

The *Escherichia coli* L-arabinose operon: Binding sites of the regulatory proteins and a mechanism of positive and negative regulation

(DNase protection/cooperativity/*araC* protein/cyclic AMP receptor protein/RNA polymerase)

SHARON OGDEN, DENNIS HAGGERTY*, CAROL M. STONER, DAVID KOLODRUBETZ, AND ROBERT SCHLEIF

Department of Biochemistry, Brandeis University, Waltham, Massachusetts 02254

Communicated by Charles Yanofsky, March 28, 1980

ABSTRACT The locations of DNA binding by the proteins involved with positive and negative regulation of transcription initiation of the L-arabinose operon in *Escherichia coli* have been determined by the DNase I protection method. Two cyclic AMP receptor protein sites were found, at positions -78 to -107 and -121 to -146, an *araC* protein-arabinose binding site was found at position -40 to -78, and an *araC* protein-fucose binding site was found at position -106 to -144. These locations, combined with *in vivo* data on induction of the two divergently oriented arabinose promoters, suggest the following regulatory mechanism: induction of the *araBAD* operon occurs when cyclic AMP receptor protein, *araC* protein, and RNA polymerase are all present and able to bind to DNA. Negative regulation is accomplished by the repressing form of *araC* protein binding to a site in the regulatory region such that it simultaneously blocks access of cyclic AMP receptor protein to two sites on the DNA, one site of which serves each of the two promoters. Thus, from a single operator site, the negative regulator represses the two outwardly oriented *ara* promoters. This regulatory mechanism explains the known positive and negative regulatory properties of the *ara* promoters.

Studies on the L-arabinose operon of *Escherichia coli* have established the following important facts on regulation of the divergently oriented *ara* promoters p_C and p_{BAD} (see Fig. 1). The activity of both promoters is stimulated by the cyclic AMP (cAMP) receptor protein (CRP) in the presence of cyclic AMP (1–4). The promoter p_{BAD} is positively regulated by *araC* protein in the presence of arabinose—i.e., the protein is required for activity of the promoter (1, 2, 5). Under noninducing conditions, the *araC* protein instead acts negatively to repress both the p_C and p_{BAD} promoters (1–3, 6). At least part of the DNA site necessary for repression of p_{BAD} lies upstream from all of the sites necessary for induction of p_{BAD} , as shown by the existence of deletions entering the *ara* regulatory region from the p_C side that abolish repression of p_{BAD} without affecting induction of p_{BAD} (6, 7).

The requisite components are now available for *in vitro* studies of the mechanism of regulation of the *ara* operon. The regulatory region DNA has been isolated, and its nucleotide sequence has been determined (8, 9) and found to contain elements similar to the RNA polymerase-binding sites seen in other *E. coli* promoters at about 10 and 35 bases before the start sites of transcription (8). The sequence also contains several stretches that resemble the CRP-binding site in the *gal* operon (10). Also available are the proteins involved in the regulation of the *ara* operon: *araC* protein (11), CRP (12), and RNA polymerase (13).

In the work reported here, we have utilized the DNA se-

quence determination methodology of Maxam and Gilbert (14) with the DNase I protection method of Galas and Schmitz (15) to determine the DNA-binding sites of the *araC* protein and CRP. These studies, data from a physiological experiment, and consideration of published data suggest the following mechanism for regulation of the operon: under induction conditions, CRP is able to bind to a DNA site to be called CRP_{BAD}. Its presence on the DNA stabilizes the binding of *araC* protein in its inducing conformation adjacent to CRP on a DNA sequence, *araI*. The presence of *araC* protein on the DNA then stimulates the binding of RNA polymerase to the DNA. This induction scenario can be blocked by the presence of *araC* protein in a repressing conformation bound to a DNA site upstream from, but partly obscuring, the CRP_{BAD} site. At the same time and from the same location, *araC* protein also partially represses p_C by blocking access of CRP to a second site, CRP_C, located on the opposite side of *araC* from CRP_{BAD}.

MATERIALS AND METHODS

DNA fragments for sequence determination and protection were obtained from plasmid pMB9*ara*440 (8) by *Eco*RI endonuclease digestion and polyethylene glycol precipitation (16). CRP was purified as described (12), and *araC* protein was purified as described (11), with the substitution of a hydroxyapatite column for the DEAE-Sephadex column. It was loaded in 0.014 M potassium phosphate buffer at pH 6.9, 10% (vol/vol) glycerol, 0.01 M L-arabinose, 1 mM EDTA, and 1 mM dithiothreitol and eluted in the same buffer with the phosphate raised to 0.13 M.

