Proc. Natl. Acad. Sci. USA
Vol. 82, pp. 1084–1088, February 1985
Biochemistry

Altered Cro repressors from engineered mutagenesis of a synthetic *cro* gene

(DNA synthesis/synthetic gene/protein–DNA interactions)

SCOTT J. EISENBEIS*, MARC S. NASOFF*†, STEWART A. NOBLE*‡, LAURENT P. BRACCO*, DAVID R. DODDS*, AND MARVIN H. CARUTHERS*§

*Department of Chemistry, University of Colorado, Boulder, CO 80309

Communicated by H. Gobind Khorana, September 21, 1984

ABSTRACT **A portion of the gene coding for the Cro repressor protein of bacteriophage λ has been chemically synthesized, incorporating base pair changes that generate restriction endonuclease sites without altering the amino acid coding sequence. These restriction endonuclease sites were used to remove small segments of the synthetic *cro* gene and the segments were replaced with duplexes carrying desired mutations. Altered Cro proteins produced by mutants constructed in this manner were then assayed for binding to λ operator O_R3 *in vivo*. Mutations directed into the region of the *cro* gene encoding the α-3 helix produced altered Cro proteins with a range of affinities for operator DNA. These changes suggest which amino acids play an important role in Cro–O_R3 complex formation.**

Recently, several new lines of investigation have provided substantial insight into the mechanism by which sequence-specific DNA-binding proteins recognize and bind to sites on double-stranded DNA. Of fundamental importance has been the elucidation of the structures of three DNA-binding proteins: the catabolite gene activator protein (CAP) from *Escherichia coli* (1), the amino-terminal fragment of cI repressor protein from bacteriophage λ (2), and Cro repressor protein, also from λ (3). All three proteins show structural similarities, the most notable of which is a protruding bihelical unit. It has been proposed that in each case one of the helices provides most of the sequence-specific contacts within the major groove of B-DNA (4–6). Furthermore, sequence homologies among other DNA-binding proteins suggest that this two-helix motif may be a common feature of DNA recognition (6–8). Cro protein, the smallest repressor characterized to date, binds as a dimer to six sites of 17 base pairs each, which are clustered into two operator regions on the phage genome (9). Refinement of the structure of Cro at high resolution has allowed a detailed model to be proposed for complex formation with operator DNA (10). The model predicts several specific contacts between each Cro monomer and base pairs in the major groove. These contacts involve amino acid side chains either within or near the α-3 helix of Cro. In addition, several sequence-independent interactions between the protein and the DNA backbone have been predicted (10).

In this report we describe the design, construction, and cloning of a synthetic *cro* gene fragment, which allow specific mutations in the coding region to be made with ease. The effect of several such mutations on Cro repressor binding to operator O_R3 is also described. The results support the current model for the interaction of Cro repressor with operator DNA.

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.

MATERIALS AND METHODS

Bacterial Strains, Bacteriophage, and Plasmids. *E. coli* strain 71-18 (Δ*lac–pro*/F′ *lacI*^{q} ZΔM15 *pro*$^+$) was lysogenized with the temperate bacteriophage λ112Δ265prmup-1, which contains the immunity region from phage 21 and a P_{rm}–*lacZ* operon fusion (11). In addition, it bears a deletion of O_R1 (Δ265) and carries a mutation (prmup-1) that increases the level of transcription initiated at P_{rm} in the absence of repressor. This lysogen was transformed with the plasmids pTR214, pTR190, or the pTR190 derivatives constructed in this work. Plasmids pTR214 and pTR190 contain the *lac* UV5 promoter operator region fused to *cro* (12).

Molecular Cloning. Isolation of plasmid, DNA cleavage with restriction endonucleases, isolation of DNA fragments, ligation with T4 DNA ligase (Bethesda Research Laboratories), and transformation of *E. coli* were carried out as described (13).

Deoxyoligonucleotide Synthesis and Enzymatic Ligations. Deoxyoligonucleotides were synthesized as previously described (14) except that purification was carried out by reversed-phase HPLC using a μBondapak C_{18} column from Waters Associates. Enzymatic ligations of synthetic deoxyoligonucleotides were completed as described (15).

Colony Hybridization. Colonies containing plasmids with synthetic inserts were screened by the method of Gergen *et al.* (16).

