JOURNAL OF BACTERIOLOGY, Aug. 1992, p. 5457–5461
0021-9193/92/165457-05$02.00/0

Vol. 174, No. 16

DNA Binding Specificity and Sequence of *Xanthomonas campestris* Catabolite Gene Activator Protein-Like Protein

QIANPING DONG AND RICHARD H. EBRIGHT*

Department of Chemistry and Waksman Institute, Rutgers University, New Brunswick, New Jersey 08855

Received 31 March 1992/Accepted 5 June 1992

The *Xanthomonas campestris* catabolite gene activator protein-like protein (CLP) can substitute for the *Escherichia coli* catabolite gene activator protein (CAP) in transcription activation at the *lac* promoter (V. de Crecy-Lagard, P. Glaser, P. Lejeune, O. Sismeiro, C. Barber, M. Daniels, and A. Danchin, J. Bacteriol. 172:5877–5883, 1990). We show that CLP has the same DNA binding specificity as CAP at positions 5, 6, and 7 of the DNA half site. In addition, we show that the amino acids at positions 1 and 2 of the recognition helix of CLP are identical to the amino acids at positions 1 and 2 of the recognition helix of CAP: i.e., Arg at position 1 and Glu at position 2.

The *Escherichia coli* catabolite gene activator protein (CAP; also referred to as cyclic AMP receptor protein) is a global regulator of gene expression (for reviews, see references 7 and 20). CAP mediates the response to glucose starvation in *E. coli*. CAP functions by binding, in the presence of the allosteric effector cyclic AMP, to specific DNA sites located in or near promoters. The consensus DNA site for CAP is 5′-AAATGTGATCTAGATCA CATTT-3′ (3, 11, 15, 23); the site is 22 bp in length and exhibits perfect twofold sequence symmetry. The crystallographic structure of CAP has been determined to 0.25-nm resolution (26), and the crystallographic structure of the CAP-DNA complex has been determined to 0.30-nm resolution (22). CAP is a dimer of two identical subunits, each of which is 209 amino acids in length and contains a helix-turn-helix DNA binding motif (for reviews on the helix-turn-helix motif, see references 19 and 24). The CAP-DNA complex has twofold symmetry: one subunit of CAP interacts with one half of the DNA site, and the other subunit of CAP interacts in a twofold-symmetry-related fashion with the other half of the DNA site. Amino acids of the second α-helix of the helix-turn-helix motif of CAP (the recognition helix of CAP) contact DNA base pairs of the DNA half site. Amino acid 1 of the recognition helix of CAP (i.e., Arg-180) forms H bonds with the guanine N7 and guanine O^6 atoms of G · C at position 5 of the DNA half site (22, 29) and determines specificity for G · C at position 5 of the DNA half site (29). Amino acid 2 of the recognition helix of CAP (i.e., Glu-181) forms an H bond with the cytosine N4 atom of G · C at position 7 of the DNA half site (10, 12, 14, 22) and determines specificity for T · A at position 6 of the DNA half site and specificity for G · C at position 7 of the DNA half site (9, 10, 12, 14).

The *Xanthomonas campestris* CAP-like protein (CLP) is a global regulator of gene expression (6). CLP regulates, directly or indirectly, a set of genes implicated in phytopathogenicity in *X. campestris* (6). The primary structure of amino acids 26 to 230 of CLP is homologous to the primary structure of CAP: 45% identical amino acids and 73% identical or conservatively substituted amino acids (6). It is likely that the overall three-dimensional structure of CLP is extremely similar to the overall three-dimensional structure of CAP. It is also likely that the overall three-dimensional structure of the specific CLP-DNA complex is extremely similar to the overall three-dimensional structure of the specific CAP-DNA complex.

