Received 8 July 1980; accepted 9 February 1981.


1. Will, C. M. in *General Relativity* (eds Hawking, S. W. & Israel, W.) (Cambridge University Press, New York, 1979).
2. Dirac, P. A. M. *Nature* **139,** 323 (1937)
3. Canuto, V. M., Adams, P. J., Hsieh, S.-H. & Tsiang, E. *Astrophys. J. Suppl.* **41,** 243 (1979).
4. Canuto, V. M. & Hsieh, S.-H. *Astrophys. J.* **237,** 613 (1980).
5. Canuto, M. & Hsieh, S.-H. *Astrophys.* (in the press).
6. Canuto, V. M., Adams, P. J., Hsieh, S.-H. & Tsiang, E. *Phys. Rev. D* **16,** 1643 (1977) (Paper I); **41,** 263 (1979) (Paper III); **41,** 301 (1979).
7. Canuto, V. M., Hsieh, S.-H. & Owen, R. J. *Mon. Not. R. astr. Soc.* **188,** 829 (1979).
8. Teller, E. *Phys. Rev.* **73,** 801 (1948).
9. Jordan, P. *The Expanding Earth* (Pergamon, New York, 1971).
10. Carey, S. W. *The Expanding Earth* (Elsevier, Amsterdam, 1976).
11. Wesson, P. S. *Cosmology and Geophysics* (Adam Hilger, Bristol, 1978).
12. McElhinny, M. W., Taylor, S. R. & Stevenson, D. J. *Nature* **271,** 316 (1978).
13. Bishop, W. T. & Landsberg, P. T. *Nature* **264,** 346 (1976).
14. Landau, L. D. & Lifshitz, E. M. *Fluid Mechanics* (Addison-Wesley, Reading, 1959).
15. Dirac, P. A. M. *General Theory of Relativity*, 25–26; 43–45 (Wiley, New York, 1975).
16. Bergmann, P. *Introduction to the Theory of Relativity* (Prentice Hall, Englewood Cliffs, 1960).
17. Einstein, A. in *Autobiographical Notes* (ed. Schilpp, P. A.) 71, (Open Court, Chicago, 1979).
18. Canuto, V. M. & Hsieh, S.-H. *Phys. Rev. Lett.* **44,** 695 (1980).
19. Bekenstein, J. D. & Meisels, A. *Phys. Rev. D* **22,** 1313 (1980).
20. Landau, L. D. & Lifshitz, E. M. *Statistical Physics*, 224 (Pergamon, New York, 1958).
21. Birch, F. *J. geophys. Res.* **57,** 227 (1952).
22. Stacey, F. D., Brennan, B. J. & Irvine, R. D. *Geophys. Surv.* (in the press).
23. Morrison, L. V. & Ward, C. G. *Mon. Not. R. astr. Soc.* **173,** 183 (1975).
24. Goad, C. C. & Douglas, B. C. *J. geophys. Res.* **83,** 2306 (1978).
25. Lambeck, K. *Phil. Trans R. Soc.* **287,** 545 (1977).
26. Van Flandern, T. C. *Astrophys. J.* (in the press).
27. Dickey, J. O., Williams, J. G. & Yoder, C. F. *EOS* **61,** 46, 939 (1980).
28. Calame, D. & Mulholland, J. D. *Science* **199,** 977 (1978).
29. Nordtvedt, K. & Will, C. M. *Astrophys. J.* **177,** 755 (1972).
30. Runcorn, S. K. *Nature* **193,** 311 (1962); **195,** 1248 (1962); **204,** 823 (1964).
31. Lambeck, K. *The Earth's Variable Rotation: Geophysical Causes and Consequences* (Cambridge University Press, 1980).
32. Scrutton, C. T. in *Tidal Friction and the Earth's Rotation* (ed. Broshe, P.) (Springer, New York, 1978).
33. Hipkin, R. G. in *Growth Rhythms and the History of the Earth's Rotation* (eds. Rosenberg, G. D. & Runcorn, S. K.) (Wiley, New York, 1975).
34. Schmidt, P. W. & Clark, D. A. *Geophys. J. R. astr. Soc.* **61,** 95 (1980).


# Structure of catabolite gene activator protein at 2.9 Å resolution suggests binding to left-handed B-DNA

**David B. McKay & Thomas A. Steitz**

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

*The 2.9 Å resolution crystal structure of* Escherichia coli *catabolite gene activator protein (CAP) complexed with cyclic AMP reveals two distinct structural domains separated by a cleft. The smaller carboxy-terminal domain is presumed to bind DNA while the amino-terminal domain is seen to bind cyclic AMP. Model building studies suggest that CAP binds to left-handed B-type DNA, contacting its major groove via two α-helices. It is possible that the CAP conversion of right- to left-handed DNA in a closed supercoil, is what activates transcription by RNA polymerase.*

REPRESSOR and activator proteins that regulate gene expression at the level of transcription recognize specific nucleotide sequences in double-stranded DNA. Although it has been suggested that specific recognition involves α-helices fitting into the major groove of B-DNA[1–3] or anti-parallel β-strands in the minor groove[4], no experimental evidence exists concerning the actual mode of sequence-specific interaction between protein and double-stranded DNA. Also, the molecular mechanism by which gene activators promote the activity of RNA polymerase, thereby switching on genes, remains unknown. We present here the structure determination of a protein that binds in a sequence-specific manner to double-stranded DNA and propose a mechanism by which the catabolite gene activator protein is able to switch on the catabolite-sensitive genes.

