Proc. Natl. Acad. Sci. USA
Vol. 87, pp. 8165–8169, October 1990
Biochemistry

Protein–DNA conformational changes in the crystal structure of a λ Cro–operator complex

(Cro protein/protein–DNA interaction/DNA bending/DNA recognition)

RICHARD G. BRENNAN*†, STEVEN L. RODERICK*, YOSHINORI TAKEDA‡, AND BRIAN W. MATTHEWS*§

*Institute of Molecular Biology, Department of Physics, and Howard Hughes Medical Institute, University of Oregon, Eugene, OR 97403; ‡Laboratory of Molecular Biology, National Cancer Institute-Frederick Cancer Research Facility, Program Resources Inc., Frederick, MD 21701

Contributed by Brian W. Matthews, July 30, 1990

ABSTRACT **The structure of a complex of bacteriophage λ Cro protein with a 17-base-pair operator has been determined at 3.9-Å resolution. Isomorphous derivatives obtained by the synthesis of site-specific iodinated DNA oligomers were of critical importance in solving the structure. The crystal structure contains three independent Cro–operator complexes that have very similar, although not necessarily identical, conformations. In the complex, the protein dimer undergoes a large conformational change relative to the crystal structure of the free protein. One monomer rotates by about 40° relative to the other, this being accomplished primarily by a twisting of the two β-sheet strands that connect one monomer with the other. In the complex, the DNA is bent by about 40° into the shape of a boomerang but maintains essentially Watson–Crick B-form. In contrast to other known protein–DNA complexes, the DNA is not stacked end-to-end. The structure confirms the general features of the model previously proposed for the interaction of Cro with DNA.**

Recognition of specific DNA binding sites by gene regulatory proteins is crucial for biological function. The Cro protein from bacteriophage λ is a prototypical example of the class of transcriptional activators and repressors that incorporates the helix-turn-helix DNA-binding motif (1–6). Crystal (7–11) and NMR (12, 13) structures of several protein–DNA complexes have confirmed the importance of the helix-turn-helix motif but, at the same time, have revealed substantial variability in the use of the motif as a DNA recognition element. There is no simple code for DNA-protein recognition (14, 15).

Here we report the crystal structure of λ Cro complexed to a tight-binding 17-base-pair operator site.

Although the resolution is moderate, the crystal includes three complexes, allowing three independent views of the Cro dimer–DNA structure.

Structure Determination

Crystals of the complex were grown as described (16). The space group is $P3_2$ with $a = b = 154.8$ Å, $c = 86.0$ Å. The 17-base-pair fragment has the consensus sequence 5′-d[T(1)-A(2)-T(3)-C(4)-A(5)-C(6)-C(7)-G(8)-C(9)-G($\bar{8}$)-G($\bar{7}$)-G($\bar{6}$)-T($\bar{5}$)-G($\bar{4}$)-A($\bar{3}$)-T($\bar{2}$)-A($\bar{1}$)], where the numbers in parentheses indicate the base-pair position. There are three protein–DNA complexes (i.e., three Cro dimers and three DNA duplexes) per asymmetric unit. About 85% of the crystal volume is occupied by solvent. Two isomorphous heavy atom derivatives were obtained by the solid-phase synthesis of site-specific iodinated DNA oligomers. In one case, thymidine-3 was replaced with 5-iododeoxyuridine and in the second, cytidine-7 was replaced with 5-iodocytidine. Crystals were also grown with DNA in which cytidine-7 was replaced with 5-bromocytidine, and a third heavy atom derivative was obtained by soaking these crystals in 2 mM K_2PtCl_4 for 2 days. Intensity data were collected at the University of California San Diego Area Detector Facility. The iodouridine derivative was solved using a space-group-specific Patterson peak search algorithm. Difference Fourier methods were employed to solve the other derivatives. As expected the two iodo derivatives had six heavy atom sites per asymmetric unit. Fifteen platinum-binding sites were identified in the bromoplatinum derivative. Selected statistics are given in Table 1. A 5-Å-resolution multiple-isomorphous-replacement electron density map immediately revealed the locations of the DNA and protein, as well as large regions of solvent. The quality of the map was further improved by solvent leveling.