Danchin and coworkers (6) have reported that CLP can substitute for CAP in transcription activation at the *E. coli* *lac* promoter in vivo. This result suggests that CLP may have a DNA binding specificity similar to, or the same as, that of CAP. It is difficult to reconcile this result with the published sequence of CLP, which indicates that the amino acids at positions 1 and 2 of the recognition helix of CLP (i.e., Ala and Gln) differ from the amino acids at positions 1 and 2 of the recognition helix of CAP (i.e., Arg and Glu). In the context of CAP, Ala is not a specificity-neutral substitution at position 1 of the recognition helix (29) and Gln is not a specificity-neutral substitution at position 2 of the recognition helix (14).

We have analyzed the DNA binding specificity of CLP at positions 5, 6, and 7 of the DNA half site, and we have redetermined the sequence of CLP. We have found that CLP has the same DNA binding specificity as CAP at positions 5, 6, and 7 of the DNA half site. In addition, we have found (consistent with the DNA binding specificity of CLP, but contrary to the published sequence of CLP [6]) that the amino acids at positions 1 and 2 of the recognition helix of CLP are identical to the amino acids at positions 1 and 2 of the recognition helix of CAP: i.e., Arg at position 1 and Glu at position 2.

DNA binding specificity of CLP. In vivo DNA binding experiments (28–31) were performed to analyze the DNA binding specificity of CLP with respect to positions 5, 6, and 7 of the DNA half site. Ten *E. coli* K-12 tester strains were utilized: XAE400, XAE451, XAE452, XAE453, XAE461, XAE462, XAE463, XAE471, XAE472, and XAE473 (Table 1). Each tester strain has two important components: (i) Δ*crp45*, a deletion of the gene that encodes wild-type CAP (4, 21), and (ii) *lacZ*, the gene that encodes β-galactosidase. In tester strain XAE400, *lacZ* is placed under the control of the wild-type *lac* promoter DNA site for CAP. In the remaining nine tester strains, *lacZ* is placed under the control of derivatives of the *lac* promoter DNA site for CAP that have G · C→A · T, G · C→C · G, and G · C→T · A substitutions at bp 5 of each DNA half site, T · A→A · T, T · A→C · G, and T · A→G · C substitutions at bp 6 of each DNA half site, or G · C→A · T, G · C→C · G, and

* Corresponding author.

TABLE 1. *E. coli* K-12 tester strains[a]

Strain	Genotype	Reference
XAE400	XA102 Δ*crp45 strA fnr1 zci*::Tn*10* [λ*i434plac5*]	30
XAE451	XA102 Δ*crp45 strA fnr1 zci*::Tn*10* [λ*i434plac5-P1*(−68*A*;−55*T*)]	30
XAE452	XA102 Δ*crp45 strA fnr1 zci*::Tn*10* [λ*i434plac5-P1*(−68*C*;−55*G*)]	30
XAE453	XA102 Δ*crp45 strA fnr1 zci*::Tn*10* [λ*i434plac5-P1*(−68*T*;−55*A*)]	30
XAE461	XA102 Δ*crp45 strA fnr1 zci*::Tn*10* [λ*i434plac5-P1*(−67*A*;−56*T*)]	31
XAE462	XA102 Δ*crp45 strA fnr1 zci*::Tn*10* [λ*i434plac5-P1*(−67*C*;−56*G*)]	31
XAE463	XA102 Δ*crp45 strA fnr1 zci*::Tn*10* [λ*i434plac5-P1*(−67*G*;−56*C*)]	31
XAE471	XA102 Δ*crp45 strA fnr1 zci*::Tn*10* [λ*i434plac5-P1*(−66*A*;−57*T*)]	30
XAE472	XA102 Δ*crp45 strA fnr1 zci*::Tn*10* [λ*i434plac5-P1*(−66*C*;−57*G*)]	30
XAE473	XA102 Δ*crp45 strA fnr1 zci*::Tn*10* [λ*i434plac5-P1*(−66*T*;−57*A*)]	30