The catabolite gene activator protein (CAP), also called the cyclic AMP receptor protein, functions in *Escherichia coli* in the regulation of several catabolite-sensitive gene operons[5,6]. Regulation by CAP is exerted at the transcriptional level with cyclic AMP acting as an allosteric effector. In the presence of a sufficient concentration of intracellular cyclic AMP, cyclic AMP forms a complex with CAP which binds to specific DNA sites near the promoters of several operons and alters the rate of their transcription by RNA polymerase. In the lactose (*lac*)[7] and arabinose[8] operons, the cyclic AMP–CAP complex is a positive regulator—it potentiates transcription. In the galactose operon, the presence of two overlapping promoters for RNA polymerase makes the situation more complex; cyclic AMP–CAP is required for initiation of transcription from one promoter, but inhibits transcription from the other. Thus, in this case it apparently acts as both a positive and negative regulator of operon expression[9].

The active form of CAP is a dimer of identical subunits of molecular weight 22,500 and 201 amino acid residues[10,11]. Proteolytic cleavage studies are consistent with the CAP subunit having two separate structural domains[12]. Results from a variety of techniques[12–17] suggest that CAP undergoes a significant conformational change on binding cyclic AMP.

To develop a structural basis for understanding (1) the cyclic AMP-induced allosteric transition in CAP, (2) the site-specific recognition of DNA by CAP, and (3) the mechanism of transcription activation by CAP, we have initiated crystallographic studies of CAP and its complexes with ligands. In this article we report the structure of the cyclic AMP–CAP complex at 2.9 Å resolution.

## Structure determination

CAP protein was purified and crystallized in the presence of 0.5 mM cyclic AMP as previously described[18]. The crystals are orthorhombic, space group $P2_12_12_1$, $a = 46.5$ Å, $b = 97.1$ Å, $c = 105.4$ Å, with one dimeric CAP molecule per asymmetric unit.

Intensities of crystallographic reflections were measured by diffractometer, using the Wyckoff scan algorithm[19]. Friedel pairs of reflections were measured on native crystals and five heavy-atom derivatives; for derivative data sets only ~75% of the total reflections, consisting of those which were most intense in the native data, were measured. Heavy-atom positions were located and refined by conventional methods[20] and the correct enantiomorph of the heavy-atom positions was determined using anomalous difference Fouriers[21]. A summary of heavy-atom derivative preparation and refinement statistics is presented in Table 1. Combined multiple isomorphous replacement and anomalous scattering phases were computed to 2.9 Å resolution and had an average figure of merit of 0.74.

Part of the electron density map is displayed in Fig. 1. Most of the polypeptide backbone is well ordered and can be traced unambiguously in both subunits; however, regions of local disorder are encountered at the amino termini, the carboxy

0028–0836/81/180744—06$01.00

**Table 1** The heavy-atom derivative preparation and refinement statistics

| | No. of heavy-atom sites | Minimum resolution (Å) | $R_{sym}$ | | Resolution limits (Å) 40.00 | 23.20 | 11.60 | 7.73 | 5.80 | 4.64 | 3.87 | 3.31 | 2.90 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Native | | 2.9 | 0.048 | | | | | | | | | | | |
| MeHg(I) | 5 | 2.9 | 0.063 | $\frac{f_{r.m.s.}}{E_{r.m.s.}}$ | 2.92 | 3.80 | 3.61 | 3.68 | 2.72 | 2.62 | 2.57 | 3.08 | | |
| | | | | $R_c$ | 0.34 | 0.36 | 0.37 | 0.35 | 0.41 | 0.53 | 0.53 | 0.46 | | |
| MeHg(II) | 3 | 2.9 | 0.044 | $\frac{f_{r.m.s.}}{E_{r.m.s.}}$ | 1.25 | 2.58 | 2.65 | 2.92 | 2.86 | 2.16 | 2.18 | 2.09 | | |
| | | | | $R_c$ | 0.54 | 0.40 | 0.44 | 0.36 | 0.40 | 0.62 | 0.64 | 0.60 | | |
| $K_2ReCl_6$(I) | 2 | 3.5 | 0.031 | $\frac{f_{r.m.s.}}{E_{r.m.s.}}$ | 1.75 | 1.22 | 1.29 | 1.70 | 1.15 | 0.82 | 0.84 | | | |
| | | | | $R_c$ | 0.37 | 0.59 | 0.66 | 0.63 | 0.71 | 0.79 | 0.98 | | | |
| $K_2ReCl_6$(II) | 3 | 3.5 | 0.032 | $\frac{f_{r.m.s.}}{E_{r.m.s.}}$ | 1.31 | 1.21 | 1.21 | 1.38 | 1.22 | 1.02 | 0.97 | | | |
| | | | | $R_c$ | 0.76 | 0.57 | 0.69 | 0.64 | 0.72 | 0.90 | 0.83 | | | |
| DMM | 5 | 5.0 | 0.041 | $\frac{f_{r.m.s.}}{E_{r.m.s.}}$ | 1.33 | 1.78 | 1.82 | 2.28 | 2.22 | | | | | |
| | | | | $R_c$ | 0.57 | 0.59 | 0.45 | 0.50 | 0.48 | | | | | |
| | | | | $\langle m \rangle$ | 0.95 | 0.93 | 0.90 | 0.91 | 0.83 | 0.75 | 0.70 | 0.65 | Overall 0.74 | |
| | | | | No. of reflections | 17 | 150 | 386 | 634 | 999 | 1383 | 2011 | 2578 | 8158 | |

Derivative soaking conditions: MeHg(I): 1 mM methylmercuri-$\beta$-mercaptoethanol, 50 mM potassium phosphate, 0.5 mM cyclic AMP, *p*H 8.0, 1 day. MeHg(II): 0.1 mM methylmercuri-$\beta$-mercaptoethanol, 50 mM potassium phosphate, 0.5 mM cyclic AMP, *p*H 8.0, 1 day. $K_2ReCl_6$(I): 1 mM $K_2ReCl_6$, 50 mM potassium phosphate, 0.5 mM cyclic AMP, 0.1 mM dithiothreitol, *p*H 8.0, 1–2 days. $K_2ReCl_6$(II): 2 mM $K_2ReCl_6$, 50 mM 2[*N*-morpholino]ethane sulphonic acid, 50 mM KCl, 0.5 mM cyclic AMP, 0.1 mM dithiothreitol, *p*H 6.4, 2 days. DMM: a drop of dimethylmercury was placed in capillary with mounted crystal and allowed to equilibrate for 3 days. $\langle m \rangle \equiv$ mean figure of merit.