The Cro–DNA crystals diffract very weakly at high Bragg angles (16). This limitation, which is probably due to the unusually high solvent content, restricts the present analysis to a nominal resolution of 3.9 Å. In the electron density map (Fig. 1), the course of the DNA backbone can be seen clearly but individual base pairs are not resolved. On the other hand the four iodine sites per DNA duplex (at base pairs 3, $\bar{3}$, 7, and $\bar{7}$) are known accurately and are very helpful in assuring the correct placement of the DNA. Similarly, prior knowledge of the Cro protein structure (1, 18) made it easy to follow the backbone of each Cro monomer in the complex, even though individual side chains cannot, in general, be distinguished. The α-helical nucleus of each Cro monomer (residues 7–40) was placed in the electron density map by visual inspection and then refined by a computer program to optimize the fit. The backbone of the C-terminal residues 54–66 appears to adopt a different conformation in the complex relative to the uncomplexed protein (see below).

Crystal Packing

Each of the three independent complexes has a different environment in the crystal. All three complexes are seen to be similar, although not necessarily identical. Each complex has twofold symmetry, with individual Cro monomers related to their mates by calculated rotations of 172°, 176°, and 179°. As was suggested by the diffraction patterns (16), there is a pseudo threefold screw axis such that the three complexes are related by ≈120° rotations and 51.4-Å translations (Fig. 2). An unexpected (19) finding is a complete lack of base–base stacking interactions between the DNA duplexes. That is to say, unlike all previously reported protein–DNA structures,

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.

†Present address: Department of Biochemistry and Molecular Biology, Oregon Health Sciences University, Portland, OR 97201-3098.
§To whom reprint requests should be addressed.

¶The atomic coordinates have been deposited in the Protein Data Bank, Chemistry Department, Brookhaven National Laboratory, Upton, NY 11973 (reference 4CRO).

Table 1. Selected heavy atom derivative statistics

Derivative	Phasing power, F_H/E	R_{iso}, %
Bromoplatinum	1.39	27.5
Iodocytidine	1.31	16.3
Iodouridine	1.61	15.0

Phasing power is the ratio of the root-mean-square heavy atom scattering to the root-mean-square lack of closure of the phase triangles (17). R_{iso} is the fractional change in the observed structure amplitudes associated with the introduction of the heavy atoms. The figure of merit (17) is 0.50 for 17,659 reflections to 3.9-Å resolution. This represents 84% of the theoretically observable reflections to this resolution.

to our knowledge (7–11, 20), there are no pseudo continuous DNA helices extending through the crystal (19). Here, the closest approach between successive duplexes is 4.5 Å–5.5 Å between pairs of 5′-hydroxyl groups. In lieu of DNA–DNA stacking there are several intercomplex protein–protein and protein–DNA interactions that stabilize the crystal lattice. Each end of the DNA duplex contacts the surface of adjacent Cro–operator complexes in the vicinity of Lys-18, Lys-21, and Arg-13. There may also be a stacking interaction between the last base pair and the side chain of Phe-14. As a result of these repeated DNA–protein–DNA–protein contacts, there are essentially endless columns of Cro–operator complexes that crystallize as an open hexagonal network enclosing solvent-filled channels (Fig. 3).