DNA fragments were ³²P end-labeled by alkaline phosphatase treatment and T4 polynucleotide kinase as described by Maxam and Gilbert (14) and in protocols provided by these authors. Size standards for the DNase-protected DNA fragments were G>A sequencing fragments in which the methylation was performed either in the buffer described in the figure legends or in the DNase I digestion buffer.

DNase I protection was performed as described (15), with the exceptions noted here and in the figure legends. DNase I digestion buffer is 50 mM NaCl/20 mM Tris-HCl, pH 8.0/3 mM MgCl₂/0.1 mM dithioerythritol/0.2 mM cAMP. DNA was incubated at 37°C for 10 min in DNase I digestion buffer, then proteins were added and the incubation was continued 10 min to permit the proteins to bind. The temperature was shifted to 20°C and DNase I was added to a final concentration of 0.13 μg/ml. After the addition of stop buffer, the DNA was prepared and separated by electrophoresis.

The strain used in induction measurements of p_C was con-

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.

Abbreviations: cAMP, cyclic AMP; CRP, cAMP receptor protein.

* Present address: Industrial Products Group, Miles Laboratories, Inc., P. O. Box 932, Elkhart, IN 40515

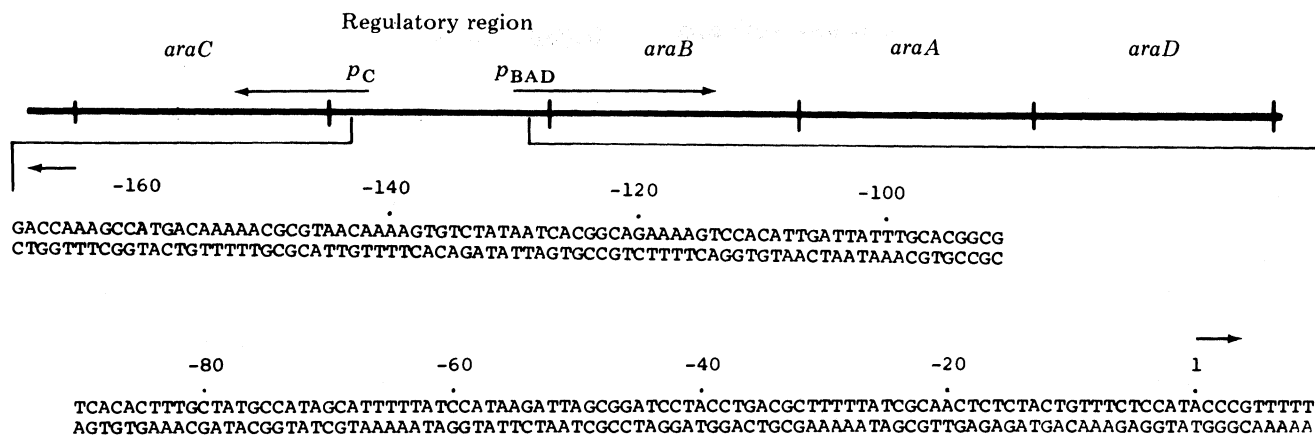


FIG. 1. Structure of the L-arabinose operon, with origins and directions of transcription of the genes, nucleotide sequence, and position notation.

structed by mating the *araB*⁻ episome, F'74 (17), into a leucine-requiring derivative of the *p*_C- β -galactosidase fusion strain FC-17 (3). The resulting strain, F'*C*⁺*B*⁻*A*⁺*D*⁺/*araC-lacZ* *B*⁺*A*⁺*D*⁻, grows on L-arabinose. In the measurements shown, cells were grown overnight in M10 medium (18) containing thiamine at 0.001%, glycerol at 0.2%, and Difco casamino acids at 1%. The cells were diluted 1:1000 and grown to 1×10^8 cells per ml, at which time the experiment began. At the indicated times, 0.35-ml samples were taken into 0.5 ml of assay buffer at 0°C containing 200 μ g of chloramphenicol per ml, and the β -galactosidase was measured as described (18).

