Proc. Natl. Acad. Sci. USA
Vol. 79, pp. 3097–3100, May 1982
Biochemistry

Structural similarity in the DNA-binding domains of catabolite gene activator and *cro* repressor proteins

(gene regulation/DNA–protein interaction/protein structure/α-helical fold)

T. A. STEITZ*, D. H. OHLENDORF†, D. B. MCKAY*, W. F. ANDERSON‡, AND B. W. MATTHEWS†

*Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, Connecticut 06511; †Institute of Molecular Biology and Department of Physics, University of Oregon, Eugene, Oregon 97403; and ‡Department of Biochemistry, University of Alberta, Edmonton, Alberta T6G 2H7 Canada

Communicated by Peter H. von Hippel, February 4, 1982

ABSTRACT It is shown that there is a structural similarity between the presumed DNA-binding regions of the *Escherichia coli* catabolite gene activator protein ("CAP") and the *cro* repressor protein ("cro") from bacteriophage λ. The correspondence between the two proteins is particularly striking for a structural unit consisting of two consecutive α-helices. The 24 α-carbon atoms that constitute the two-helical structural units in the two proteins can be superimposed with a root-mean-square disagreement of 1.1 Å. It is shown that this agreement is very unlikely to be due to a chance correspondence. For both CAP activator and cro repressor proteins it is the second α-helix of the two-helical unit that has been proposed to bind within the major groove of left-handed or right-handed B DNA, respectively [McKay, D. B. & Steitz, T. A. (1981) *Nature (London)* 290, 744–749; Anderson, W. F., Ohlendorf, D. H., Takeda, Y. & Matthews, B. W. (1981) *Nature (London)* 290, 754–758]. The structural correspondence between CAP and cro seen here, together with other recent evidence of sequence homologies between cro, CAP, and other proteins that bind double-stranded DNA, suggests that the two-helical unit is likely to be a common feature of many DNA-binding proteins. The results also suggest that some principles of specific protein–double-stranded DNA interaction may be general and include recognition via α-helices fitting into the major groove of the DNA.

Recently, the structures of two proteins that recognize specific nucleotide sequences in double-stranded DNA have been determined (1, 2). One of these proteins, the λ phage protein "cro," acts as a repressor; that is, it prevents transcription by RNA polymerase (3). The other protein, the *Escherichia coli* catabolite gene activator protein ("CAP"), functions primarily as an activator of transcription by RNA polymerase, although it can in certain systems also function as a repressor (4, 5). In this paper, we examine the similarities and differences in the structures of these two proteins and the way in which they appear to interact with double-stranded DNA.

Cro repressor is a tetramer in the crystal but probably acts as a dimer in solution. Each of the subunits is identical and has a molecular weight of approximately 7,351 (6). Model building suggests that the repressor binds to its operator DNA in the B form with a twofold symmetry axis of the protein coincident with that of the DNA (2). A pair of twofold-related α-helices of the repressor lie within successive major grooves of the DNA and are proposed to be a major determinant in recognition and binding. The centers of these two helices are 34 Å apart and have a tilt relative to the line connecting their centers that is appropriate for interaction with right-handed DNA.

CAP is a dimer of chemically identical 22,500 molecular weight subunits, with each subunit consisting of two distinct structural domains (1). The larger, amino-terminal, domain is observed to bind cyclic AMP within its interior, whereas the smaller, carboxyl-terminal, domain is presumed to interact with DNA. As in the case of cro repressor, the two DNA-binding domains of CAP each contain a protruding α-helix. Likewise, these two α-helices are 34 Å apart, with their helix axes related by a local protein twofold axis. However, the helices have a tilt relative to the line connecting their centers that is opposite to that observed for cro, and it has been proposed that CAP interacts with DNA via these two α-helices interacting in two successive major grooves of left-handed B DNA (1). The difference in the arrangement of the presumed DNA-binding helices in the respective cro and CAP dimers is shown in Fig. 1. As can be seen, the twofold-related α-helices are the same distance apart in the two proteins but are tilted in opposite directions.

