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 *pc* and *pBAD* (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 *pBAD* 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 *pc* and *pBAD* promoters (1–3, 6). At least part of the DNA site necessary for repression of *pBAD* lies upstream from all of the sites necessary for induction of *pBAD*, as shown by the existence of deletions entering the *ara* regulatory region from the *pc* side that abolish repression of *pBAD* without affecting induction of *pBAD* (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 sequence 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 *pc* 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 pBBara-440 (8) by EcoRI 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 di-thiothreitol 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 methylating 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 *pc* was con-

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

* Present address: Industrial Products Group, Miles Laboratories, Inc., P.O. Box 932, Elkhart, IN 46515
structed by mating the $araB^-$ episome, F'74 (17), into a leucine-requiring derivative of the $pcC^-$galactosidase fusion strain FC-17 (8). The resulting strain, $F^+ C^+ B^+ A^+ D^+$ /$\Delta araC-lacZ\ B^+ A^+ D^-$, 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 $1 \times 10^8$ 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 $0^\circ C$ containing 200 $\mu$g of chloramphenicol per ml, and the $\beta$-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 $pcC$, position $-120$ to $-146$, and one from 78 to 107 base pairs from $p_{BAD}$, position $-78$ to $-107$ (Fig. 2A 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. 2C 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 $pcC$ 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 $pcC$ promoters; therefore, induction by arabinose might both induce $p_{BAD}$ and derepress $pcC$. Several years ago Casadaban addressed this very question after noting that both ara promoters were in the same region (3). He constructed a $pcC^{-\beta}$-galactosidase fusion strain to enable convenient assay of $pcC$ activity, but his measurements failed to reveal any arabinose-mediated activation of $pcC$. 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 $pcC$ in an $araC^{-}\ B^+ A^+ D^+$ strain is nearly the same in the presence or absence of arabinose. We therefore explored the response of $pcC$ immediately after the addition of arabinose. Fig. 3 shows that $pcC$ is indeed derepressed by the addition of arabinose. The initial activity of $pcC$, as demonstrated by the slope of the curve for $\beta$-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 $\beta$-galactosidase level has almost doubled, the rate of $pcC$ expression falls back to its preinduction value. Subsequently (data not shown), $pcC$ retains this preinduction level so that the level of $\beta$-galactosidase per cell is diluted by cell growth back to its preinduction level.

The expression of $pcC$ 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 $pcC$ 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 $pcC$ 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<sub>BAD</sub>, CRP<sub>C</sub>, ara<sub>I</sub>, and ara<sub>O</sub> by DNase I digestion. Arrows indicate 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<sub>C</sub> by DNase I digestion. Buffer was 10 mM Tris-Cl, pH 7.8/50 mM KCl/10 mM MgCl<sub>2</sub>/2.5 mM CaCl<sub>2</sub>/0.1 mM dithiothreitol/30% (v/v) glycerol/0.2 mM cAMP. DNA was labeled at the BamHI 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 μg/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 μg/ml. (B) Protection of CRP<sub>C</sub> from DNase I digestion. DNA was labeled at the BamHI site at position -43. Lanes 1-3 contained CRP, a different preparation than that used in part A, at 12 μg/ml; DNA was digested with DNase I at 0.10, 0.12, and 0.2 μg/ml. Lane 4 contained no other added protein and its DNA was digested with DNase I at 0.13 μg/ml. (C) Protection of ara<sub>I</sub> from DNase I digestion. DNA was labeled at the EcoRI 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 μg/ml. Lane 1 is DNase I with no other added protein and lane 2 contained 2.5 μl of ara<sub>C</sub> protein and 100 mM 1-arabinose. (D) Protection of ara<sub>O</sub> from DNase I digestion. DNA was labeled at the BamHI site at position -43. Lane 1 is the G>A sizing standard. DNA was digested with DNase I at 0.1 μg/ml. Lane 2 contained no other added protein and lane 3 contained 2.5 μl of ara<sub>C</sub> protein and 100 mM D-fucose.

Experiments performed with ara<sub>C</sub> protein plus CRP indicate that the presence of the CRP–cAMP complex enhances the binding of ara<sub>C</sub> protein. Similarly, electron microscopy previously indicated that RNA polymerase detectably binds to p<sub>BAD</sub> only when CRP and ara<sub>C</sub> proteins are present. The microscopy also showed that CRP–cAMP stimulated the binding of RNA polymerase to the p<sub>C</sub> promoter about 5-fold. This information, plus the observation contained in this paper that addition of arabinose induces both p<sub>C</sub> and p<sub>BAD</sub>, suggests the mechanism sketched in Fig. 4: ara<sub>C</sub> protein in the repressing state binds to the ara<sub>O</sub> sequence. In this position ara<sub>C</sub> blocks access of CRP to the CRP<sub>C</sub> and CRP<sub>BAD</sub> sites located on either side. Thus p<sub>BAD</sub> remains fully off, but because RNA polymerase retains significant “unaided” activity on the p<sub>C</sub> promoter (3), a residual level of ara<sub>C</sub> protein synthesis remains. Upon the addition of arabinose, ara<sub>C</sub> protein leaves the ara<sub>O</sub> site and permits the entry of CRP molecules to both CRP-binding sites. Such binding of CRP to the CRP<sub>C</sub> sequence stimulates the activity of RNA polymerase on p<sub>C</sub>. Similarly, the binding of CRP to the CRP<sub>BAD</sub> DNA sequence leads to greater occupancy by ara<sub>C</sub> protein in its inducing state to a site adjacent to the CRP on the ara<sub>I</sub> sequence. In turn, the presence of ara<sub>C</sub> protein increases occupancy by RNA polymerase of the promoter DNA and hence stimulates transcription from p<sub>BAD</sub>.

This mechanism directly explains the facts enumerated in the Introduction on regulation of p<sub>BAD</sub> and p<sub>C</sub>: (i) positive and negative control by ara<sub>C</sub> of p<sub>BAD</sub>; (ii) CRP requirement for induction of p<sub>BAD</sub>; (iii) negative control on p<sub>BAD</sub> is exerted upstream from positive control; (iv) ara<sub>C</sub> protein represses p<sub>C</sub>; and (v) the synthesis of ara<sub>C</sub> 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 ara<sub>C</sub> protein. Furthermore, this derepression
should be cAMP dependent. Indeed, such derepression is observed, as shown in Fig. 3. Unexpectedly, the arabinose-induced derepression of pC ceased after about 15 min. A likely cause for this shutdown appears to be catabolic 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-R. If this difference is also expressed in vivo, then the metabolism of arabinose, which is known to impose partial catabolic repression (19, 20), could effectively catabolically repress pC back to its araC-repressed activity while having little effect on pBAD. The experiments reported in Fig. 3 also showed that derepression of pC 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-R 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^*, 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^* 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 pC and pBAD 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 shutdown of pC 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 proteins 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 pBAD.

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 pC, 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
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.