., Jeffrey, A. & Sauer, R. T. (1982) *Nature (London)* **298**, 441–443.
38. Ohlendorf, D. H., Anderson, W. F., Takeda, Y. & Matthews, B. W. (1983) *J. Biomol. Struct. Dyn.* **1**, 553–563.