RESULTS

Location of the CRP_C and CRP_{BAD} Binding Sites. Binding of a protein to a specific DNA sequence can block or enhance cleavage of the phosphate backbone by DNase I (15). The location of the sites with altered cleavage frequencies is then determined by electrophoresis of the DNase I digestion products: cleavage positions are determined by parallel electrophoresis of the same DNA that has been prepared in accordance with the DNA sequence determination methods developed by Maxam and Gilbert. In such experiments with DNase I, we have located two CRP-binding regions: one between 20 and 50 base pairs from *p*_C, position -121 to -146, and one from 78 to 107 base pairs from *p*_{BAD}, position -78 to -107 (Fig. 2 A and B). In one of the experiments shown (Fig. 2A), almost complete protection of CRP_{BAD} was achieved, and protection and enhancement of cleavage sites in CRP_C were only barely detectable. In the other experiment shown (Fig. 2B), the original data showed more complete protection of CRP_{BAD} than CRP_C, although the reproduced data show the differences less well. From these observations we conclude that the *in vitro* affinity of CRP protein is higher for CRP_{BAD} than it is for CRP_C.

Location of the *araC* Protein Induction and Repression Binding Sites. By similar DNase I experiments, the *araC* protein-binding sites have been located (Fig. 2 C and D). *araC*, in the presence of arabinose, identifies a binding region from positions -40 to -78, and a region less well protected from -106 to -144 (Fig. 2C). When the same DNase I experiments are performed with *araC* protein in the presence of the anti-inducer D-fucose, the region -106 to -144 is protected, and to a lesser extent the region -40 to -78 (Fig. 2D). Thus the sequence from -40 to -78 is identified as the *araI* site, and the region from -106 to -144 is identified as the *araO* site in accordance with the nomenclature of Englesberg (6).

No evidence of other *araC* protein-binding sites, either arabinose- or fucose-dependent, was seen in the region +45 to

-170. The presence of cAMP was irrelevant for C protein-binding in the absence of CRP, but necessary for CRP binding. At less than saturating concentrations of *araC* protein, the presence of CRP-cAMP enhanced the apparent binding of *araC* protein in the presence of arabinose (data not shown).

Induction of the *p*_C Promoter. The binding sites determined above suggested the mechanism of positive and negative regulation of *p*_{BAD} to be discussed in the next section. Additionally, the partial overlapping of *araO* with both CRP_{BAD} and CRP_C suggested that the repression complex of *araC* protein with DNA might repress both the *p*_{BAD} and *p*_C promoters; therefore, induction by arabinose might both induce *p*_{BAD} and derepress *p*_C. Several years ago Casadaban addressed this very question after noting that both *ara* promoters were in the same region (3). He constructed a *p*_C- β -galactosidase fusion strain to enable convenient assay of *p*_C activity, but his measurements failed to reveal any arabinose-mediated activation of *p*_C. On the basis of the DNase I protection findings, we have reexamined this question.

Our initial experiments confirmed Casadaban's findings that the steady-state expression of *p*_C in an *araC*⁺*B*⁺*A*⁺*D*⁺ strain is nearly the same in the presence or absence of arabinose. We therefore explored the response of *p*_C immediately after the addition of arabinose. Fig. 3 shows that *p*_C is indeed derepressed by the addition of arabinose. The initial activity of *p*_C, as demonstrated by the slope of the curve for β -galactosidase level, is about 5-fold higher than the activity just preceding the addition of arabinose. At about 15 min after arabinose addition, by which time the β -galactosidase level has almost doubled, the rate of *p*_C expression falls back to about its preinduction value. Subsequently (data not shown), *p*_C retains this preinduction value so that the level of β -galactosidase per cell is diluted by cell growth back to its preinduction level.

The expression of *p*_C appears to be strongly catabolite sensitive. As shown in Fig. 3, adding glucose to the medium 30 min before adding arabinose drastically reduces the magnitude of the *p*_C derepression. Additionally, as shown, a high concentration of cAMP in the medium reverses the catabolite repression caused by glucose.