[a] The *E. coli* K-12 tester strains are derivatives of strain XA102 [Δ(*lac-proAB*)*X111 argE*(Am) *metB ara rpoB nal* Su2 (5)]. The *crp* and *strA* markers are from strain CA8445 (*HfrH* Δ*crp45* Δ*cya854 strA thi* [21]). The *fnr1* and *zci*::Tn*10* markers are from strain ECL323 (*fnr1 zci*::Tn*10* Δ*lacU169 araD139 motA strA thi* [16]). The *fnr1* marker was included to prevent Fnr-dependent transcription of *lacP1*(−68*T*;−55*A*) (30). To analyze the DNA binding specificity of CAP, plasmid pXZCRP (31) was introduced into each of the tester strains; the resulting strains were designated XAE400/CRP, XAE451/CRP, etc. To analyze the DNA binding specificity of CLP, plasmid pQDCLP was introduced into each of the tester strains; the resulting strains were designated XAE400/CLP, XAE451/CLP, etc. Plasmid pXZCRP encodes CAP under the control of the *crp* promoter (31). Plasmid pQDCLP encodes CLP under control of the *crp* promoter. Plasmid pQDCLP was constructed in two steps. In step one, the 24-bp *Hin*dIII-*Xba*I polylinker DNA segment of plasmid pUC118 (25) was replaced by the 1.2-kb *Hin*dIII-*Xba*I *clp* structural gene DNA segment from plasmid pDIA5100 (6), thereby constructing plasmid pUC118-CLP. In step two, the 1.7-kb *Hin*dIII-*Bam*HI *crp* structural gene DNA segment of plasmid pYZCRP (32) was replaced by the 1.2-kb *Hin*dIII-*Bam*HI *clp* structural gene DNA segment of plasmid pUC118-CLP.

G · C→T · A substitutions at bp 7 of each DNA half site (Fig. 1).

To analyze the DNA binding specificities of CAP and CLP, plasmid pXZCRP (31) and plasmid pQDCLP (Table 1), respectively, were introduced into each of the 10 tester strains. For each resulting plasmid-bearing strain, the differential rate of β-galactosidase synthesis was determined by the method of Miller (18). Cultures were grown in LB medium (18) containing 100 μg of ampicillin per ml. The data were corrected for background, i.e., for the differential rate of β-galactosidase synthesis in the absence of CAP or CLP.

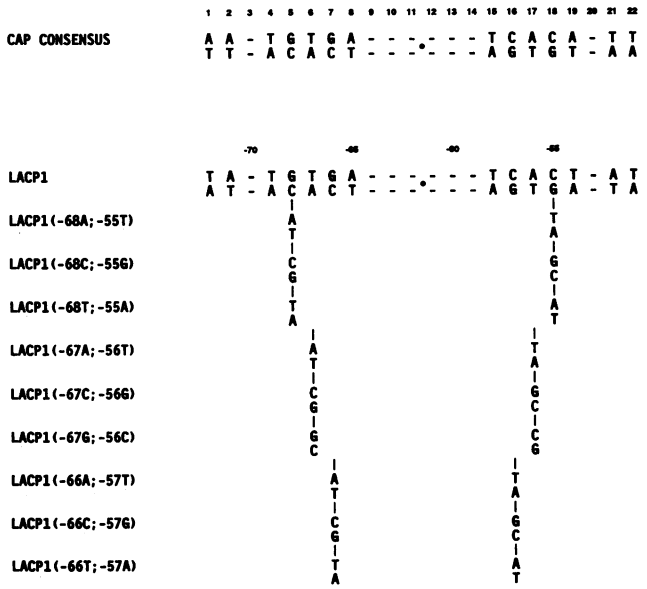

FIG. 1. DNA sites utilized in this study: the consensus DNA site for CAP (3, 11, 15, 23) and the wild-type *lac* promoter DNA site for CAP (bp −72 to −50 with respect to the start point of the *lac* promoter [8]). The symmetric A · T, C · G, and T · A substitutions at bp 5 of the DNA half site, the symmetric A · T, C · G, and G · C substitutions at bp 6 of the DNA half site, and the symmetric A · T, C · G, and T · A substitutions at bp 7 of the DNA half site are indicated beneath the sequence.