DNA Sequence Analysis. DNA sequences were confirmed by the method of Maxam and Gilbert (17).

RESULTS

Design of the Synthetic *cro* Gene Fragment. The plasmid pTR190 carries the wild-type *cro* gene under the control of the *lac* UV5 promoter (12). A deletion between the *lac* Shine–Dalgarno sequence and the *cro* initiation codon results in a plasmid that directs the synthesis of an extremely small amount of Cro protein (<0.0006% of the total soluble protein) (12). The *cro* gene as cloned in pTR190 contains a *Bgl* II site at the position coding for amino acid 22 of Cro and has a *Sal* I site approximately 300 base pairs downstream from the *cro* translational termination codon. Our plan was to remove this fragment, which includes DNA coding for amino acids 22–66 of Cro, from pTR190 and replace it with a synthetic duplex. The sequence of the duplex is shown in Fig. 1. The sequence design includes *Bgl* II and *Sal* I cohe-

Abbreviations: bp, base pairs; IPTG, isopropyl β-D-thiogalactopyranoside.

†Present address: Genetics Institute, 225 Longwood Avenue, Boston, MA 02115.
‡Present address: Glaxo Group Research Ltd., Greenford, Middlesex, UB6 OHE England.
§To whom reprint requests should be addressed.

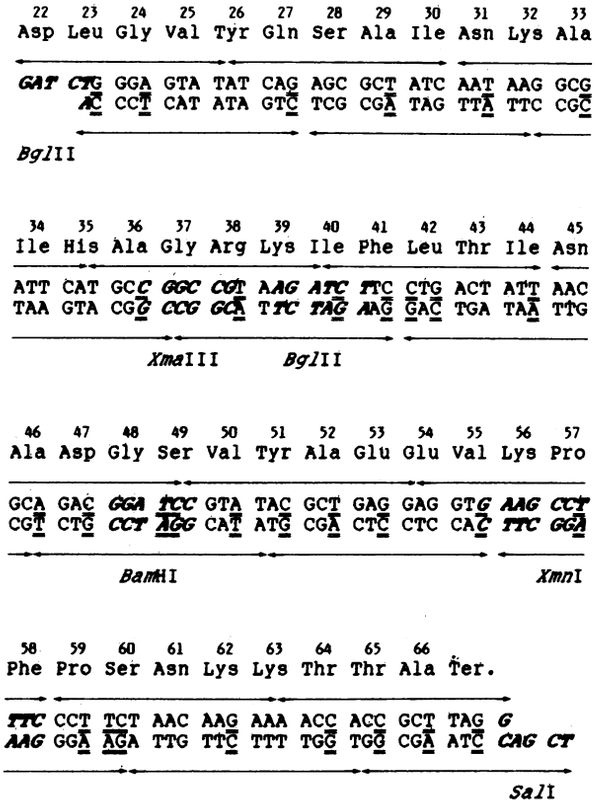

FIG. 1. The synthetic DNA duplex coding for amino acids 22–66 of Cro. Included in the design are *Bgl* II and *Sal* I cohesive ends and restriction endonuclease recognition sequences, which are indicated by bold italic type. The 31 base pair changes from the wild-type *cro* sequence are doubly underlined. These changes do not alter the amino acid sequence of Cro. Bars ending with arrows above and below the sequence indicate the length of the chemically synthesized deoxyoligonucleotides.

sive termini and internal restriction sites for *Bgl* II, *Bam*HI, *Xmn* I, and *Xma* III. These sites allow easy access to various regions of the *cro* gene, which can be replaced with small synthetic duplexes carrying desired mutations. For example, digestion with *Bgl* II allows access to the region coding for amino acids 22–40, which includes the α-3 helix of Cro (amino acids 27–36), that has been proposed to make specific contacts with base pairs of operator DNA (10). The 31 base pair changes designed into the synthetic *cro* gene fragment were introduced without changing the wild-type amino acid coding sequence or dramatically altering the codon usage pattern (Fig. 1). These changes were made in order to generate restriction sites and to prevent misligation of the synthetic deoxyoligonucleotides due to direct or inverted repeat sequences.