DISCUSSION

In *Results* we have presented data that locate the DNA-binding sites of the proteins involved in the joint regulation of the arabinose operon promoters *p*_C and *p*_{BAD}. Determined were the two binding sites of CRP protein, CRP_C and CRP_{BAD}, and the binding sites *araI* and *araO* of *araC* protein in its inducing and repressing states. Additionally, DNase I protection exper-

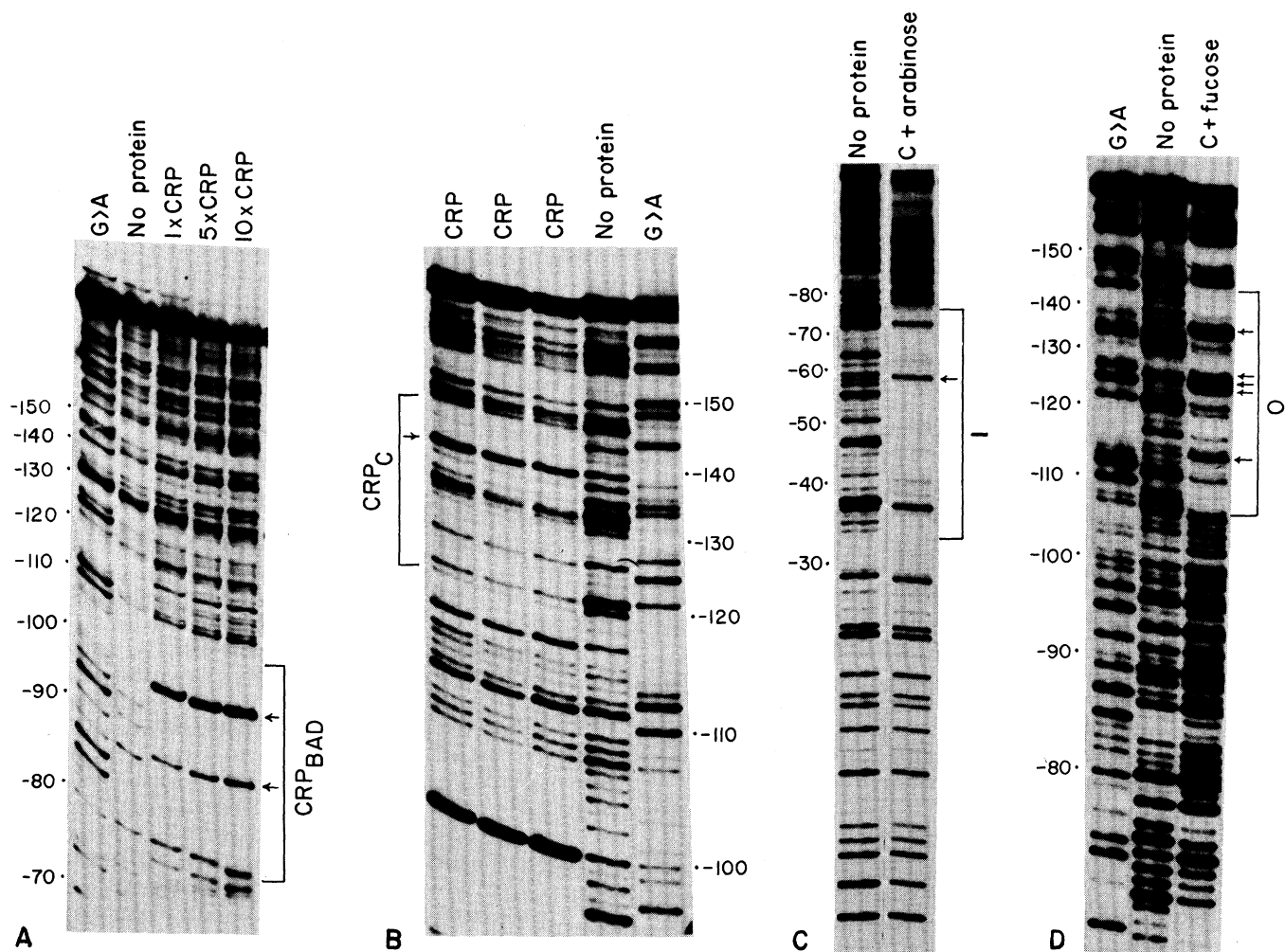


FIG. 2. Locations of CRP_{BAD} , CRP_C , $araI$, and $araO$ by DNase I. Indicated are the regions containing bases with protected or enhanced cleavages. Arrows indicate the bases with cleavage enhanced by the presence of the protein. (A) Protection of CRP_{BAD} from DNase I digestion. Buffer was 10 mM Tris-HCl, pH 7.9/50 mM KCl/10 mM $MgCl_2$ /2.5 mM $CaCl_2$ /0.1 mM dithiothreitol/30% (vol/vol) glycerol/0.2 mM cAMP. DNA was labeled at the *Bam*HI site at position -43. Lane 1 is G>A sequencing size control. For lanes 2-5 DNA was digested with DNase I at 0.13 μ g/ml. Lane 2 is DNase I digestion with no other protein added. Lanes 3-5 are DNase I digestion with CRP present at concentrations of 6.5, 32.5, and 65 μ g/ml. (B) Protection of CRP_C from DNase I digestion. DNA was labeled at the *Bam*HI site at position -43. Lanes 1-3 contained CRP, a different preparation than that used in part A, at 12 μ g/ml; DNA was digested with DNase I at 0.10, 0.12, and 0.2 μ g/ml. Lane 4 contained no other added protein and its DNA was digested with DNase I at 0.13 μ g/ml. (C) Protection of $araI$ from DNase I digestion. DNA was labeled at the *Eco*RI end at position +48. The G>A sizing lane is not shown because it was substantially fainter and required longer exposure. DNA was digested with DNase I at 0.2 μ g/ml. Lane 1 is DNase I with no other added protein and lane 2 contained 2.5 μ l of *araC* protein and 100 mM L-arabinose. (D) Protection of $araO$ from DNase I digestion. DNA was labeled at the *Bam*HI site at position -43. Lane 1 is the G>A sizing standard. DNA was digested with DNase I at 0.1 μ g/ml. Lane 2 contained no other added protein and lane 3 contained 2.5 μ l of *araC* protein and 100 mM D-fucose.