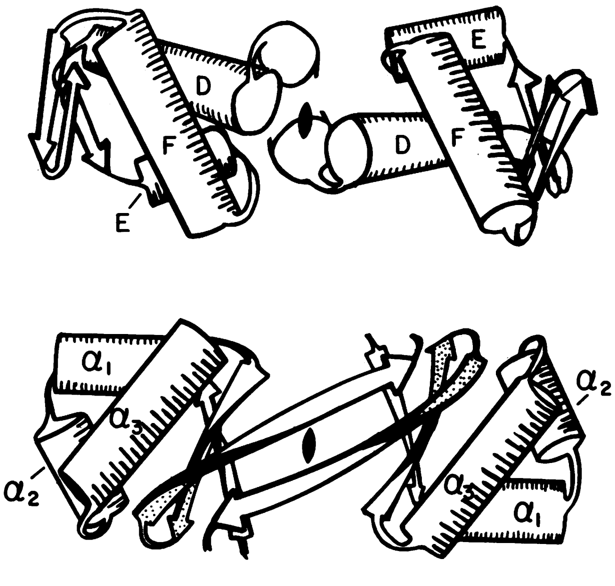

FIG. 1. Schematic drawings comparing the backbone conformations in the presumed DNA binding domain of CAP (*Upper*) and cro repressor (*Lower*). Helix nomenclature is as in refs. 1 and 2. In each case, dimers of the protein are depicted as seen along their respective twofold symmetry axes with the presumed DNA-binding α-helices (F in CAP and α_3 in cro) towards the viewer. The difference in the tilt of these α-helices is apparent.

Comparison of cro and CAP structures

The backbone structures of the cro repressor and the DNA-binding domains of CAP were compared in our laboratories, using the Evans & Sutherland Picture System II and MMS-X computer graphics systems (7). In these comparisons we first

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.

Abbreviation: rms, root mean square.

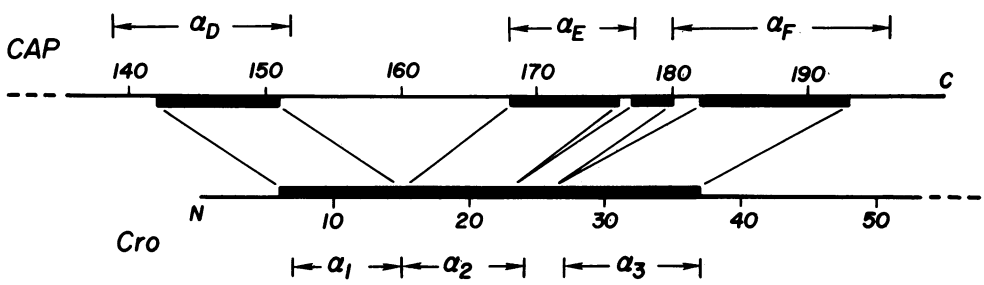

FIG. 2. Diagram showing those parts of the backbone of CAP that can be approximately superimposed on the backbone of cro. The connected solid bars indicate α-carbon atoms that are structurally equivalent in the two proteins. For the 31 equivalences, R_{C_α} = 3.1 Å. The diagram also shows that the D, E, and F α-helices of CAP approximately correspond to the α_1, α_2, and α_3 helices of cro.

looked for structural correspondence between one carboxyl-terminal domain of CAP and one monomer of cro, ignoring, for the moment, the difference described above in the respective quaternary structures. The atomic coordinates of cro have been refined to a crystallographic R-factor of 0.27 at 2.2 Å (unpublished). The coordinates of CAP have been obtained from a model in which the amino acid sequence has been fit to a 2.9-Å resolution map of CAP and the coordinates regularized (unpublished).

The major similarity between the two proteins is the existence of three α-helices in sequence (Fig. 1). These helices are labeled D, E, and F in CAP and are called α_1, α_2, and α_3 in cro. It is the F helix in CAP and the α_3 helix in cro that have been proposed, in the respective cases, to interact in the major groove of B DNA.

In order to quantitate the agreement between the two proteins we used the procedure developed by Rossmann and Argos (8, 9). Starting with an approximate alignment of the helices described above, the two proteins are rotated and translated to optimize the agreement between them. Where necessary, appropriate "deletions" are made in order to maximize the number of "equivalent" α-carbon atoms in the two structures. The results of this comparison are shown diagrammatically in Fig. 2. Altogether 31 "equivalent" atoms were found with a root-mean-square (rms) difference of 3.1 Å. As indicated in Fig. 2, and as can also be seen in the superposition of the two molecules in Fig. 3, most of the equivalent residues lie within the three consecutive α-helices. The correspondence of the first helices (α_1 of cro with D of CAP) is not particularly good, but the structural similarity of the second and third helices in the respective proteins (α_2 and α_3 of cro with E and F of CAP) is substantially better. It should also be noted that the superimposed α_3 and F helices are the presumed DNA-binding helices.