$$R_{sym} \equiv \frac{\Sigma|I_i - \bar{I}|}{\Sigma\bar{I}} \qquad f_{r.m.s.} \equiv \left\{\frac{\Sigma f_H^2}{n}\right\}^{1/2} \qquad E_{r.m.s.} \equiv \left\{\frac{\Sigma(F_{PH} - |\mathbf{F}_P + \mathbf{f}_H|)^2}{n}\right\}^{1/2} \qquad R_c \equiv \frac{\Sigma||F_{PH} - F_P| - f_H|}{\Sigma|F_{PH} - F_P|}$$

where $I$ = diffraction intensity, $f_H$ = calculated heavy-atom structure factor amplitude, $F_P$ = structure factor amplitude of native crystals, and $F_{PH}$ = structure factor amplitude of derivative crystals. Summations are overall reflections, $n$.

termini and a coiled structure in the smaller domain marked A and A'.

Approximate $\alpha$-carbon positions were obtained by placing markers on map sections and measuring their coordinates. Polarity of the polypeptide chain was established from the tilt of the larger side chains on the $\alpha$-helices. Although many side chains are clearly visible in this map, no attempt has yet been made to place the several fragments of amino acid sequence obtained by Schlesinger. Cyclic AMP was identified because it was not continuously connected with the polypeptide backbone, included the most dense feature in the map (presumably the phosphate) and was too large to be an amino acid side chain.

## Subunit morphology

The CAP subunit is folded into two distinct structural domains (Fig. 2). The larger amino-terminal domain has approximate overall dimensions of $25 \times 30 \times 35$ Å. It contains ~135 amino acid residues and consists primarily of a pair of helices (A and B, Fig. 2) which lead into an antiparallel $\beta$-strand folded into an eight-strand $\beta$-roll; also, a 40 Å-long $\alpha$-helix C connects the two domains and partially closes off one end of the interior pocket formed by the $\beta$-roll. This helix extends from the small domain to the distal extremity of the large domain rather than forming a short, proximal bridge between the two domains. The topology of the eight-strand $\beta$-roll, sometimes referred to as the 'Swiss roll', has recently been observed, with minor variations, in three protein structures: the 'shell' domain of tomato bushy stunt virus protein[22], Southern bean mosaic virus protein[23] and influenza haemagglutinin[46].

The carboxy-terminal domain is smaller, with approximate dimensions $20 \times 20 \times 30$ Å. It has about 65 amino acid residues and contains three distinct $\alpha$-helices, D, E and F; the D helix is ~24 Å long, the F helix ~22 Å long, whereas the E helix is rather short. Much of the small domain consists of irregular structure, some of which is partially disordered. Although the electron density marked AA' in Fig. 1 and corresponding to the region circled in Fig. 2 is well above the background level of the solvent region, it is not sufficiently well defined to place $\alpha$-carbon positions with complete assurance. Thus, the polypeptide path shown in Fig. 2 at the carboxy terminus of the D and F helices illustrates only the general folding of the subunit in these regions and the connections within the circled region may have to be altered when amino acid sequence data are fitted to the electron density map. As the number of amino acid residues in our model is very nearly that expected from amino analysis[10,11], no significant portion of the molecule can be missing from the model.

Although there is a distinct cleft between the two domains of the CAP subunit, noncovalent contacts occur between them that presumably stabilize the relative orientation of the domains. Both subunits contain one similar region of noncovalent contact between their large and small domains: the end loop connecting strands 4 and 5 of the $\beta$-roll in the large domain is in contact with a loop of coil in the small domain. In addition, in one of the subunits, there is a second point at which the carboxyl end of the D helix may be within noncovalent contact distance of strand 5 of the $\beta$-roll of the large domain.

## Dimer morphology

As the structures of the two subunits of the CAP dimer differ, there is no unique dyad axis relating them (Figs 3–5). Qualitatively, the cleft of one subunit appears 'closed' to the point of extensive van der Waal's contact between the large and small domains while the cleft of the second subunit is 'open'; that is, the two domains have different relative orientations in the two subunits. Quantitatively, if the two large domains are super-

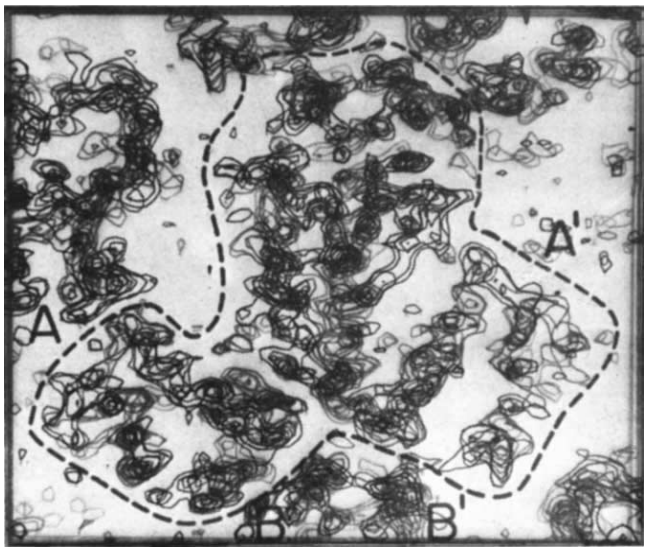


**Fig. 1** Part of the electron density map, 81 Å horizontal by 71 Å vertical by 10 Å depth. Only positive contours are shown; the contour interval is approximately one-sixth the maximum peak height. The dashed line circumscribes one CAP molecule. The two oblong lobes near the bottom of the map are the small domains of the molecule; the two helices running approximately vertically near the centre of the map are the C helices. A, A', coiled structure in the smaller domain; B, B', positions of tight intermolecular contact between small domains and a large domain of a neighbouring molecule.