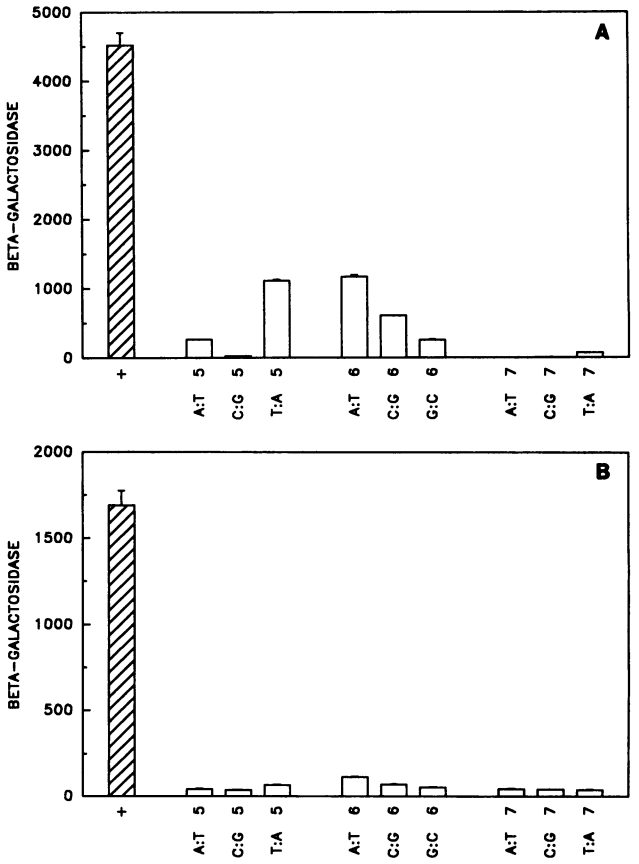

FIG. 2. Lac expression. (A) Data for CAP. (B) Data for CLP. Data are presented for the wild-type *lac* promoter (hatched bar) and for the derivatives of the *lac* promoter having the symmetric A · T, C · G, and T · A substitutions at bp 5 of the DNA half site for CAP, the symmetric A · T, C · G, and G · C substitutions at bp 6 of the DNA half site for CAP, and the symmetric A · T, C · G, and T · A substitutions at bp 7 of the DNA half site for CAP (open bars; sequences in Fig. 1). Data are corrected for background, i.e., for the differential rate of β-galactosidase synthesis in the absence of CAP or CLP (32 ± 1 U). Data are shown as the mean ± 1 standard error of the mean.