iments performed with *araC* protein plus CRP indicate that the presence of the CRP-cAMP complex enhances the binding of *araC* protein. Similarly, electron microscopy previously indicated that RNA polymerase detectably binds to p_{BAD} only when CRP and *araC* proteins are present (4). The microscopy also showed that CRP-cAMP stimulated the binding of RNA polymerase to the p_C promoter about 3-fold. This information, plus the observation contained in this paper that addition of arabinose induces both p_C and p_{BAD} , suggests the mechanism sketched in Fig. 4: *araC* protein in the repressing state binds to the $araO$ sequence. In this position *araC* blocks access of CRP to the CRP_C and CRP_{BAD} sites located on either side. Thus p_{BAD} remains fully off, but because RNA polymerase retains significant "unaided" activity on the p_C promoter (3), a residual level of *araC* protein synthesis remains. Upon the addition of arabinose, *araC* protein leaves the $araO$ site and permits the entry of CRP molecules to both CRP-binding sites. Such binding of

CRP to the CRP_C sequence stimulates the activity of RNA polymerase on p_C . Similarly, the binding of CRP to the CRP_{BAD} DNA sequence leads to greater occupancy by *araC* protein in its inducing state to a site adjacent to the CRP on the *araI* sequence. In turn, the presence of *araC* protein increases occupancy by RNA polymerase of the promoter DNA and hence stimulates transcription from p_{BAD} .

This mechanism directly explains the facts enumerated in the *Introduction* on regulation of p_{BAD} and p_C : (i) positive and negative control by *araC* of p_{BAD} ; (ii) the CRP requirement for induction of p_{BAD} ; (iii) negative control on p_{BAD} is exerted upstream from positive control; (iv) *araC* protein represses p_C ; and (v) the synthesis of *araC* protein can be stimulated by CRP (3, 4).

The mechanism leads to the prediction that under certain conditions the addition of arabinose to cells should derepress the synthesis of *araC* protein. Furthermore, this derepression