DNA Conformation

For the most part, the DNA conformation of those bases involved in protein recognition (base pairs 1–6 and $\bar{1}$–$\bar{6}$) can be described as typically B-form. The complex, therefore, supports the idea (1, 3, 7–10, 18, 21) that recognition of the operator sequence by the protein is by specific contacts between the side chains of the protein and the base pairs in the major groove of the DNA (i.e., by "direct readout"). There is no evidence that the DNA backbone in the recognition region is significantly distorted, as would be diagnostic for indirect recognition (11, 14). In the middle of the operator, however, the DNA is bent through an angle of about 40°. In association with this bending, the middle five C+G-rich base pairs appear to be overwound and kinks are apparent between base-pairs 6 and 7 (and between base-pairs $\bar{6}$ and $\bar{7}$). One measure of the degree of bending can also be obtained from the iodine–iodine heavy atom positions at base-pairs 3 and $\bar{3}$, which are on the side of the DNA that faces the protein. In the three complexes, the iodine–iodine distances are 37.1 Å, 36.7 Å, and 37.3 Å (average, 37.0 Å). This is about 2.9 Å less than would be anticipated for straight B-DNA. NMR studies of Cro–O_R3 complexes (22–24) have indicated bending or distortion near the center of the DNA and theoretical calculations (25) suggest that electrostatic interactions favor Cro binding to a bent form of DNA.

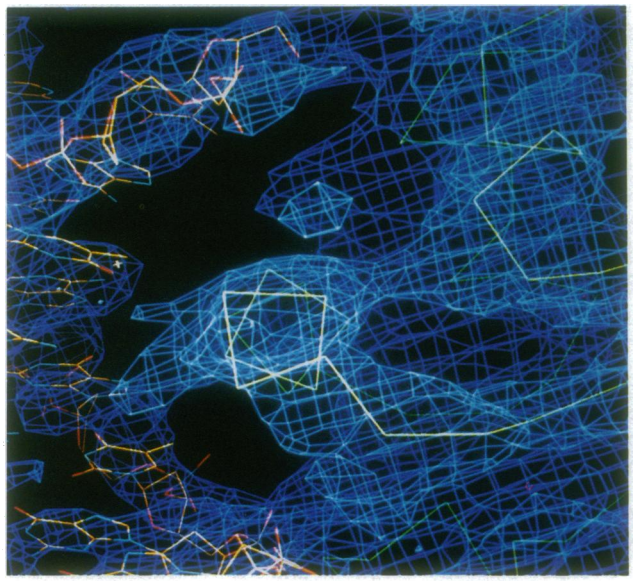


FIG. 1. Section of the electron density map for the Cro–operator complex, at 3.9-Å resolution, after solvent flattening. The view shows one of the six "recognition helices" in the crystal structure (in the center, seen end-on) within the major groove of the DNA. The electron density for the phosphate backbone (at the upper left) is clear but for the base pairs (at the lower left) is poorly defined. The α-carbon backbone of the Cro protein (shown center and right) was taken from the structure of the uncomplexed protein (1, 18) and fitted to the electron density map by a combination of manual and automatic procedures (see text).