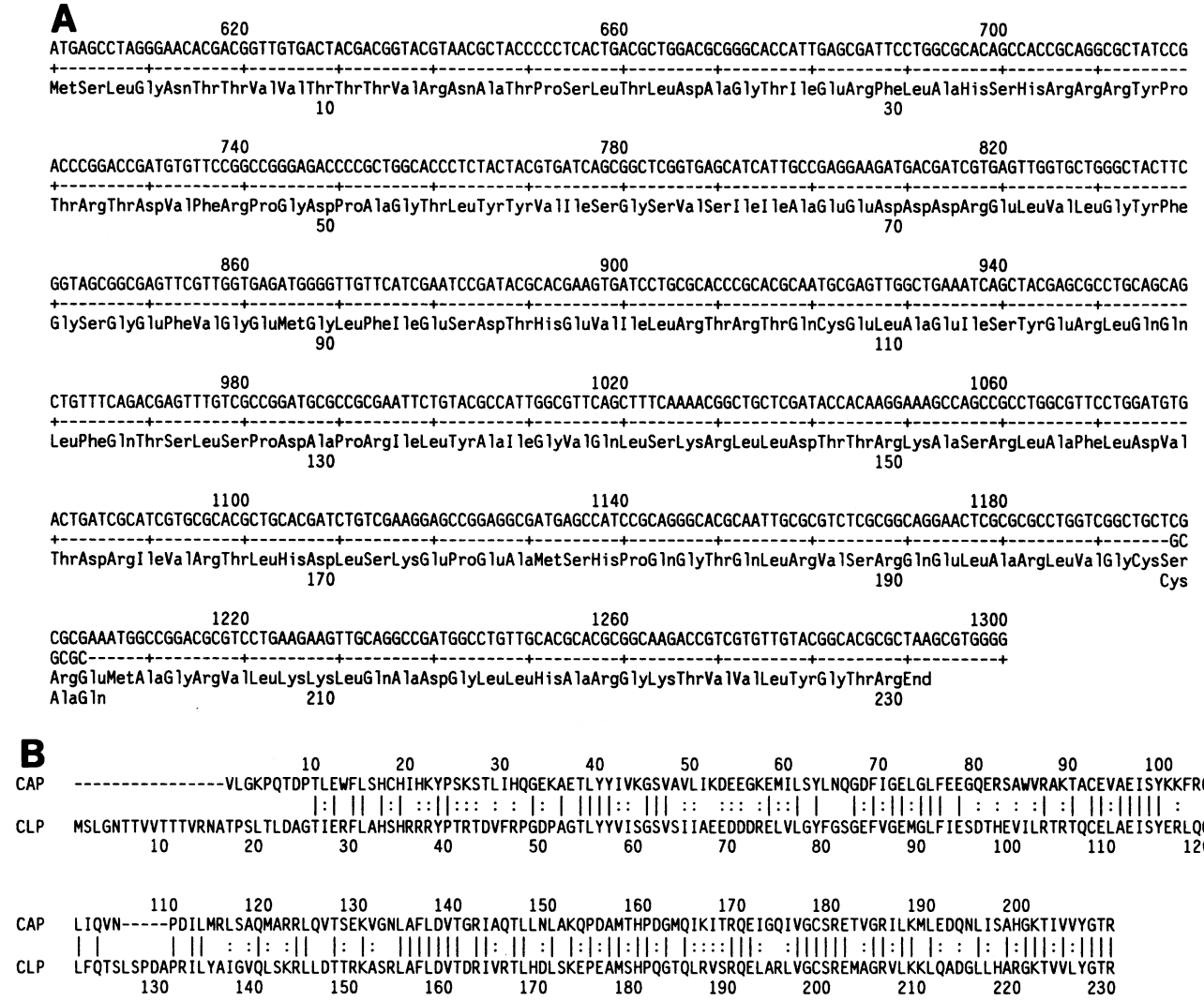

FIG. 3. (A) Nucleotide sequence and inferred amino acid sequence of the gene encoding CLP. Residues that differ in the previously reported nucleotide sequence and inferred amino acid sequence (6) are indicated beneath the sequence. Numbering of residues is as in reference 6. (B) Comparison of the inferred amino acid sequences of CAP and CLP. Identical and conservatively substituted amino acids are indicated by lines and colons, respectively.

The results are presented in Fig. 2. CLP, like CAP, exhibits strong specificity for G · C at position 5 of the DNA half site (G · C ≫ T · A > A · T > C · G), moderate to strong specificity for T · A at position 6 of the DNA half site (T · A ≫ A · T > C · G > G · C), and strong specificity for G · C at position 7 of the DNA half site (G · C ≫ A · T/C · G/T · A).

Sequence of CLP. The nucleotide sequence of positions 1160 to 1300 of the top strand and positions 1170 to 1300 of the bottom strand of the *clp* structural gene was redetermined, using double-stranded DNA of plasmid pDIA5100 (6) (numbering of nucleotide positions as in reference 6). In addition, the nucleotide sequence of the entire top strand of the *clp* structural gene was determined, using single-stranded DNA (25) prepared from plasmid pQDCLP.