imposed by least-squares minimization of differences in corresponding $\alpha$-carbon positions, a subsequent rotation of the small domain of one subunit by 28° is required to bring it into coincidence with the small domain of the second subunit. The 'hinge' region of the molecule appears to consist of approximately four to five amino acids in the region connecting the C helix to the small domain. This kind of conformational difference between chemically identical subunits has been widely observed, such as in yeast hexokinase[24], tomato bushy stunt virus coat protein[22], immunoglobulins[25] and alcohol dehydrogenase[26].

Interactions between the two subunits are through direct contact along virtually the entire length of the two C helices, and contact of the $\beta$-roll of one subunit with the central region of the C helix of the other. No direct contact between the two small domains is apparent.

Although the entire dimer does not possess an overall 2-fold axis, the two pairs of domains are each related by separate rotation axes. These two rotation axes have been obtained by computing the rotation required for least-squares minimization of the difference in corresponding $\alpha$-carbon positions of each domain. The results are summarized in Fig. 3. The large domains appear to be related by 2-fold symmetry (computed rotation $179.2 \pm 1.1°$); the small domains may have a small distortion from dyad symmetry (computed rotation $186.7 \pm 3.7°$); the angle between the rotation axis relating the large domains and that relating the small domains is 13°. The extensive contact between large domains suggests that these two domains maintain the same relative orientation in the presence and absence of various ligands. The observed asymmetry is, therefore, probably generated by an alteration in the orientation of a small domain relative to the large domains.

## Cyclic AMP binding

The electron density map contains density for two cyclic AMP molecules of equal and full occupancy per CAP dimer, as anticipated from data on cyclic AMP binding to CAP in solution[27]. The cyclic AMP is completely buried within the interior of the $\beta$-roll of the large domain. The peak of highest electron density in the cyclic AMP molecule and the entire map is presumed to be the phosphate. Further indication of the orientation of the nucleotide is provided by crystals soaked in a solution of 8-bromoadenosine-3', 5'-cyclic monophosphate (8-Br-cyclic AMP). The cyclic AMP seems to be oriented with the phosphate bound in the loop connecting strand 6 to strand 7 in the $\beta$-roll, with the nucleotide base in proximity of the two C helices. Whether the cyclic AMP makes direct contact with one or both of the C helices, and whether the two cyclic AMP molecules per dimer are bound identically cannot be determined conclusively at this stage. The finding[27] that cyclic AMP shows negative cooperativity in its binding to CAP in solution may explain the formation of the asymmetric dimer observed in the crystal.

## CAP domain structure and function

The crystal structure of CAP shows that the subunits consist of two distinct structural domains which may be generally true for allosterically controlled gene regulatory proteins. Limited proteolysis sudies of *E. coli lac* repressor[28–30], bacteriophage $\lambda$ repressor[31] and CAP[12,13] suggest that these proteins have a small DNA-binding domain and a larger domain that binds allosteric effectors and/or holds the subunits together. In all three cases it now seems that there is little, if any, direct interaction between the two DNA-binding domains.

Proteolytic cleavage of the *lac* repressor in appropriate conditions yields two separate protein fragments[29,30]: a tetrameric core that binds inducers, and a smaller monomeric domain having 51–59 amino acids, depending on the protease used. The small domain retains nonspecific[30] and operator-specific[32] DNA-binding activity but does not bind inducers. Similarly, proteolytic cleavage of the $\lambda$ repressor yields a dimeric core devoid of DNA-binding activity and monomeric domains which bind both nonspecific and $\lambda$ operator DNA[31].

It has been demonstrated that limited proteolysis of CAP yields an amino-terminal fragment of subunit MW 12,500, which still forms dimers and binds cyclic AMP but has lost its cyclic AMP-dependent DNA-binding activity[12]. Isolation of a second intact carboxy-terminal fragment and determination of its cyclic AMP- and DNA-binding activities have not been

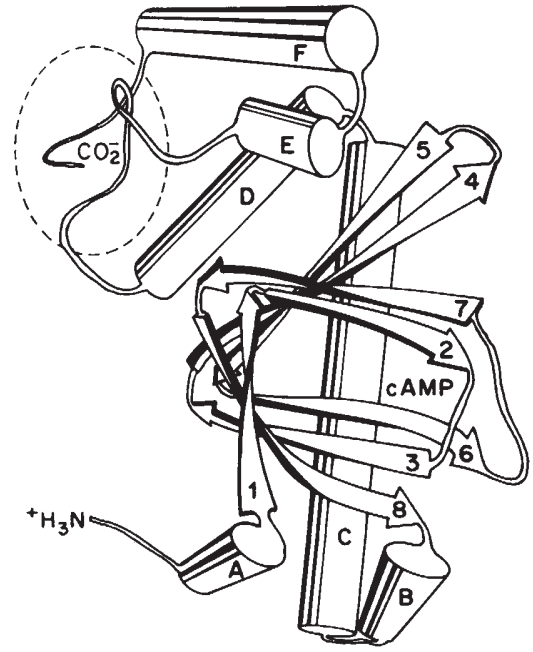


**Fig. 2** Schematic drawing of the CAP promoter in the closed conformation. The numbered arrows represent $\beta$-sheet; the lettered cylinders $\alpha$-helices. The binding site for cyclic AMP in the $\beta$-roll is labelled. The dashed line circumscribes the region which is not well ordered in the electron density map.