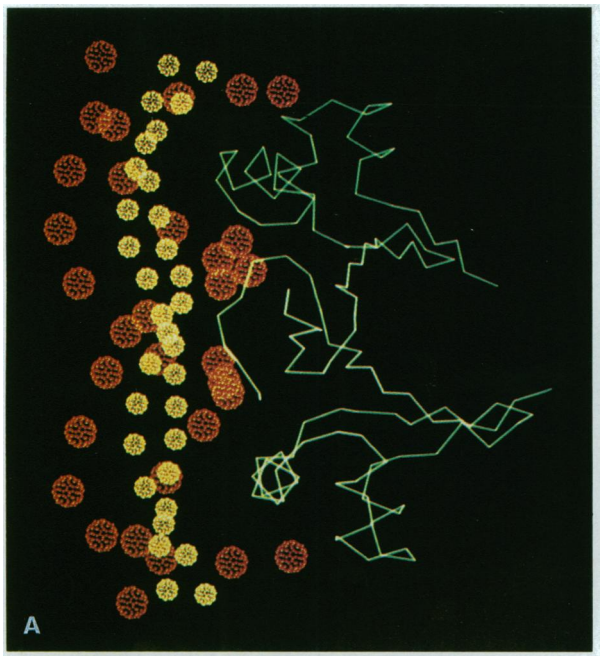
Protein Conformation

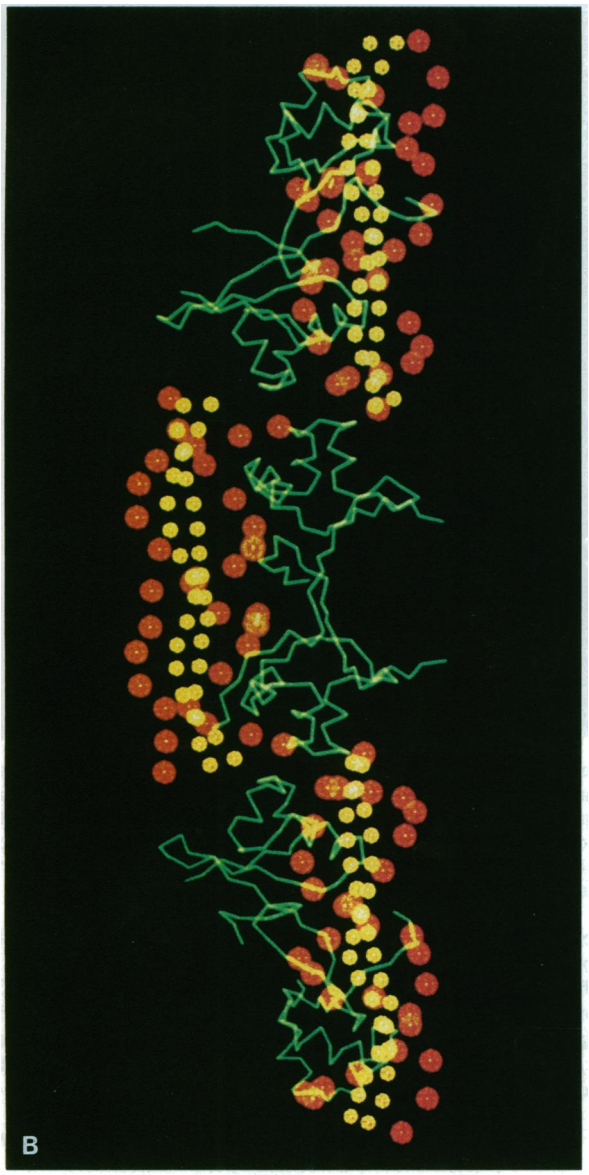
The electron density map suggests that the conformation of residues 7–50, which constitute most of the core of an individual Cro monomer, remains essentially unchanged on DNA binding. Within each of the three Cro dimer duplexes, however, one monomer rotates relative to the other by 35°, 48°, or 51°. This large quaternary change induced by DNA binding is accomplished primarily by a twisting in the vicinity of residues 52–54 within the β-sheet strands that connect one Cro monomer with another. Chemical modification (26) and NMR (24, 27) studies also show evidence for conformational changes in the vicinity of residues 52–54, including increased solvent accessibility of Tyr-51 and an increase in the exchange rate of the amide of Lys-56 upon specific DNA binding.

Specific Protein–DNA Interaction

Since the structure of Cro was first reported (1) and a model proposed (18) for its interaction with DNA the protein has been the subject of intense scrutiny (e.g., refs. 21–24, 26–33). These studies have supported, and the present work confirms, the essential features of the proposed model. In particular, the α_3 recognition helix of the protein is located within the major groove of the DNA and recognizes a sequence-specific site by direct interactions with the edges of the base pairs. As in other regulatory protein–DNA crystal structures, the side chains of the α_2-helix are important in the orientation and binding of the helix-turn-helix motif. The present work also confirms the suggestion from model building (18) that the binding of Cro to DNA might be associated with "hinge bending" of the protein and bending of the DNA.