The results are presented in Fig. 3. The actual nucleotide sequence of the gene encoding CLP is different from the previously reported nucleotide sequence (6). The actual

	180	181	182	183	184	185	186	187	188
CAP RECOGNITION HELIX	**ARG**	**GLU**	THR	VAL	GLY	**ARG**	ILE	LEU	LYS
CLP RECOGNITION HELIX	**ARG**	**GLU**	MET	ALA	GLY	**ARG**	VAL	LEU	LYS
	201	202	203	204	205	206	207	208	209

FIG. 4. Recognition helixes of CAP (1, 4, 19, 24) and CLP. In the structure of the CAP-DNA complex, three amino acids of CAP contact DNA base pairs of the DNA half site: i.e., amino acids 1, 2, and 6 of the recognition helix (9, 10, 12, 14, 22, 29). These amino acids of CAP and the corresponding amino acids of CLP are indicated in boldface type.

	156	157	158	159	160	161	162
CAP ACTIVATING REGION	ALA	MET	THR	HIS	PRO	ASP	GLY
PROPOSED CLP ACTIVATING REGION	ALA	MET	SER	HIS	PRO	GLN	GLY
	177	178	179	180	181	182	183

FIG. 5. Transcription-activating region of CAP (2, 13, 27, 28) and proposed transcription-activating region of CLP.

nucleotide sequence has CGCGCG at nucleotide positions 1198 to 1203, whereas the previously reported sequence had GCGCGC at nucleotide positions 1198 to 1203. The correction to the nucleotide sequence of the gene encoding CLP results in three corrections to the inferred amino acid sequence of CLP: i.e., Ser at amino acid 200, Arg at amino acid 201, and Glu at amino acid 202. Remarkably, the corrections to the inferred amino acid sequence of CLP involve positions −1, 1, and 2 of the recognition helix of CLP and increase the extent of amino acid identity with CAP (Fig. 3B and 4).

Implications. Our results indicate that CAP and CLP have the same DNA binding specificity at positions 5, 6, and 7 of the DNA half site. In addition, our results indicate that all three amino acids of CAP that contact DNA base pairs in the CAP-DNA complex (9, 10, 12, 14, 22, 29)—amino acids 1, 2, and 6 of the recognition helix of CAP—are conserved in CLP (Fig. 4). We propose that the equivalent amino acids of CAP and CLP make equivalent contacts in the respective protein-DNA complexes. Thus, we propose that amino acid 1 of the recognition helix of CLP (i.e., Arg-201) forms H bonds with the guanine N7 and guanine O^6 atoms of G · C at position 5 of the DNA half site in the CLP-DNA complex, that amino acid 2 of the recognition helix of CLP (i.e., Glu-202) forms an H bond with the cytosine N4 atom of G · C at position 7 of the DNA half site in the CLP-DNA complex, and that amino acid 6 of the recognition helix of CLP (i.e., Arg-206) forms an H bond with the thymine O^4 atom of A · T at position 8 of the DNA half site in the CLP-DNA complex.

Our results confirm that CLP, like CAP, is able to activate transcription at the *lac* promoter (6). In addition, our results indicate that CLP, like CAP, activates transcription at the *lac* promoter from a DNA site centered at position −61.5 relative to the transcription start point. (Our results indicate that transcription activation at the *lac* promoter by CLP is sensitive to twofold-symmetry-related substitutions at positions −55 and −68, positions −56 and −67, and positions −57 and −66 [Fig. 1 and 2].) Evidence has been presented that amino acids 156 to 162 of CAP are critical for transcription activation at the *lac* promoter by CAP but are not critical for DNA binding by CAP, and it has been suggested that these amino acids make direct protein-protein contacts with *E. coli* RNA polymerase in transcription activation at the *lac* promoter (2, 13, 17, 27, 28). Amino acids 156 to 162 of CAP are conserved in CLP (five of seven identical amino acids; seven of seven identical or conservatively substituted amino acids [Fig. 5]). We propose that the equivalent amino acids of CLP make equivalent direct protein-protein contacts with *E. coli* RNA polymerase in transcription activation at the *lac* promoter, and we speculate that the equivalent amino acids of CLP make equivalent direct protein-protein contacts with *X. campestris* RNA polymerase in transcription activation at *X. campestris* promoters.