Chemical and Enzymatic Synthesis of the Synthetic *cro* Gene Fragment. Twenty oligomers were chemically synthesized by using the rapid solid-phase phosphoramidite approach (14). The ligation plan for assembling the *cro* gene fragment from these oligomers is shown in Fig. 2. The initial ligation was divided into three parts to minimize fragment mismatches and to maximize yield. The right half of the duplex formed readily (section C), while the left half was assembled in two reactions (A and B). Subsequent ligation of the products isolated from the initial reactions (A, B, and C) yielded a final product containing the entire 135-base-pair (bp) duplex.

Cloning the Synthetic *cro* Gene Fragment. The plan for constructing a plasmid carrying the synthetic *cro* gene fragment is shown in Fig. 3. The 135-bp duplex containing a *Bgl* II cohesive sequence at one end and a *Sal* I cohesive sequence at the other end was ligated into pTR190 from which the DNA fragment between the *Bgl* II and *Sal* I sites had previously been removed. This ligation mixture was used to transform *E. coli* strain 71-18 lysogenized with λ112Δ265prmup-1. The cells were plated on MacConkey medium containing ampicillin (20 μg/ml) and supplemented with 1 mM isopropyl β-D-thiogalactopyranoside (IPTG). Several white (non-

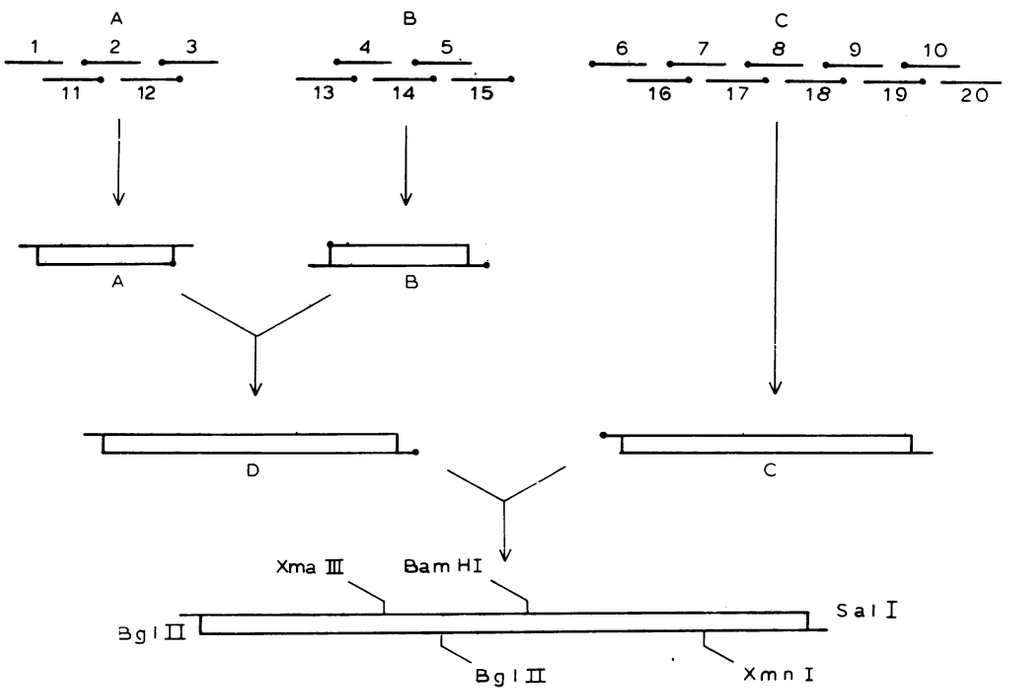

FIG. 2. Ligation plan for preparing the synthetic *cro* gene fragment. Three separate ligations (yielding sections A, B, and C) were carried out initially to maximize efficiency. Products from the initial ligations were purified by gel electrophoresis and ligated to form the final product. —●, Synthetic pieces that were phosphorylated prior to ligation. The segments are oriented so that the *Bgl* II and *Sal* I cohesive ends are schematically shown at the left and right ends, respectively.