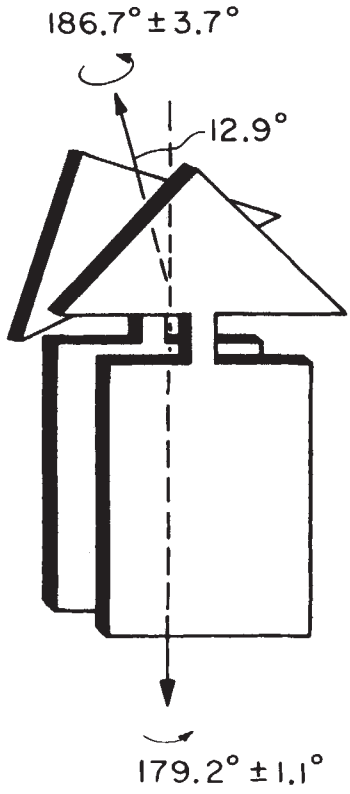


**Fig. 3** Schematic diagram of the molecular symmetry of the CAP dimer. The rectangles represent the large N-terminal domains, the triangles the small C-terminal domains.

reported. Nevertheless, by analogy with the *lac* and λ repressors one might presume that the small domain of CAP binds to DNA while the larger domain binds the cyclic AMP and forms dimers.

## Interaction of CAP with DNA

It has not proved possible to observe DNA binding to these crystals. Crystals soaked in solutions containing short (6–10 base pairs) oligodeoxyribonucleotides become substantially disordered. Because the small domains form tight intermolecular contacts with a large domain of a neighbouring molecule (regions B and B′, Fig. 1), the most plausible DNA-binding surface of the small domains is inaccessible to oligonucleotides. Thus, direct study of DNA binding in this crystal form is probably impossible. To determine experimentally the structure of CAP bound to DNA, it will be necessary to co-crystallize CAP with an appropriate oligo-DNA.

We have turned to model building to identify the DNA-binding site on CAP. Although we have not yet incorporated side chains into our CAP model, three general characteristics should be exhibited by the DNA binding site. (1) We expect a large degree of general structural complementarity between DNA and the potential binding site. (2) The correct CAP–DNA model must provide a structural basis for explaining the chemical protection[33,34] and 'footprinting' (A. Schmitz, personal communication) data obtained on the CAP site of the *lac* operon. (3) Because the sequence of the CAP binding site at the *lac* promoter has a high degree of 2-fold symmetry, CAP probably binds to DNA with the approximate 2-fold axis that relates the two small domains coincident with the approximate 2-fold axis in the CAP site. Our conclusion is that the only model of interaction between a CAP dimer and DNA that adequately meets these three criteria requires that CAP binds to a left-handed, B-type DNA.

We have examined several possible modes of binding, and find only two ways in which a DNA molecule could interact with both small domains of a single CAP dimer with the approximate 2-fold axes of DNA and CAP coincident. Both ways involve the 22 Å-long F helices that are a dominant structural feature of the small domains, and are related by the approximate 2-fold axis between the small domains, are nearly parallel to each other and have their helix axes separated by about 34 Å. Furthermore, these helices are inclined at ~65–70° to the line connecting their centres.

The most complementary mode of interaction between the DNA and CAP is suggested by the 34 Å spacing and tilt of the 2-fold-related F α-helices. These two α-helices can fit snugly into two successive major grooves of left-handed B-DNA[35] (Fig. 4). In addition, the extended polypeptide chain is correctly positioned to interact with the sugar–phosphate backbone on either side of the two α-helices (Figs 4, 5*c*). Placing the α-helices in the major groove accounts nicely for the L8 and L29 mutations[36], methylation protection at the $N^7$ of two pairs of guanines and the UV light-induced cross-linking of two symmetry-related thymines[33,34], because all these residues then interact with these helices (Fig. 5*c*). Furthermore, the phosphate moieties implicated by chemical modification to be interacting with CAP are almost all near the polypeptide backbone (Fig. 5*c*). In this way the protein can interact with at least 18–22 base pairs of DNA in a manner consistent with the footprinting data (refs 8, 34, 37 and A. Schmitz, personal communication) which show CAP protection spanning 25 base pairs.

Left-handed B-DNA has recently been shown to be a stereochemically acceptable conformation[35]. Note that CAP cannot interact with left-handed Z-DNA[38,39], because the pitch is too large (45 Å) and there are no major grooves for the helices to fill. Furthermore, a purine–pyrimidine alternating sequence, requisite for formation of the Z-DNA helix, is not present in any known CAP recognition sites. Nevertheless, the known crystal

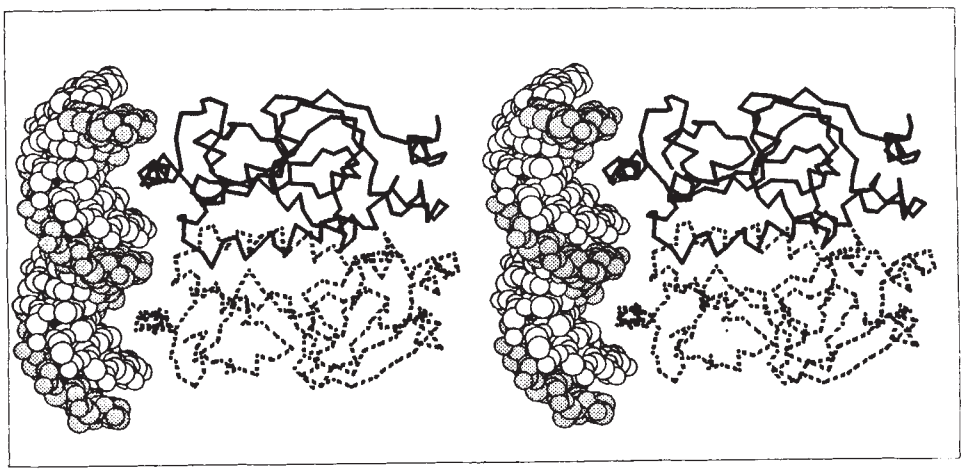

**Fig. 4** Stereo drawing of the α-carbon backbone of the CAP dimer interacting with 22 base pairs of left-handed B-type DNA. One CAP subunit is drawn with dashed lines, the other subunit with solid lines. The coordinates for the DNA were derived from parameters given in ref. 37 and are presented in a space-filling representation using standard van der Waals radii. Note the extraordinary degree of general structural complementarity which extends over 18–20 base pairs (65–70 Å) of DNA.