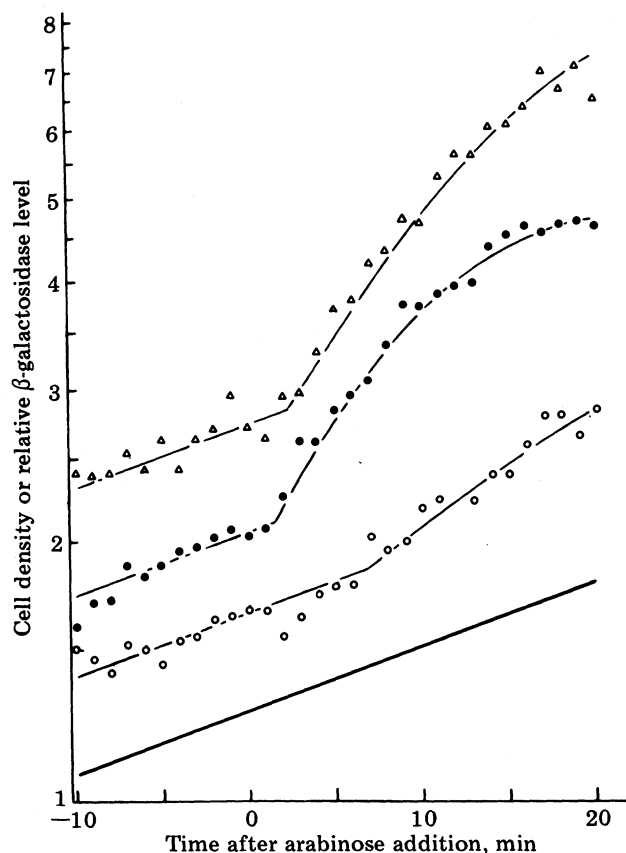


FIG. 3. Activity of the p_C promoter as assayed by β -galactosidase in the p_C - $lacZ$ fusion strain $F^+C^+B^-A^+D^+/p_C$ - $lacZ$ $A^+B^+D^-$. Time is after the addition of arabinose to 0.2% to all cultures. Plotted are cell density (—), β -galactosidase level in the culture with only arabinose added (\bullet — \bullet), β -galactosidase level in the culture with glucose added to 0.2% at -30 min (\circ — \circ), and β -galactosidase level in the culture with glucose added to 0.2% and cAMP added to 5 mM at -30 min (Δ — Δ).

should be cAMP dependent. Indeed, such derepression is observed, as shown in Fig. 3. Unexpectedly, the arabinose-induced derepression of p_C ceases after about 15 min. A likely cause for the shutoff appears to be catabolite repression generated by the catabolism of arabinose. The experiment reported in *Results* showed that CRP-cAMP possesses a substantially lower apparent affinity for CRP_C than for CRP_{BAD}. If this difference is also expressed *in vivo*, then the metabolism of arabinose, which is known to impose partial catabolite repression (19, 20), could effectively catabolite repress p_C back to its $araC$ -repressed activity while having little effect on p_{BAD} . The experiments reported in Fig. 3 also showed that derepression of p_C is indeed highly responsive to catabolite repression and cAMP. Similar differential effects of cAMP on different operons *in vivo* have been observed previously (20, 21).

The mechanism proposed explains a substantial body of additional information that has accumulated on regulation of the arabinose operon. For example, it has been possible to isolate CRP-independent Ara^+ mutants (22, 23). These mutants can grow on arabinose in the absence of CRP, and it has been found that the mutations conferring this property lie in $araC$ protein itself. These can be most easily understood as alterations in the $araC$ protein such that it may bind to the $araI$ site in the absence of CRP—i.e., tighter-binding mutants.

Mutants have been isolated that express the arabinose operon in the absence of $araC$ protein (24, 25). Amongst other possibilities, such mutations could create a new RNA polymerase

binding site adjacent to CRP_{BAD} or convert the $araC$ -binding site, I , to another CRP-binding site. From the relevant sequences (8, 10) such results appear possible with one or two base changes. Mutations affecting $araC$ protein have also been observed that permit expression of the arabinose operon in the absence of the normal inducer, arabinose (26, 27). Paradoxically, these mutations, C^c , are not dominant to the wild-type $araC$ protein. That is, uninduced cells containing both proteins possess basal levels of arabinose enzymes, whereas uninduced C^c mutants possess levels characteristic of wild-type fully induced cells. The dominance of the repressing form of $araC$ protein has long been hard to rationalize; however, it can now be seen to be a natural consequence of the mechanism: occupancy of the repression site blocks induction.

A number of questions relevant to the proposed mechanism remain to be answered. Delineation of the RNA polymerase-binding sites on p_C and p_{BAD} by DNase protection has not yet been completed. We have not physically demonstrated that $araC$ protein bound to the $araO$ site directly blocks the entry of CRP to its binding sites, although this seems probable in light of the apparent overlapping binding regions. Also to be determined in the future are the actual association and dissociation rates for all of the protein-DNA complexes. The relative values of these rates are necessary for determination of the most probable order of protein binding. No evidence has yet been presented that determines whether the cooperativity between the proteins in DNA binding is a result of protein-protein interactions or whether it is mediated via the DNA. It is yet to be explained how CRP can stimulate the activity of RNA polymerase on the lac and gal promoters from different positions with respect to the start of transcription (10, 28), and additionally, on the ara operon, stimulate the binding of $araC$ protein. This apparent plasticity raises the question of whether any protein bound to DNA in the correct position can stimulate the DNA binding of another protein. It has not yet been shown that the sole cause for the shutoff of p_C induction about 15 min after arabinose addition is catabolite repression generated by metabolism of arabinose.