The structures of the rest of the subunits or domains are different in the two proteins. In the case of cro much of the rest of the protein forms an antiparallel β-sheet structure and an extended carboxyl-terminal arm that holds the subunits together. In the case of CAP the remaining polypeptide chain of the small domain is folded into what appears to be three antiparallel strands. Unlike the situation in cro, there is no interaction at all between the two small DNA-binding domains of CAP. Rather, the CAP dimer is held together entirely by the second, larger domain. In both molecules there is a very small portion of the structure that appears to be less well defined in the electron density maps and may or may not interact with DNA upon formation of a complex.

A two-helix supersecondary structure in CAP and cro

Because of the apparent close structural homology between the helices α_2 and α_3 of cro and the helices E and F of CAP we determined the agreement between these two-helical structural units. We found that the 24 consecutive α-carbon atoms 13–36 of cro superimposed within 1.1 Å on residues 166–189 of CAP. To estimate the error due to imprecision of coordinates, α-carbons 166–189 of one domain of CAP were superimposed on the corresponding α-carbons of the other domain; the rms difference in atomic coordinates was 0.7 Å in this case. For cro, the corresponding discrepancy is 0.4 Å. Therefore, the remarkable structural correspondence between CAP and cro, shown in Fig. 4, approaches the experimental error of the coordinates. It should be noted that the alignment of α-carbon atoms in cro and CAP for the 24-atom comparison is not exactly the same as for the whole-domain alignment shown in Fig. 2.

An estimate of the significance of the observed agreement between the two α-helices in CAP and cro was obtained in two ways. First, the empirical structure agreement probability plot of Remington and Matthews (10) shows that an agreement of 1.1 Å between 24 contiguous α-carbon atoms is significant at the level of about 3.5σ and is, therefore, quite unusual. Second, as a further test of the significance of the structural agreement between CAP and cro, we carried out a systematic search through all the proteins listed in the Brookhaven Protein Data Bank (11) in order to see if there were backbone segments of other proteins that agreed with the 24-residue segment of cro

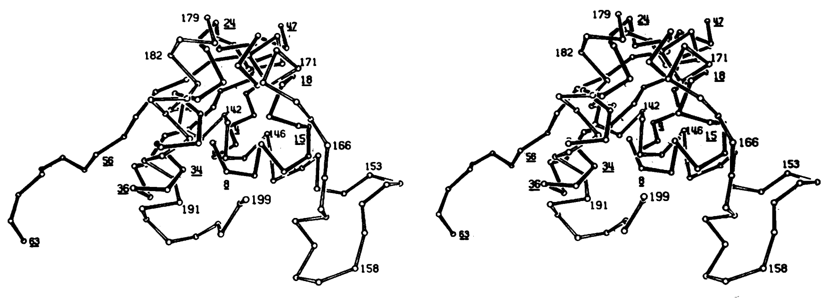

FIG. 3. Stereo drawing showing the backbone of CAP (open bonds) superimposed on the backbone of cro (solid bonds and residue numbers underlined). The presumed DNA-binding helices are labeled 24–36 (cro) and 179–191 (CAP). For cro protein, the extended carboxyl-terminal arm 56–63 is thought to be important in stabilizing the cro dimer, and the carboxyl-terminal residues 63–66 are disordered in the crystals (2).

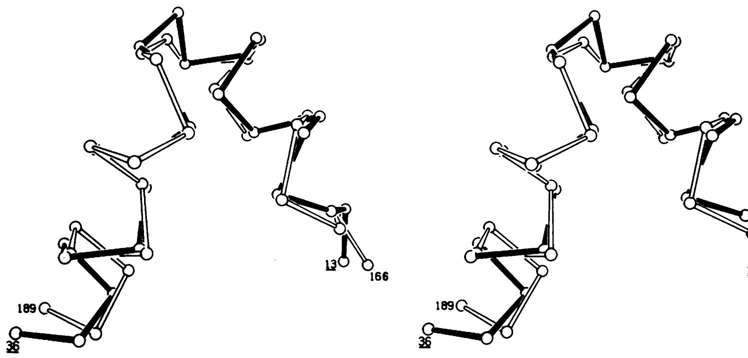

FIG. 4. Stereo drawing showing the close structural correspondence between the E and F helices of CAP (open bonds) and the α_2 and α_3 helices of cro (solid bonds).