Nucleotide sequence accession number. The GenBank accession number for the nucleotide sequence shown in Fig. 3A is M92289.

We thank V. de Crecy-Lagard and A. Danchin for plasmid pDIA5100. We also thank Y. W. Ebright for synthesis of oligodeoxyribonucleotide primers.

This work was supported by Public Health Service grant GM41376 from the National Institutes of Health to R.H.E. and by a New Jersey Molecular and Cellular Biology Predoctoral Fellowship to Q.D.

REFERENCES

1. **Aiba, H., S. Fujimoto, and N. Ozaki.** 1982. Molecular cloning and nucleotide sequencing of the gene for *E. coli* cAMP receptor protein. Nucleic Acids Res. **10:**1345–1360.
2. **Bell, A., K. Gaston, R. Williams, K. Chapman, A. Kolb, H. Buc, S. Minchin, J. Williams, and S. Busby.** 1990. Mutations that alter the ability of the *Escherichia coli* cyclic AMP receptor protein to activate transcription. Nucleic Acids Res. **18:**7243–7250.
3. **Berg, O., and P. von Hippel.** 1988. Selection of DNA binding sites by regulatory proteins. The binding specificity of cyclic AMP receptor protein to recognition sites. J. Mol. Biol. **200:** 709–723.
4. **Cossart, P., and B. Gicquel-Sanzey.** 1982. Cloning and sequence of the crp gene of *Escherichia coli*. Nucleic Acids Res. **10:**1363–1378.
5. **Coulondre, C., and J. Miller.** 1977. Genetic studies of the *lac* repressor. III. Additional correlation of mutational sites with specific amino acid residues. J. Mol. Biol. **117:**525–575.
6. **de Crecy-Lagard, V., P. Glaser, P. Lejeune, O. Sismeiro, C. Barber, M. Daniels, and A. Danchin.** 1990. A *Xanthomonas campestris* pv. campestris protein similar to catabolite activation factor is involved in regulation of phytopathogenicity. J. Bacteriol. **172:**5877–5883.
7. **de Crombrugghe, B., S. Busby, and H. Buc.** 1984. Cyclic AMP receptor protein: role in transcription activation. Science **224:** 831–838.
8. **Dickson, R., J. Abelson, P. Johnson, W. Reznikoff, and W. Barnes.** 1977. Nucleotide sequence changes produced by mutations in the *lac* promoter of *Escherichia coli*. J. Mol. Biol. **111:**65–75.
9. **Ebright, R., P. Cossart, B. Gicquel-Sanzey, and J. Beckwith.** 1984. Mutations that alter the DNA sequence specificity of the catabolite gene activator protein of *E. coli*. Nature (London) **311:**232–235.
10. **Ebright, R., P. Cossart, B. Gicquel-Sanzey, and J. Beckwith.** 1984. Molecular basis of DNA sequence recognition by the catabolite gene activator protein: detailed inferences from three mutations that alter DNA sequence specificity. Proc. Natl. Acad. Sci. USA **81:**7274–7278.
11. **Ebright, R., Y. Ebright, and A. Gunasekera.** 1989. Consensus DNA site for the *Escherichia coli* catabolite gene activator protein (CAP): CAP exhibits a 450-fold higher affinity for the consensus DNA site than for the *E. coli lac* DNA site. Nucleic Acids Res. **17:**10295–10305.
12. **Ebright, R., A. Kolb, H. Buc, T. Kunkel, J. Krakow, and J. Beckwith.** 1987. Role of glutamic acid-181 in DNA-sequence recognition by the catabolite gene activator protein (CAP) of *Escherichia coli*: altered DNA-sequence-recognition properties of [Val181]CAP and [Leu181]CAP. Proc. Natl. Acad. Sci. USA **84:**6083–6087.
13. **Eschenlauer, A. and W. Reznikoff.** 1991. *Escherichia coli* catabolite gene activator protein mutants defective in positive control of *lac* operon transcription. J. Bacteriol. **173:**5024–5029.
14. **Gunasekera, A.** 1991. Ph.D. thesis. Rutgers University, New Brunswick, N.J.
15. **Gunasekera, A., Y. Ebright, and R. Ebright.** DNA-sequence determinants for binding of the *Escherichia coli* catabolite gene activator protein (CAP). J. Biol. Chem., in press.
16. **Kuritzkes, D., X.-Y. Zhang, and E. C. C. Lin.** 1984. Use of Φ(*glp-lac*) in studies of respiratory regulation of the *Escherichia coli* anaerobic *sn*-glycerol-3-phosphate dehydrogenase genes (*glpAB*). J. Bacteriol. **157:**591–598.
17. **Li, X.-M. and J. Krakow.** 1988. Monoclonal antibodies that inhibit activation of transcription by the *Escherichia coli* cyclic AMP receptor protein. J. Biol. Chem. **263:**3448–3453.
18. **Miller, J.** 1972. Experiments in molecular genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
19. **Pabo, C., and R. Sauer.** 1984. Protein-DNA recognition. Annu. Rev. Biochem. **53:**293–321.
20. **Pastan, I., and S. Adhya.** 1976. Cyclic adenosine 5′-monophosphate in *Escherichia coli*. Bacteriol. Rev. **40:**527–551.
21. **Sabourin, D., and J. Beckwith.** 1975. Deletion of the *Escherichia coli crp* gene. J. Bacteriol. **122:**338–340.