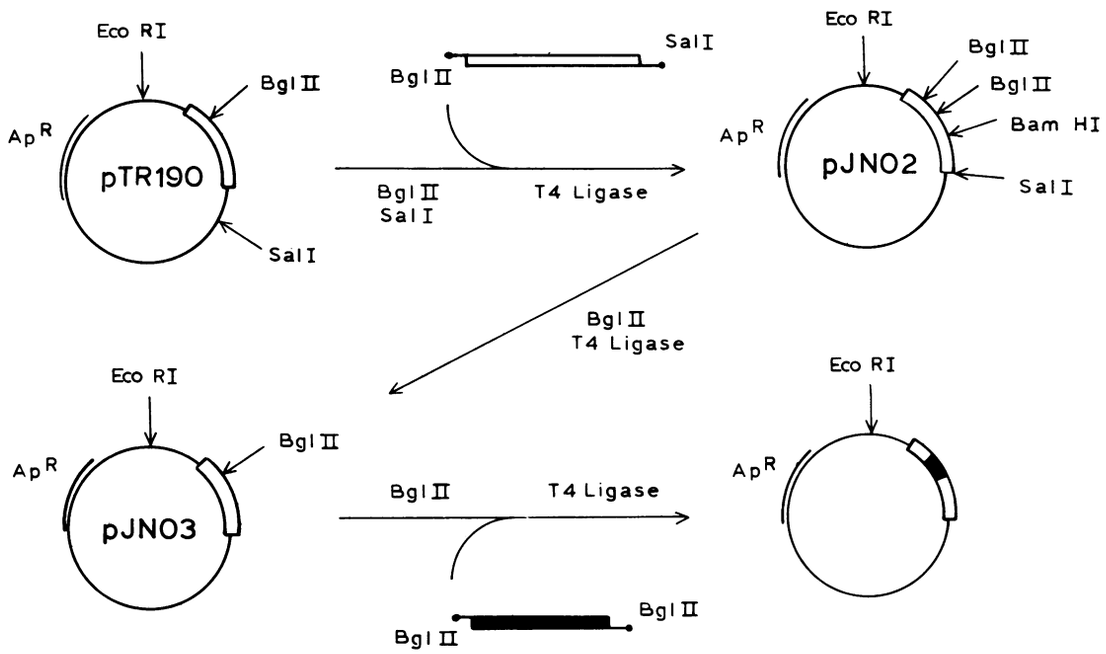

FIG. 3. Construction of a plasmid, pJN02, that carries the synthetic *cro* gene fragment and the strategy for directing mutations into the region of *cro* encoding the α-3 helix (see text). Ap^R, ampicillin resistance.

lactose-utilizing) colonies were obtained that were assumed to carry the synthetic *Bgl* II/*Sal* I fragment. The basis for this assumption was that cells carrying a functional *cro* gene will produce Cro protein that binds to O_R3 and turns off synthesis of β-galactosidase in this system (Fig. 4). Plasmid DNA from one of these clones was isolated and characterized by restriction mapping. DNA sequence analysis showed that this plasmid, pJN02, carried the synthetic duplex containing *Bgl* II and *Sal* I cohesive ends and restriction sites for *Bgl* II, *Bam*HI, *Xmn* I, and *Xma* III.

Incorporation of Mutations into the Synthetic *cro* Gene. The strategy for directing mutations into the *cro* gene region encoding the α-3 helix is depicted in Fig. 3. Plasmid pJN02 was cleaved with *Bgl* II and the resulting mix was ligated. After transformation, a plasmid (pJN03) missing the 49-bp *Bgl* II fragment was isolated. Several 49-bp synthetic duplexes, each containing a desired mutation, were constructed. The small duplexes were each assembled by ligating eight oligomers, 12–14 nucleotides in length, to form a 49-bp duplex with *Bgl* II cohesive ends. Each duplex was inserted into the *Bgl* II site of plasmid pJN03, and colonies carrying plasmids with synthetic inserts were identified by colony hybridization as shown in Fig. 5. Positive colonies were rescreened with a synthetic probe that spans the vector–insert junction to determine which clones contained hybrid plasmids with the synthetic fragment inserted in the proper orientation (Fig. 5). Under the appropriate conditions, only those plasmids having the insert in the proper orientation will hybridize to this probe. Finally, each insert having the proper orientation was characterized by DNA sequencing. A total of 11 mutations have been directed into *cro* by this approach. Each of these mutations alters the Cro protein sequence by a single amino acid.