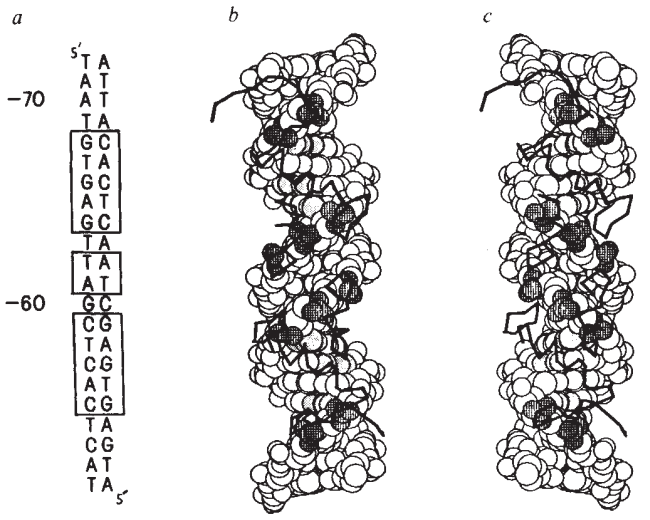


**Fig. 5** *a*, Sequence of the CAP-binding region of the *lac* operon. Boxes enclose regions of 2-fold symmetry in the sequence. *b*, Space-filling model of right-handed B-DNA, with proximal regions of the protein $\alpha$-carbon backbone superimposed. Contact sites predicted from chemical modification studies[33,34] are shaded. *c*, The same for left-handed B-DNA.

structure of left-handed Z-DNA and the stereochemical possibility of left-handed B-DNA suggest that the energy difference between left- and right-handed DNA may not be large and that stabilization of a left-handed DNA by a protein might be possible.

In contrast, a complementary interaction between CAP and right-handed DNA is not possible (Fig. 6). In this case the F $\alpha$-helices are skewed relative to the major groove and make contact only with a portion of the sugar–phosphate backbone. No protein can project into either groove to make specific contacts. This model does not account for the methylation protection, mutations or cross-linking observed with bases in the major groove.

Structural rearrangements of CAP which would allow it to bind with its F helices in the major groove of right-handed B-DNA are highly implausible. Rotation of the F helices by 60–70°, while constraining their centre-to-centre distance at 34 Å, would be required. Rotation of the small domains about the hinge points changes the centre-to-centre spacing of the F helices and hence is excluded. Rotation of the individual helices by 60–70° about their centre points would maintain the 34 Å spacing, but would require a major refolding of the small domains. As the crystals are grown in the conditions in which CAP binds specifically to DNA in solution, we expect that the observed crystal structure is in the active DNA-binding conformation. By analogy with virtually all other known protein examples, there is no reason to expect that the structure of CAP in the crystals differs significantly from its structure in solution.

The only other model of DNA bound to CAP that we seriously considered was right-handed B-DNA placed in the shallow depression formed primarily by the two parallel F helices (model not shown). In this model the long axis of the DNA runs parallel to the F helices and makes minor contacts with the proximal ends of the D helices. Such a model seems less plausible because (1) the length of protein contact along the DNA helix would be only 35 Å or 10 base pairs, (2) contact of CAP along the DNA phosphate backbone would be along alternate sides of the helix, rather than along one face, and (3) there would be no structural complementarity between protein and either the major or minor groove of DNA.

Thus, we conclude that only left-handed DNA, bound to CAP as shown in Fig. 4, accommodates all the relevant data. At least two experimentally verifiable predictions are made by this model. First, the inclusion of amino acid side chains in the protein model should provide a DNA-binding site which is complementary to left-handed B-DNA in hydrogen bond donors and acceptors and charge interactions, as well as in overall shape as shown here. Second, specific binding of active CAP to appropriate CAP sites should reduce the winding number in closed circular DNA by about four, although nonspecific binding of either active or inactive CAP would probably not. Both experiments are in progress.

A possible mechanism by which cyclic AMP might exert its allosteric effect on specific CAP binding to DNA is suggested by the model of CAP–DNA (Fig. 4) and the two different orientations of the small domains observed in the two subunits. If the removal of cyclic AMP changes the orientation of the small domain of one subunit relative to its large domain, then the DNA-binding site would be destroyed. It would not be unreasonable for the two subunits to have the same structure (either both 'open' or both 'closed') in the absence of cyclic AMP and the dimer a 2-fold symmetry axis. A ligand-induced change in relative domain orientation has been demonstrated in hexokinase[25] and alcohol dehydrogenase[26].