On the basis of model building, a number of specific interactions were proposed between amino acids of the protein and base pairs within the major groove of the DNA. Some of these proposed interactions have, however, been in dispute. In the initial model (18), it was proposed that Ser-28 contacts base-pair T·A(3) and Lys-32 contacts base-pairs G·C(4) and T·A(5). Ptashne and coworkers (30, 31) have, however, argued on the basis of *in vivo* binding studies that Ser-28 contacts base-pair G·C(4) and Lys-32 contacts base-pairs T·A(5) and G·C(6). [These changes would make interactions for λ Cro correspond to those for λ repressor (10–31), which recognizes the same operator sites.] On the other hand, Takeda *et al.* (21) and Benson and Youderian (33) have presented *in vitro* and *in vivo* binding data that support the proposed interactions in original model. At the present resolution, it is not possible to unequivocally resolve these differences. The electron density map plus the known iodo-substitution sites permit the operator DNA to be placed with





an estimated uncertainty of about 1 Å. The placement of the known backbone of the protein within the electron density is more accurate (estimated error, ±0.5 Å). The individual side chains are, however, not resolved and the positions of the distal atoms, especially, are uncertain. Based on inspection of the six independent Cro monomers in the crystal structure, it appears that Ser-28 contacts base-pair 3, 4, or both. Lys-32 is closest to base-pair 5, probably also contacts base-pair 6 and might reach base-pair 4. There is some indication that the contacts may vary somewhat for the six independent Cro monomers in the crystal but, at the present stage of the analysis, it is not known if these are real differences due to the asymmetry of the operator and/or differences in crystalline environment or if they simply reflect experimental errors. If, indeed, Ser-28 contacts base-pairs A·T(3) and G·C(4) and Lys-32 contacts base-pairs G·C(4), T·4(5), and G·C(6) (possibly through water molecules), as consistent with the present analysis, it might serve to reconcile the binding data of Ptashne and coworkers (30, 31), Takeda *et al.* (21), and Benson and Youderian (33).

The proposed (18, 21) contacts involving Gln-27 and Arg-38 are consistent with the present structure. In the original model for Cro binding (18), it was proposed that Tyr-26 contacted base-pair 1 since this seemed to be the only interaction that was geometrically feasible. The present structure, however, suggests that it is actually Thr-17, in the α_2-helix, that contacts base-pair 1. (In the model Thr-17 contacted the phosphate backbone adjacent to base-pair 1 rather than the base pair itself.) These changes in the model now explain the observations (28, 34) that some substitutions of Tyr-26 cause a relatively small reduction in the affinity of Cro for operator DNA. Interaction between the γ-methyl of the Thr-17 and the methyl group of the thymine in T·C(7) could account, in part, for Cro's preference (21, 33) for T·A versus C·G at position 1.

The apparent direct contact between base-pair 1 and Thr-17 in the α_2-helix of Cro is distinctly different from the contacts observed for λ repressor (10), even though these two proteins bind the same operator sequences. It provides further evidence that the helix-turn-helix units in λ Cro and λ repressor do not align on the DNA in exactly the same way (35) and that there is not a simple one-on-one recognition code between amino acids and base pairs (14, 35, 36).

Finally, the determination of the structure of the complex provides information on the role of the C-terminal residues 60–66, Ser-Asn-Lys-Lys-Thr-Thr-Ala. These residues are disordered in the native protein (1) and, in the proposed complex (17), were postulated to lie along the minor groove of the DNA with the side chains interacting with the phosphate backbone. In the observed structure, Ser-60, Lys-62, and possibly Thr-64 do, indeed, seem to be in a position to interact with the phosphate backbone but, as a result of the conformational changes during complex formation, residues 61–66 also lie against the protein. [In other words, these residues do not "wrap around" the DNA like the N-terminal

Fig. 2. (*A*) Overall conformation of one of the three Cro–operator complexes. The protein α-carbon backbone is shown in green. For the DNA, the orange spheres indicate the positions of the phosphate backbone and the smaller yellow spheres show the bottom of the major and minor grooves. The direction of view is chosen to be along the axis of one of the recognition α-helices (at the bottom). In the crystal structure of the uncomplexed protein, the two recognition helices in the Cro dimer are essentially parallel (1). In the complexes, however, the second recognition helix (at the top) is rotated 35°–51° relative to the other. (*B*) Overall arrangement of the three independent Cro–operator complexes within the crystal structure (conventions as in *A*). Each complex is related to its neighbor by a translation of about 51 Å and a rotation of about 120°. There is no end-to-end stacking of the DNA strands.