***In Vivo* Cro-Binding Assay.** *E. coli* strain 71-18 (λ112Δ265-prmup-1) carrying plasmids containing either wild-type or synthetic *cro* genes were grown in M9 glucose medium (18) supplemented with ampicillin (20 μg/ml) and, where indicated, IPTG (1 mM) and assayed for β-galactosidase activity (18). Growth of the strains in the presence of IPTG inactivates *lac* repressor, resulting in the induction of Cro protein synthesis. Cro, in turn, shuts off transcription of *lacZ* initiated at P_{rm} (Fig. 4). An initial experiment was carried out to compare the operator binding of Cro made from the synthetic *cro* gene carried by pJN02 to that made from the wild-type sequences carried by pTR190 and pTR214 (Fig. 6). The latter plasmid directs the synthesis of at least 1000-fold more Cro (1% of the soluble protein) than pTR190 (12). Strains carrying plasmids pTR190, pJN02, or pTR214 all show a similar reduction in β-galactosidase activity in response to induction of Cro protein by IPTG, indicating that O_R3-saturating levels of Cro can be produced by all three strains. Cells carrying pJN03 (with the DNA coding for the α-3 helix of Cro deleted) are unable to turn off β-galactosidase expression when induced by IPTG. The slight level of apparent repression of β-

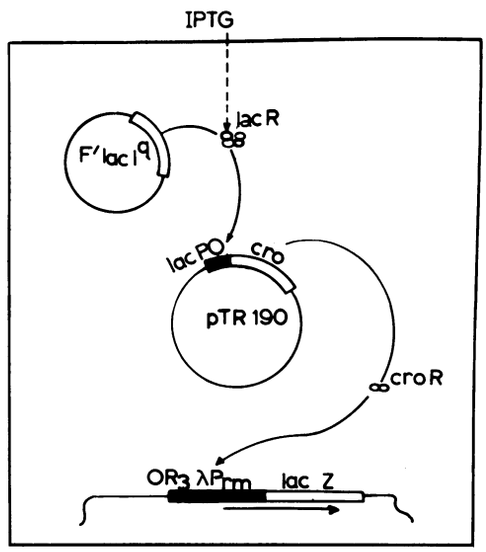

FIG. 4. *In vivo* system for analysis of Cro binding to O_R3. An *E. coli* strain carrying an F' *lacI*^{q} was lysogenized with λ112Δ265prmup-1 and transformed with pTR214, pTR190, or a synthetic derivative of pTR190 described in the text. Cro binding to O_R3 is determined indirectly by measuring the activity of β-galactosidase (the *lacZ* product) in the presence or absence of IPTG.

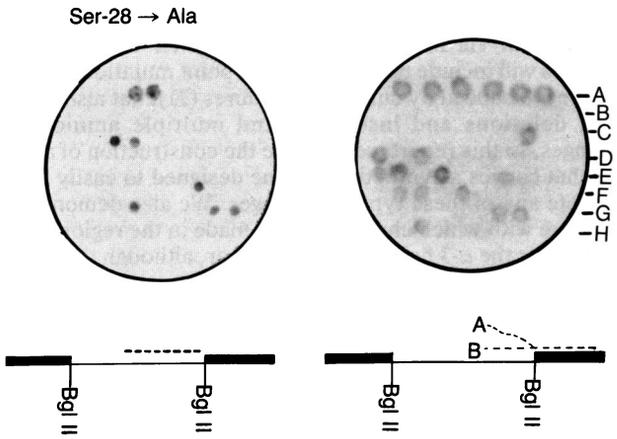

FIG. 5. Identification of strains containing plasmids that carry mutated inserts in the proper orientation. Ampicillin-resistant strains were first screened for the presence of an insert by colony hybridization with a 28-nucleotide probe. (*Left*) Detection of an insert coding for a Ser → Ala change at position 28. The hybridization is shown schematically below the autoradiograph. Clones that were positive for the presence of an insert were then rescreened with a 20-nucleotide probe that spans one of the vector–insert junctions (*Right*). Row A is the positive control (pJN02). Row B is the negative control (pJN03), and rows C–H represent various mutants. Plasmids with inserts in the proper orientation hybridize to the entire probe sequence, while inserts in the opposite orientation allow hybridization to only half of the probe. These two alternative hybridization results are shown below the right autordiograph (strands B and A, respectively).