as well as or better than had been observed for CAP. This search involved 21,540 comparisons of the two-helical cro segment with all possible 24-residue segments from 134 coordinate files. The *best* structural correspondence obtained from this search was for a part of the backbone of hen egg-white lysozyme and had an rms difference of 2.8 Å for the 24 contiguous α-carbon atoms. The average structure agreement for this exhaustive search through virtually every known protein structure was 6.9 Å, with a standard deviation of 1.7 Å [i.e., close to the values expected for a 24-residue comparison (10)]. It is striking to find that no other comparison even approaches the value of 1.1 Å for cro and CAP.

Thus, we conclude that identical two-helix supersecondary structures exist in CAP and cro and that this motif does not occur in any other protein in the protein structure data bank. Furthermore, the identity in structure between cro and CAP exists precisely in the region proposed in the two structures to bind specifically to DNA.

Specific cro and CAP binding to DNA

The structural correspondence between CAP and cro described here gives additional support to the proposed models for the binding of these two proteins to DNA (1, 2). In particular, we have proposed that it is the F helix of CAP and the α_3 helix of cro that bind within the major groove of double-stranded DNA (1, 2). Further, in both CAP and cro (12) the amino-proximal region of the E (or α_2) helix is capable of making specific contacts with the phosphate backbone of DNA. Now, independent of those proposals, we find that cro and CAP have a precise structural correspondence between the α_2–α_3 and E–F helical units. In each case, the α_3 and F helices protrude from the surfaces of the proteins and are separated by 34 Å from an (approximately) twofold-related helix in another subunit. This is, at minimum, consistent with the postulate that these helical units have similar functions in the two proteins. This proposal is further strengthened by the evidence from sequence homology that the same two-helical unit very likely occurs in a number of other proteins that also bind double-stranded DNA (see below).

However, it has to be emphasized that the comparisons made here are for isolated subunits of cro and CAP and therefore do not pertain to the question of the binding of right-handed or left-handed DNA as has been proposed for cro (2) and CAP (1). The reasons for favoring a left-handed or right-handed DNA conformation come from the relative arrangement of the DNA-binding helix of one subunit relative to the DNA-binding helix in the second subunit *in the dimers of cro and CAP* (Fig. 1), *and not from the polypeptide conformation within a single monomer.* Thus, individual monomers of cro and CAP have a common α_2–α_3 (or E–F) helical conformation, but the relative arrangements of these structural units within the dimers of CAP and cro are very different (Fig. 1), leading to the different models for DNA binding. The tilt that the two pairs of proposed DNA binding α-helices make with respect to the line connecting their centers differs in a "mirror image" fashion (Fig. 1). That is, the hand of the DNA to which these helices are complementary is different in the two cases. The difference in the tilt of the α-helices can be attributed to a difference in the subunit interaction in the proteins. Examination of Fig. 1 shows that the difference in the hand of the DNA to which each protein is complementary can be changed (at least in principle) by sliding one subunit relative to the other along the direction of the F or α_3 helices, or, alternatively, by rotating one subunit relative to the other by about 60°.

A general two-helix motif for DNA recognition

Sequence comparisons (refs. 12 and 13; unpublished) suggest that the two-helical DNA-binding fold observed in CAP and cro probably occurs in a number of other DNA-binding proteins. On the basis of amino acid sequence comparisons and DNA gene sequence comparisons, it appears that parts of the *c*I and *c*II proteins from bacteriophage λ, the repressor protein from the *Salmonella* phage P22, and 434-cro, the cro-like repressor from phage 434, are all homologous with the helical DNA-binding region of cro (13). In addition, there is also an apparent amino acid and gene sequence homology between *lac* repressor of *E. coli* and the above proteins. In this case it appears that the first 26 or so amino-terminal residues of *lac* repressor may fold similarly to the α_2 and α_3 helices of cro (12). Furthermore, the recently determined gene sequence of CAP shows a striking homology to *lac* repressor on the level of both the DNA and protein sequence in the 24 amino acid region of the two-helical fold described here (unpublished observations).