The protection experiments with $araC$ protein showed incomplete selectivity in binding to $araI$ and $araO$. It is possible, then, that arabinose or fucose bound to the protein only slightly alters its affinity for either of the sites. Alternatively, because $araC$ protein is highly labile, it is possible that the purified protein used in these experiments was damaged or sluggish in shifting between inducing and repressing states. The findings presented here raise the possibility that if a cell possessed a great excess of $araC$ protein, a substantial quantity of the protein might always remain in the repressing state. This could prevent induction of the operon. An excess of $araC$ protein could also have a different effect. Experiments presented earlier showed that $araC$ protein is able to bind to the I site in the absence of CRP. Thus an excess of $araC$ protein might instead alleviate the requirement for CRP in induction of p_{BAD} .

Finally, we note that the failure of high-resolution electron microscopy to reveal the CRP-DNA complexes (29) may be attributed to their occurring at low frequency, staining inadequately for definitive identification, or both. Also, the protein-DNA complex observed in the earlier electron microscope work, which was thought to be a repression complex (29), appears more likely to be RNA polymerase bound at p_C , as was later suggested (4). Still unavailable is an explanation for the role of $araC$ protein in the formation of protein-DNA complexes.

We have described a mechanism of gene regulation in which induction occurs only when all of a set of proteins are present and able to bind to a string of sites on DNA. That there exists