galactosidase synthesis in the presence of IPTG is due to inhibition by IPTG of the hydrolysis of the chromogenic substrate *o*-nitrophenyl β-D-galactopyranoside used in the assay (data not shown). To demonstrate that β-galactosidase levels can be directly correlated with Cro binding to O_R3, a strain was lysogenized with the phage λ112Δ265prmup-1 $O_R3$rl (11), which carries a mutation in O_R3 that reduces the affinity of Cro for this site. Subsequent induction of this strain transformed with pTR190 failed to turn off β-galactosidase synthesis (data not shown).

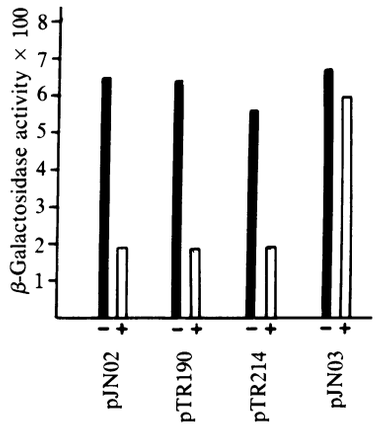

FIG. 6. Comparison of β-galactosidase activity in strains carrying wild-type and synthetic *cro* genes. pJN02 carries the synthetic *cro* gene that was constructed from pTR190, a plasmid that has the wild-type *cro* sequence (12). pTR214 also carries the wild-type sequence but produces higher levels of Cro protein than pTR190. pJN03 carries the synthetic *cro* sequence but has a deletion of the 49-bp fragment coding for the α-3 helix of Cro. Cells were grown in the presence of 1 mM IPTG (+) or without IPTG (−), and β-galactosidase activity was measured as described by Miller (18).

Strains carrying plasmids with specific mutations in the region of *cro* encoding the α-3 helix were also assayed for O_R3 binding (Fig. 7). The 11 mutants, which produce amino acid substitutions at six different positions, can be divided into three classes with respect to their ability to bind to O_R3. The substitutions Tyr-26 → Asp, Gln-27 → Leu, Gln-27 → Cys, Gln-27 → Arg, Ser-28 → Ala, and Ala-33 → Lys allow levels of β-galactosidase activity similar to the level from a pJN03-carrying strain, indicating a markedly diminished ability to bind O_R3. Strains producing Cro proteins that carry the substitutions Tyr-26 → Lys, His-35 → Arg, Ala-36 → Thr, and Ala-36 → Lys show β-galactosidase activity levels that are similar to wild-type Cro produced by pJN02, indicating that these substitutions do not drastically decrease O_R3 binding capacity, although small differences in binding may be masked by saturating levels of Cro. However, an intermediate level of activity is seen in the strain carrying the Tyr-26 → Phe substitution.

DISCUSSION

The recent determination of the structure of λ Cro protein has allowed a detailed model to be proposed for complex formation with operator DNA (10). The model predicts that several amino acids in or near the α-3 helix of Cro make sequence-specific hydrogen bonds and van der Waals contacts. Among the amino acids participating in this site-specific recognition are Tyr-26, Gln-27, and Ser-28. The model proposes that the hydroxyl group of Tyr-26 donates a hydrogen bond to the O4 of a thymine, while the side chain amide of Gln-27 donates a hydrogen bond to the N7 and accepts a hydrogen bond from the N6 of an adenine. In addition, Gln-27 appears to make a van der Waals contact with a thymine methyl group. Two hydrogen bonds are also formed between the side-chain hydroxyl of Ser-28 and an adenine.