These data taken together suggest that a similar motif of α-helices will be found in many of the proteins that bind specifically to double-stranded DNA. One common component of the specific recognition of the DNA sequence is likely to be provided by the amino acid side chains of an α-helix that protrudes from the surface of the protein and fits into the major groove of B DNA. We would anticipate that the structures of many other proteins that specifically recognize double-stranded DNA sequences would have at least this two-helix motif and further that some DNA or protein sequence similarity would exist.

CAP is different from all the other DNA-binding proteins listed above in that its presumed DNA binding region is toward the carboxyl terminus of the molecule. Cro and 434-cro are both very small proteins of about 70 amino acid residues, and in this case the polypeptide folds to form a single DNA-binding domain. Association of a pair of monomers about a twofold symmetry axis then yields a dimer in which the twofold axis of the protein can align with the local twofold symmetry axis normal to the DNA, thereby doubling the area of interaction between protein and DNA (1, 2). In the case of the larger proteins *c*I, P22, and *lac* repressors, the amino-terminal part of the polypeptide folds to form a DNA-binding "headpiece" whereas the carboxyl part of the molecule forms an essentially separate domain (14–17). The addition of this second domain adds another level of sophistication to the function of the protein. It is not unlikely that the first double-stranded DNA-binding proteins to evolve were relatively small and had elements in common with the cro protein we see today. Subsequently, additional domains were added in different instances, as a result of which the basic DNA-binding function could be modified. Thus, the structural fold observed in cro and the DNA-binding domain of CAP—two helices folded in such a way that one protrudes from the surface of the protein and hence could penetrate the major groove of a DNA helix—may be a general motif for sequence-specific recognition of DNA by proteins.

We thank Terry Gray for his help with implementing M. G. Rossmann's comparison program and for instruction in its use. This work was supported in part by grants from the National Science Foundation (PCM-8014311 and PCM-775111), National Institutes of Health (GM20066, GM28138, and GM22778), and the M. J. Murdock Charitable Trust (to B.W.M. and T.A.S.); from the Medical Research Council Group on Protein Structure and Function of Canada (to W.F.A.); and by postdoctoral fellowships from the U.S. Public Health Service (to D.B.M. and D.H.O.).

1. McKay, D. B. & Steitz, T. A. (1981) *Nature (London)* **290**, 744–749.
2. Anderson, W. F., Ohlendorf, D. H., Takeda, Y. & Matthews, B. W. (1981) *Nature (London)* **290**, 754–758.
3. Echols, H. (1971) *Annu. Rev. Genet.* **6**, 157–190.
4. Zubay, G., Schwartz, D. & Beckwith, J. (1970) *Proc. Natl. Acad. Sci. USA* **66**, 104–110.
5. Musso, R. E., DiLauro, R., Adhya, S. & de Crombrugghe, B. (1977) *Cell* **12**, 847–854.
6. Hsiang, M. W., Cole, R. D., Takeda, Y. & Echols, H. (1977) *Nature (London)* **270**, 275–277.
7. Molnar, C. E., Barry, C. D. & Rosenberg, F. U. (1976) *Technical Memorandum No.* 229 (Computer Systems Laboratory, Washington Univ., St. Louis, MO).
8. Rossmann, M. G. & Argos, P. (1976) *J. Mol. Biol.* **105**, 75–96.
9. Rossmann, M. G. & Argos, P. (1977) *J. Mol. Biol.* **109**, 99–129.
10. Remington, S. J. & Matthews, B. W. (1980) *J. Mol. Biol.* **140**, 77–79.
11. Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F., Jr., Brice, M. C., Rodgers, J. R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977) *J. Mol. Biol.* **112**, 535–542.
12. Matthews, B. W., Ohlendorf, D. H., Anderson, W. F. & Takeda, Y. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 1428–1432.
13. Anderson, W. F., Takeda, Y., Ohlendorf, D. H. & Matthews, B. W., *J. Mol. Biol.*, in press.
14. Ptashne, M., Jeffrey, A., Johnson, A. D., Maurer, R., Meyer, B. J., Pabo, C. O., Roberts, T. M. & Sauer, R. T. (1980) *Cell* **19**, 1–11.
15. Echols, H. (1980) in *The Molecular Genetics of Development*, eds. Loomis, W. & Leighton, T. (Academic, New York), pp. 1–16.
16. Sauer, R. T., Pan, J., Hopper, P., Hehir, K., Brown, J. & Poteete, A. R. (1981) *Biochemistry* **20**, 3591–3598.
17. Miller, J. H. & Reznikoff, W. S., eds. (1978) *The Operon* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).