22. **Schultz, S., G. Shields, and T. Steitz.** 1991. Crystal structure of a CAP-DNA complex: the DNA is bent by 90 degrees. Science **253:**1001–1007.
23. **Stormo, G., and G. Hartzell.** 1989. Identifying protein-binding sites from unaligned DNA fragments. Proc. Natl. Acad. Sci. USA **86:**1183–1187.
24. **Takeda, Y., D. Ohlendorf, W. Anderson, and B. Matthews.** 1983. DNA-binding proteins. Science **221:**1020–1026.
25. **Viera, J., and J. Messing.** 1987. Production of single-stranded plasmid DNA. Methods Enzymol. **153:**3–11.
26. **Weber, I., and T. Steitz.** 1987. Structure of a complex of catabolite gene activator protein and cyclic AMP refined at 2.5 Å resolution. J. Mol. Biol. **198:**311–326.
27. **Williams, R., A. Bell, G. Sims, and S. Busby.** 1991. The role of two surface exposed loops in transcription activation by the *Escherichia coli* CRP and FNR proteins. Nucleic Acids Res. **19:**6705–6712.
28. **Zhang, X.** 1991. Ph.D. thesis. Rutgers University, New Brunswick, N.J.
29. **Zhang, X., and R. Ebright.** 1990. Identification of a contact between arginine-180 of the catabolite gene activator protein (CAP) and base pair 5 of the DNA site in the CAP-DNA complex. Proc. Natl. Acad. Sci. USA **87:**4717–4721.
30. **Zhang, X., and R. Ebright.** 1990. Substitution of 2 base pairs (1 base pair per DNA half site) within the *Escherichia coli lac* promoter DNA site for catabolite gene activator protein places the *lac* promoter in the FNR regulon. J. Biol. Chem. **265:**12400–12403.
31. **Zhang, X., A. Gunasekera, Y. Ebright, and R. Ebright.** 1991. Derivatives of CAP having no solvent-accessible cysteine residues, or having a unique solvent-accessible cysteine residue at amino acid 2 of the helix-turn-helix motif. J. Biomol. Struct. Dyn. **9:**463–473.
32. **Zhou, Y., X. Zhang, and R. Ebright.** 1991. Random mutagenesis of gene-sized DNA molecules by use of PCR with *Taq* DNA polymerase. Nucleic Acids Res. **19:**6052.