Seven of the amino acid substitutions we describe here occur at positions 26, 27, and 28. The substitution of Asp for Tyr-26 decreases the ability of Cro to bind to O_R3 and turn off β-galactosidase synthesis. However, when Tyr-26 is replaced with Phe, an intermediate level of β-galactosidase activity is observed. This substitution replaces the hydroxyl group of tyrosine with a hydrogen atom while preserving the remaining side-chain character. This intermediate level of β-galactosidase activity may represent the loss of a single hy-

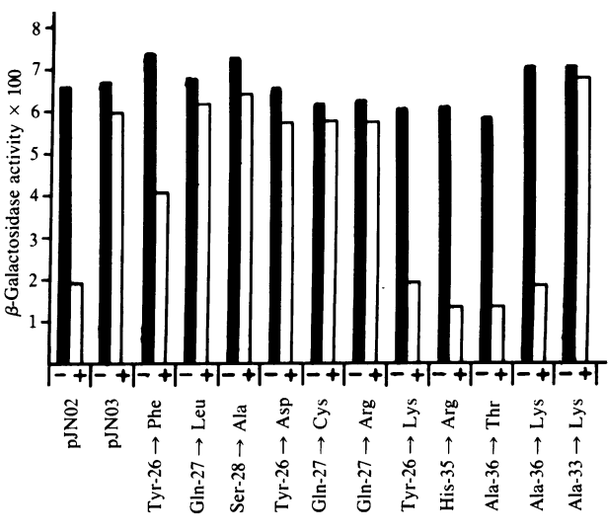

FIG. 7. Comparison of β-galactosidase activity in strains carrying plasmids with specific mutations in the region of *cro* encoding the α-3 helix. The amino acid substitution resulting from each mutation is indicated below the appropriate pair of bars, which are as in Fig. 6.

drogen bond as proposed by the model. Surprisingly, when Tyr-26 is replaced with a Lys, β-galactosidase activity is reduced to a level similar to that found in a strain carrying the wild-type *cro* sequence. Computer modeling of this amino acid substitution indicates that it is possible that an interaction between the Lys side chain and the phosphate backbone replaces the hydrogen bond contributed by Tyr at this position. The substitution of Leu, Cys, or Arg for Gln at position 27 reduces Cro binding and suggests that the glutamine side chain may, in fact, participate in multiple interactions in the major groove of operator DNA. This interpretation is consistent with the conservation of glutamine at analogous positions in other DNA-binding proteins (5, 6). The Ser-28 → Ala substitution replaces the hydroxyl group of serine with a hydrogen atom. In this case the effect on O_R3 binding is more pronounced than that seen with the Tyr-26 → Phe substitution. However, Ser-28 is implicated in a bidentate hydrogen bonding arrangement and this change may result in the loss of two hydrogen bonds.

His-35 is predicted by the model to have a sequence-independent interaction with the phosphate backbone. When Arg is placed at this position little difference in Cro binding can be detected by β-galactosidase activity. Computer modeling of this substitution suggests that Arg may also interact with the phosphate backbone but at a position different from that for His. Two substitutions at position 36, Ala-36 → Thr and Ala-36 → Lys also produce variant Cro proteins capable of turning off β-galactosidase synthesis. The model does not predict a role for Ala-36 in operator complex formation. However, this position is appropriate for making contacts with base pairs either directly, in the case of the Lys substitution, or mediated by a water molecule, in the case of the Thr substitution. Whether or not these contacts are actually made is unknown.

The final amino acid substitution, Ala-33 → Lys, decreases the ability of Cro to bind to O_R3. Ala-33 resides on the hydrophobic face of the α-3 helix, and substituting a charged group at this position may therefore be expected to have a disruptive effect on Cro structure and operator binding. We expect further analysis of the other variant Cro proteins described here to show that these amino acid substitutions do not disrupt Cro structure or the steady-state level of Cro but rather owe any change in O_R3 binding to a specific change in the ability to recognize operator DNA.

It is apparent from comparison with plasmid pTR214 that both pTR190 and pJN02 produce O_R3-saturating levels of Cro under induced conditions. Therefore, it is doubtful whether a variant Cro protein with a higher affinity for O_R3 could be detected in this system under fully induced conditions. Further experiments are necessary to characterize the Cro proteins with amino acid substitutions which result in binding that, by β-galactosidase activity, appears similar to wild-type Cro.

Recently, additional Cro repressor mutants defective in operator binding have been isolated (A. Pakula and R. Sauer, personal communication). Included among this group are mutations which result in the amino acid substitutions Gln-27 → His and Ser-28 → Arg. These results are consistent with our observation that amino acids at these positions are critical to operator recognition and support the proposed model for Cro–operator interaction (10). It is extremely interesting that Gln-27 and Ser-28 are conserved at analogous positions in λ repressor (6–8) and have been proposed to interact in sequence-specific manner (19). It has been found that substitutions at these two positions in repressor substantially reduce operator binding (20).