## Formation of left-hand B-DNA could activate transcription initiation by RNA polymerase

We suggest that CAP activates transcription by destabilizing, or melting, a right-handed DNA helix within the promoter region. The significantly different positions of the known CAP-binding sites relative to the promoter in the *lac*[33,34], *gal*[37] and *ara*[8] operons rule out the possibility of a unique protein–protein interaction between CAP and RNA polymerase. Also, the observation that inhibition of DNA gyrase activity *in vivo* inhibits the expression of catabolite-sensitive genes[40], and the more recent *in vitro* observation that in the *lac* operon, CAP stimulation of polymerase activity is most effective when the

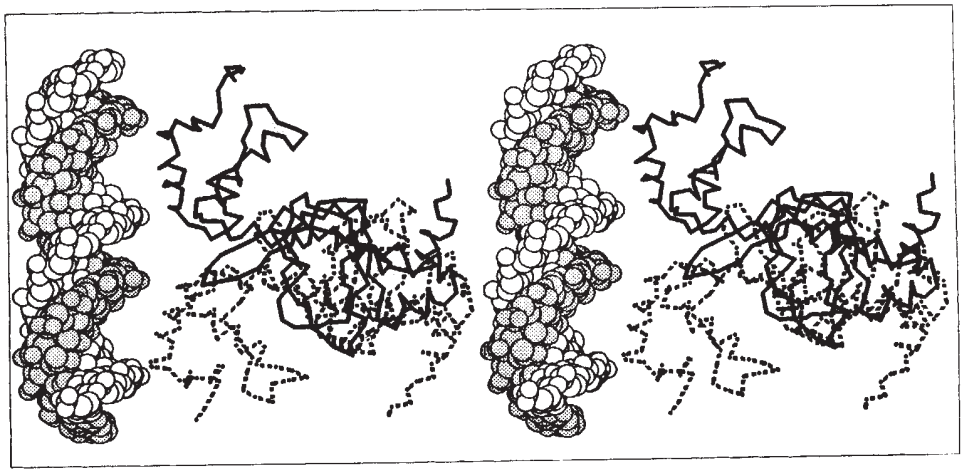

**Fig. 6** Stereo drawing of a model of CAP and right-handed B-DNA[45] in approximately the same orientation as in Fig. 4. In this figure the DNA has to be moved further from the protein to avoid overlap of the F helices and the sugar–phosphate backbone. The contact surface between these macromolecules is strikingly less than in Fig. 4.

promoter and CAP site are in negatively supercoiled rather than relaxed linear DNA (W. R. McClure, personal communication), demonstrate that CAP is most effective when it is energetically favourable to unwind, and remove negative supercoils from a closed circular DNA helix[41].

Thus, although we do not know specific details of the transition, we propose that CAP traps specific regions of DNA in a left-handed helical conformation, and that the right-to-left helix transition denatures nearby regions of the helix in the promoter[42], possibly the region specifically unwound by RNA polymerase itself[43].

Anderson, Ohlendorf, Takeda and Matthews[44] have recently determined the course of the polypeptide backbone of λ cro protein and find that this repressor protein has two α-helices separated by 34 Å that can fit well into the major groove of right-handed B-DNA.

These data suggest the following general model for proteins that regulate gene transcription: (1) Two DNA-binding domains related by (approximate) 2-fold symmetry bind to one side of the DNA helix at the site of (approximate) 2-fold symmetry. (2) Base-specific recognition can be accomplished by α-helices of the protein binding in the major groove of B-type DNA. (3) Repressors, by binding to right-handed B-DNA, stabilize the right-handed helix and act as 'anti-melting' proteins; CAP protein, by binding to a left-handed helix, destabilizes the helix and facilitates binding and initiation by RNA polymerase.

This work was supported by USPHS grant GM-22778, NSF grant PCM-77-25111 and by postdoctoral fellowships from the American Cancer Society and US Public Health Service to D.B.M. We acknowledge assistance from Dr Harold W. Wyckoff, Paul Pepin, Gerald Johnson, John Kennedy, Art Perlo, Steve Anderson, Michael Fried and Melody Lane.

Received 3 November 1980; accepted 24 February 1981.