FIG. 3. Arrangement of Cro–operator complexes within the crystal structure. Within each protein monomer the individual amino acids are shown as spheres, basic (blue), acidic (red), nonpolar (white), and polar noncharged (green). The orange and yellow spheres follow the phosphate backbone and the major and minor grooves of the operator DNA. In contrast to other known crystals of protein–operator complexes, the 17-base-pair DNA segments are not stacked end-to-end. The view direction is down the crystallographic *c* axis and a layer of molecules 1-unit-cell (86.0 Å) thick is included. The horizontal row of molecules extending across the bottom includes nine Cro–operator complexes and corresponds to three unit cells (i.e., 3 × 154.8 Å). The open spaces are solvent-filled channels that extend throughout the crystals.

arms of λ repressor (10).] These observations are consistent with the studies of Caruthers and coworkers (37) on C-terminal deletion and site-directed mutants of Cro in which positions 60, 62, and 64 were shown to be more important than positions 61, 63, and 65.

We are especially grateful to Drs. K. I. Varughese, C. Nielsen, and Ng. Xuong at the University of California San Diego Area Detector Facility for their help in measuring the x-ray diffraction data. We also thank Dr. M. Lindorfer, Biotechnology Laboratory, University of Oregon, for excellent technical assistance and Debbie Holland for expert help in the preparation of figures; Fig. 3 was prepared by Debbie Holland and B.W.M. The coordinates of uncomplexed Cro protein were provided by Drs. D. H. Ohlendorf and D. E. Tronrud (unpublished data). This work was supported in part by grants from the National Institutes of Health (to R.G.B. and B.W.M.) and the Lucille P. Markey Charitable Trust (to B.W.M.).