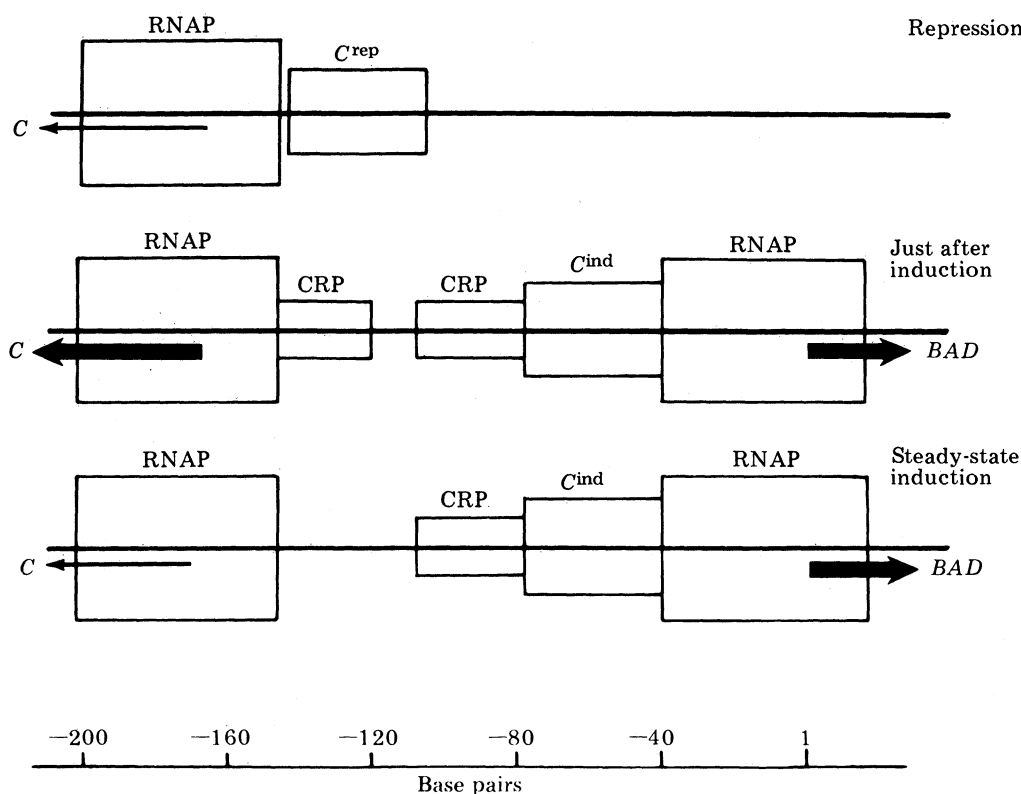


FIG. 4. Proposed mechanism of *ara* operon regulation. Represented is the binding status of the regulatory proteins before, immediately after, and several minutes after induction by the addition of arabinose. The positions of the binding regions are to scale, but not the sizes of the proteins. Also depicted are the activities of the leftward p_C and rightward p_{BAD} promoters in the three conditions. RNAP, RNA polymerase; C^{rep} and C^{ind} , *araC* protein in repressing and inducing conformations.

a requirement for the complete set of proteins necessitates that the binding of the proteins to DNA be cooperative. Such a regulatory mechanism is versatile and easily accommodates situations in which large numbers of inducers must all be present in order for a gene to be induced.

We gratefully acknowledge Allen Maxam for his advice on the sequence determination procedures and Ann Favreau for technical assistance. This work was supported by U.S. Public Health Service Research Grant GM-18277 and Training Grant GM-212 from the National Institutes of Health.

- Greenblatt, J. & Schleif, R. (1971) *Nature (London) New Biol.* **233**, 166–170.
- Wilcox, G., Neuris, P., Bass, R. & Englesberg, E. (1974) *J. Biol. Chem.* **249**, 2946–2952.
- Casadaban, M. (1976) *J. Mol. Biol.* **104**, 557–566.
- Hirsh, J. & Schleif, R. (1977) *Cell* **11**, 545–550.
- Sheppard, D. & Englesberg, E. (1967) *J. Mol. Biol.* **25**, 443–454.
- Englesberg, E., Squires, C. & Meronk, F. (1969) *Proc. Natl. Acad. Sci. USA* **62**, 1100–1107.
- Schleif, R. & Lis, J. (1975) *J. Mol. Biol.* **95**, 417–431.
- Smith, B. & Schleif, R. (1978) *J. Biol. Chem.* **253**, 6931–6933.
- Greenfield, L., Boone, T. & Wilcox, G. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 4724–4728.
- Taniguchi, T., O'Neill, M. & de Crombrughe, B. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 5090–5094.
- Steffen, D. & Schleif, R. (1977) *Mol. Gen. Genet.* **157**, 333–339.
- Eilen, E., Pampeno, C. & Krakow, J. (1978) *Biochemistry* **17**, 2469–2473.
- Burgess, R. & Jendrisak, J. (1975) *Biochemistry* **14**, 4634–4638.
- Maxam, A. & Gilbert, W. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 560–564.
- Galas, D. & Schmitz, A. (1978) *Nucleic Acids Res.* **5**, 3157–3170.
- Lis, J. & Schleif, R. (1975) *Nucleic Acids Res.* **2**, 383–389.
- Schleif, R. (1972) *Proc. Natl. Acad. Sci. USA* **69**, 3479–3484.
- Haggerty, D., Oeschger, M. & Schleif, R. (1978) *J. Bacteriol.* **135**, 775–781.
- Katz, L. & Englesberg, E. (1971) *J. Bacteriol.* **107**, 34–52.
- Lis, J. & Schleif, R. (1973) *J. Mol. Biol.* **79**, 149–162.
- Prisvant, M. & Lazdunski, C. (1975) *Biochemistry* **14**, 1821–1825.
- Heffernan, L., Bass, R. & Englesberg, E. (1976) *J. Bacteriol.* **126**, 1119–1131.
- Heffernan, L. & Wilcox, G. (1976) *J. Bacteriol.* **126**, 1132–1135.
- Englesberg, E., Sheppard, D., Squires, C. & Meronk, F. (1969) *J. Mol. Biol.* **43**, 281–298.
- Gielow, L., Lagen, M. & Englesberg, E. (1971) *Genetics* **69**, 289–302.
- Englesberg, E., Irr, J. & Lee, N. (1965) *J. Bacteriol.* **90**, 946–956.
- Nathanson, N. & Schleif, R. (1975) *J. Mol. Biol.* **96**, 185–199.
- Majors, J. (1977) Dissertation (Harvard Univ., Cambridge, MA).
- Hirsh, J. & Schleif, R. (1976) *Proc. Natl. Acad. Sci. USA* **73**, 1518–1522.