1. Zubay, G. & Doty, P. *J. molec. Biol.* **1,** 1–20 (1959).
2. Sung, M. T. & Dixon, G. H. *Proc. natn. Acad. Sci. U.S.A.* **67,** 1616–1623 (1970).
3. Adler, K. *et al. Nature* **237,** 322–327 (1972).
4. Church, G. M., Sussman, J. L. & Kim, S. H. *Proc. natn. Acad. Sci. U.S.A.* **74,** 1458–1462 (1977).
5. Zubay, G., Schwartz, D. & Beckwith, J. *Proc. natn. Acad. Sci. U.S.A.* **66,** 104–110 (1970).
6. Epstein, W., Rothman-Denes, L. B. & Hesse, J. *Proc. natn. Acad. Sci. U.S.A.* **72,** 2300–2304 (1975).
7. deCrombrugghe, B. *et al. Nature new Biol.* **231,** 139–142 (1971).
8. Ogden, S., Haggerty, D., Stoner, C. M., Kolodrubetz, D. & Schleif, R. *Proc. natn. Acad. Sci. U.S.A.* **77,** 3346–3350 (1980).
9. Musso, R. E., DiLauro, R., Adhya, S. & deChrombrugghe, B. *Cell* **12,** 847–854 (1977).
10. Anderson, W. B., Schneider, A. B., Emmer, M., Perlman, R. L. & Pastan, I. *J. biol. Chem.* **246,** 5929–5937 (1971).
11. Riggs, A. D., Reiness, G. & Zubay, G. *Proc. natn. Acad. Sci. U.S.A.* **68,** 1222–1225 (1971).
12. Krakow, J. G. & Pastan, I. *Proc. natn. Acad. Sci. U.S.A.* **70,** 2529–2533 (1973).
13. Eilen, E., Pampeno, G. & Krakow, J. S. *Biochemistry* **17,** 2469–2473 (1978).
14. Wu, C.-W. & Wu, F. Y.-H. *Biochemistry* **13,** 2573–2578 (1974).
15. Wu, F. Y.-H., Nath, K. & Wu, C.-W. *J. molec. Biol.* **100,** 459–472 (1976).
16. Eilen, E. & Krakow, J. S. *J. molec. Biol.* **114,** 47–60 (1977).
17. Kumar, S. A., Murthy, N. S. & Krakow, J. S. *FEBS Lett.* **109,** 121–124 (1980).
18. McKay, D. B. & Fried, M. G. *J. molec Biol.* **139,** 95–96 (1980).
19. Wyckoff, H. W. *et al. J. molec. Biol.* **27,** 563–578 (1967).
20. Rossmann, M. G. *Acta crystallogr.* **A32,** 774–777 (1976).
21. Strahs, G. & Kraut, J. *J. molec. Biol.* **35,** 503–512 (1968).
22. Harrison, S. C., Olson, A. J., Schutt, C. E., Winkler, F. K. & Bricogne, G. *Nature* **276,** 368–373 (1978).
23. Abad-Zapatero, C. *et al. Nature* **286,** 33–39 (1980).
24. Bennett, W. S. Jr & Steitz, T. A. *Proc. natn. Acad. Sci. U.S.A.* **75,** 4848–4852 (1978).
25. Amzel, M. & Poljak, R. J. *A. Rev. Biochem.* **48,** 961–998 (1979).
26. Eklund, H., Samama, J.-P., Wallen, L. & Branden, C.-I. *J. molec. Biol.* (in the press).
27. Takahashi, M., Blazy, B. & Baudras, A. *Biochemistry* **19,** 5124–5130 (1980).
28. Platt, T., Files, J. G. & Weber, K. *J. biol. Chem.* **248,** 110–121 (1973).
29. Geisler, N. & Weber, K. *Biochemistry* **16,** 938–943 (1977).
30. Jovin, T. M., Geisler, N. & Weber, K. *Nature* **269,** 668–672 (1977).
31. Pabo, C. O., Saver, R. T., Sturtevant, J. N. & Ptashne, M. *Proc. natn. Acad. Sci. U.S.A.* **76,** 1608–1612 (1979).
32. Ogata, R. T. & Gilbert, W. *Proc. natn. Acad. Sci. U.S.A.* **75,** 5851–5854 (1978).
33. Simpson, R. B. *Nucleic Acids Res.* **8,** 759–766 (1980).
34. Majors, J. thesis, Harvard Univ. (1979).
35. Gupta, G., Bansal, M. & Sasisekharan, V. *Proc. natn. Acad. Sci. U.S.A.* **77,** 6486–6490 (1980).
36. Dickson, R. C., Abelson, J., Johnson, P., Reznikoff, W. S. & Barnes, W. M. *J. molec Biol.* **111,** 65–79 (1977).
37. Taniguchi, T., O'Neill, M. & deCrombrugghe, B. *Proc. natn. Acad. Sci. U.S.A.* **76,** 5090–5094 (1979).
38. Wang, A. H.-J. *et al. Nature* **282,** 680–686 (1979).
39. Drew, H., Takano, T., Tanaka, S., Itakura, K. & Dickerson, R. E. *Nature* **286,** 567–573 (1980).
40. Sanzey, B. *J. Bact.* **138,** 40–47 (1979).
41. Cozzarelli, N. R. *Science* **207,** 953–960 (1980).
42. Dickson, R. G., Abelson, J., Barnes, W. M. & Reznikoff, N. S. *Science* **187,** 27–35 (1975).
43. Siebenlist, U. *Nature* **279,** 651–652 (1979).
44. Anderson, W. F., Ohlendorf, D. H., Takeda, Y. & Matthews, B. W. *Nature* **290,** 754–757 (1981).
45. Arnott, S. & Hukins, D. W. L. *Biochem. biophys. Res. Commun.* **47,** 1504–1509 (1972).
46. Wilson, I. A., Skehel, J. J. & Wiley, D. C. *Nature* **289,** 366–373 (1981).

# Arrangements and rearrangements of sequences flanking the two types of rDNA insertion in *D. melanogaster*

**Heli Roiha, J. Ross Miller, Lesley C. Woods & David M. Glover**

Cancer Research Campaign, Eukaryotic Molecular Genetics Research Group, Department of Biochemistry, Imperial College of Science and Technology, London SW7 2AZ, UK

*The two types of* Drosophila melanogaster *rDNA insertion occur at different sites in the 28S gene, separated by 51 nucleotide pairs. A short segment of rDNA is deleted at the site of the 5 kilobase type I insertion, whereas there is no such deletion of rDNA sequences which flank the type II insertion. Type I insertion elements in tandem arrays are flanked by very short segments of the 28S rRNA gene.*

A LARGE proportion of the ribosomal genes of *Drosophila melanogaster* contain an intervening sequence in the 28S rRNA gene[1–4]. There are two types of intervening sequence which have been termed the type I and type II insertions. The type I insertions are found in about 60% of the rDNA units of the X chromosome[5–7] and the type II insertions in about 15% of the rDNA units of both the X and Y chromosomes[6,8–10]. Physical mapping studies showed that these two non-homologous types of insertion are apparently located within the rDNA at identical sites, suggesting that there might be a structural feature of this region which could promote either the acquisition or loss of these elements. A high proportion of type I sequences is also present in the chromocentral heterochromatin[7,11], linked together in tandem arrays[12]. As far as is known, the type II insertions are only found in ribosomal gene repeats.

Intervening sequence elements have subsequently been reported in the distal part of the gene for the large rRNA of several other organisms[13–18]. The nucleotide sequences at the junctions between the gene and the intervening sequence have been determined for the 28S gene from *Tetrahymena*[19], the chloroplast rDNA from *Chlamydomomas*[20] and the mitochondrial rDNA of yeast[21]. In none of these cases is it possible to recognize a consensus sequence which might serve as a recognition site for enzymes which process the primary rRNA

0028–0836/81/180749—05$01.00