1. Anderson, W. F., Ohlendorf, D. H., Takeda, Y. & Matthews, B. W. (1981) *Nature (London)* **290**, 754–758.

2. McKay, D. B. & Steitz, T. A. (1981) *Nature (London)* **290,** 744–749.
3. Pabo, C. O. & Lewis, M. (1982) *Nature (London)* **298,** 443–447.
4. Schevitz, R. W., Otwinowski, Z., Joachimiak, A., Lawson, C. L. & Sigler, P. B. (1985) *Nature (London)* **317,** 782–786.
5. Brennan, R. G. & Matthews, B. W. (1989) *Trends Biochem. Sci.* **14,** 286–290.
6. Brennan, R. G. & Matthews, B. W. (1989) *J. Biol. Chem.* **264,** 1903–1906.
7. Anderson, J. E., Ptashne, M. & Harrison, S. C. (1987) *Nature (London)* **326,** 846–852.
8. Aggarwal, A. K., Rodgers, D., Drottar, M., Ptashne, M. & Harrison, S. C. (1988) *Science* **242,** 899–907.
9. Wolberger, C., Dong, Y., Ptashne, M. & Harrison, S. C. (1988) *Nature (London)* **335,** 789–795.
10. Jordan, S. R. & Pabo, C. O. (1988) *Science* **242,** 893–899.
11. Otwinowski, R. W., Schevitz, R. W., Zhang, R.-G., Lawson, C. L., Joachimiak, A., Marmorstein, R. Q., Luisi, B. F. & Sigler, P. B. (1988) *Nature (London)* **335,** 321–329.
12. Boelens, R., Scheek, R. M., van Boom, J. H. & Kaptein, R. (1987) *J. Mol. Biol.* **193,** 213–216.
13. Lammerichs, R. M. J. N., Boelens, R., van der Mavel, G. A., van Boem, J. H., Kaptein, R., Buck, F., Fara, B. & Rutenjans, H. (1989) *Biochemistry* **28,** 2985–2991.
14. Matthews, B. W. (1988) *Nature (London)* **335,** 294–295.
15. Sarai, A. & Takeda, Y. (1989) *Proc. Natl. Acad. Sci. USA* **86,** 6513–6517.
16. Brennan, R. G., Takeda, Y., Kim, J., Anderson, W. F. & Matthews, B. W. (1986) *J. Mol. Biol.* **188,** 115–118.
17. Blow, D. M. & Crick, F. H. C. (1959) *Acta Cryst.* **12,** 794–802.
18. Ohlendorf, D. H., Anderson, W. F., Fisher, R. G., Takeda, Y. & Matthews, B. W. (1982) *Nature (London)* **298,** 718–723.
19. Anderson, J., Ptashne, M. & Harrison, S. C. (1984) *Proc. Natl. Acad. Sci. USA* **81,** 1307–1311.
20. McClarin, J. A., Frederick, C. A., Wang, B. C., Greene, P., Boyer, H. W., Grable, J. & Rosenberg, J. M. (1986) *Science* **234,** 1526–1542.
21. Takeda, Y., Sarai, A. & Rivera, Y. M. (1989) *Proc. Natl. Acad. Sci. USA* **86,** 439–443.
22. Kirpichnikov, M. P., Hahn, K. D., Buck, F., Ruterjans, H., Chernov, B. K., Kurochkin, A. V., Skryabin, K. G. & Bayev, A. A. (1984) *Nucleic Acids Res.* **21,** 3551–3561.
23. Metzler, W. J. & Lu, P. (1989) *J. Mol. Biol.* **205,** 149–164.
24. Lee, S. J., Shirakawa, M., Akutsu, H., Kyogoku, Y., Shiraishi, M., Kitano, K., Shin, M., Ohtuska, E. & Ikehara, M. (1987) *EMBO J.* **6,** 1129–1135.
25. Matthew, J. B. & Ohlendorf, D. H. (1985) *J. Biol. Chem.* **260,** 5860–5862.
26. Takeda, Y., Kim, J., Caday, C. G., Steers, E., Jr., Ohlendorf, D. H., Anderson, W. F. & Matthews, B. W. (1986) *J. Biol. Chem.* **261,** 8608–8616.
27. Leighton, P. & Lu, P. (1987) *Biochemistry* **26,** 7262–7271.
28. Eisenbeis, S. J., Nasoff, M. S., Noble, S. A., Bracco, L. P., Dodds, D. R. & Caruthers, M. H. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 1084–1088.
29. Pakula, A. A., Young, V. B. & Sauer, R. T. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 8829–8833.
30. Hochschild, A. & Ptashne, M. (1986) *Cell* **44,** 925–933.
31. Hochschild, A., Douhan, J., III, & Ptashne, M. (1986) *Cell* **47,** 807–816.
32. Benson, N., Sugiono, P. & Youderian, P. (1988) *Genetics* **188,** 21–29.
33. Benson, N. & Youderian, P. (1989) *Genetics* **121,** 5–12.
34. Pakula, A. A. & Sauer, R. T. (1989) *Proteins: Struct. Funct. Genet.* **5,** 202–210.
35. Ohlendorf, D. H., Anderson, W. F., Lewis, M., Pabo, C. O. & Matthews, B. W. (1983) *J. Mol. Biol.* **169,** 757–769.
36. Anderson, W. F., Takeda, Y., Ohlendorf, D. H. & Matthews, B. W. (1982) *J. Mol. Biol.* **159,** 745–751.
37. Hubbard, A. J., Eisenbeis, S. J., Bracco, L. P., Gayle, R. B., Beaton, G. & Caruthers, M. H. (1990) *Biochemistry*, in press.