We perceive that the understanding of the recognition process between Cro protein and operator DNA will be facilitated by the introduction of a large number of changes in Cro protein via mutations in the *cro* gene. Ideally, these changes will include not only specific point mutations, which can be introduced by current procedures (21), but also amino acid deletions and insertions and multiple amino acid changes. In this report we describe the construction of a vector that carries a synthetic *cro* gene designed to easily accomodate any of these types of changes. We also demonstrate the ease with which changes can be made in the region of *cro* coding for the α-3 helix of Cro repressor, although the procedure described here is applicable to any portion of the protein that is encoded by the synthetic DNA fragment.

We thank M. Ptashne for providing the phages used in this study. We also thank R. Sauer, J. Engels, B. W. Matthews, and D. Ohlendorf for helpful discussions, R. Sauer for communication of unpublished results, B. W. Matthews for access to the Cro coordinates and computer graphics facilities, and D. Tronrud for help with computer modeling of amino acid substitutions. This is paper XVIII in the series "Studies on Gene Control Regions." Paper XVII is ref. 22. This research was supported by National Institutes of Health Grant GM21120.

1. McKay, D. B. & Steitz, T. A. (1981) *Nature (London)* **290,** 744–749.
2. Pabo, C. O. & Lewis, M. (1982) *Nature (London)* **298,** 443–447.
3. Anderson, W. F., Ohlendorf, D. H., Takeda, Y. & Matthews, B. W. (1981) *Nature (London)* **290,** 754–758.
4. Matthews, B. W., Ohlendorf, D. H., Anderson, W. F. & Takeda, Y. (1982) *Proc. Natl. Acad. Sci. USA* **79,** 1428–1432.
5. Weber, I. T., McKay, D. B. & Steitz, T. A. (1982) *Nucleic Acids Res.* **10,** 5085–5102.
6. Sauer, R. T., Yocum, R. R., Doolittle, R. F., Lewis, M. & Pabo, C. O. (1982) *Nature (London)* **298,** 447–451.
7. Anderson, W. F., Takada, Y., Ohlendorf, D. H. & Matthews, B. W. (1982) *J. Mol. Biol.* **159,** 745–751.
8. Steitz, T. A., Ohlendorf, D. H., McKay, D. B., Anderson, W. F. & Matthews, B. W. (1982) *Proc. Natl. Acad. Sci. USA* **79,** 3097–3100.
9. 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.
10. Ohlendorf, D. H., Anderson, W. F., Fisher, R. G., Takeda, Y. & Matthews, B. W. (1982) *Nature (London)* **298,** 718–723.
11. Maurer, R., Meyer, B. J. & Ptashne, M. (1980) *J. Mol. Biol.* **139,** 147–161.
12. Roberts, T. M., Kacich, R. & Ptashne, M. (1979) *Proc. Natl. Acad. Sci. USA* **76,** 760–764.
13. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).
14. Caruthers, M. H. (1982) in *Chemical and Enzymatic Synthesis of Gene Fragments*, eds. Gassen, H. G. & Lang, A. (Verlag Chemie, Weinheim, GDR), pp. 71–79.
15. Sgaramella, V. & Khorana, H. G. (1972) *J. Mol. Biol.* **72,** 427–444.
16. Gergen, J. P., Stern, R. H. & Wensink, P. C. (1979) *Nucleic Acids Res.* **7,** 2115–2136.
17. Maxam, A. M. & Gilbert, W. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 560–564.
18. Miller, J. H. (1972) *Experiments in Molecular Genetics* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY), pp. 352–355.
19. Lewis, M., Jeffrey, A., Wang, J., Ladner, R., Ptashne, M. & Pabo, C. O. (1982) *Cold Spring Harbor Symp. Quant. Biol.* **47,** 435–440.
20. Nelson, H. M. C., Hecht, M. H. & Sauer, R. T. (1982) *Cold Spring Harbor Symp. Quant. Biol.* **47,** 441–449.
21. Zoller, M. J. & Smith, M. (1982) *Nucleic Acids Res.* **10,** 6487–6500.
22. Mandecki, W. & Caruthers, M. H. (1985) *Gene*, in press.