THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 253, No. 19, Issue of October 10, pp. 6931-6933, 1978
Printed in U.S.A.

Nucleotide Sequence of the L-Arabinose Regulatory Region of *Escherichia coli* K12*

(Received for publication, March 16, 1978)

Brenda R. Smith and Robert Schleif

From the Department of Biochemistry, Brandeis University, Waltham, Massachusetts 02154

The nucleotide sequence of a 250-base pair segment of L-arabinose operon DNA containing the 150-base pair regulatory region has been determined. This segment includes the promoter for rightward *ara*BAD transcription, p_{BAD}, the promoter for leftward *ara*C transcription, p_C, and sites responsible for repression.

The L-arabinose operon of *Escherichia coli* is subject to both positive and negative regulation (1-3). The presence of arabinose stimulates synthesis of the arabinose catabolic enzymes B, A, and D several hundredfold by increasing the transcription initiation frequency at the promoter p_{BAD} (4, 5). This stimulated synthesis requires *ara*C protein, the catabolite activator protein CAP with cyclic AMP, and RNA polymerase. In the absence of arabinose or in the presence of the structural analogue D-fucose, *ara*C protein represses the activity of p_{BAD} by acting at a site upstream from p_{BAD} (6). The promoter for synthesis of *ara*C mRNA p_C, is also in the *ara* regulatory region, but it is oriented in the opposite direction to p_{BAD} (7) (Fig. 1). *In vitro* transcription experiments and high resolution electron microscopy studies have indicated that *ara*C and *ara*BAD mRNA begin at sites that are about 150 base pairs apart (8, 9). Synthesis of *ara*C messenger is stimulated by CAP and it is repressed by *ara*C protein itself.

EXPERIMENTAL PROCEDURES

Enzymes—Restriction nuclease *Hae* III (*Haemophilus aegyptius*), *Hha* I (*Haemophilus haemolyticus*), *Eco* RI (*Escherichia coli* RY13), *Bam* HI (*Bacillus amyloliquefaciens* H), and T4 polynucleotide kinase were purchased from New England Biolabs. The *Thermoplasma acidophilum* restriction enzyme *Tac* I which cleaves CGCG was a gift from David McConnell from Harvard University. Bacterial alkaline phosphatase was purchased from Worthington.

Source of the DNA and Sequencing Strategy—DNA fragments for sequencing were obtained from the plasmid pMB9*ara*440 and a λ*para*BAD phage. The phage was isolated as described by Lis and Schleif (10). The *ara* DNA on the plasmid is the central 440-base pair fragment containing the *ara* regulatory region that is produced by *Hae* III digestion of a 1000-base pair *ara* DNA fragment originating from the heteroduplex 114/116 (8, 11). This DNA was ligated into the plasmid pMB9 so as to recreate *Eco* RI cleavage sites at each end (12). All manipulations were performed using P1-EK1 containment procedures in accordance with the National Institutes of Health guidelines.

DNA sequencing by the Maxam-Gilbert technique (13) was inward from both *Eco* RI cleavage sites (Fig. 3*b*)[1] and outward in both directions from the *Bam* HI and *Tac* I sites. Sequence information could be unambiguously read for about 110 base pairs from each position. The sequences at the *Bam* HI and *Tac* I sites were confirmed by sequencing across them from external sites. Except for the region from the *Tac* I site toward the *ara*C gene, both strands of the 250-base pair segment were sequenced.

RESULTS

Using the Maxam-Gilbert DNA sequencing methods, we have sequenced a 250-nucleotide segment containing the L-arabinose regulatory region (Fig. 2). It includes the promoters for *ara*C and *ara*BAD synthesis, p_C and p_{BAD}, as well as the site or sites responsible for repression of each promoter (8).

Parts of the two promoters may be identified from previous work on the *ara* operon which approximately locates the promoters and by comparison to the sequences known for other promoters (14-18). High resolution electron microscopy located complexes of RNA polymerase on the DNA to an overall precision of ±20 base pairs (8, 9). *In vitro* transcription located the start sites of mRNA from p_{BAD} with precision ±3 base pairs and the start site from p_C with precision ±30 base pairs. Finally, the sequence following p_{BAD} closely agrees with the sequence of the 5′ end of *ara*BAD messenger (19).

Pribnow (14) observed that promoters often contain sequences similar to $TAT^{A}_{G}ATG$ located 6 to 8 residues before the start of transcription. The Pribnow sequence TCTACTG precedes the *ara*C mRNA by 6 base pairs and the sequence TCTACTG precedes the *ara*BAD mRNA by 10 nucleotides. The sequence ACACTTT has been found to occur about 30

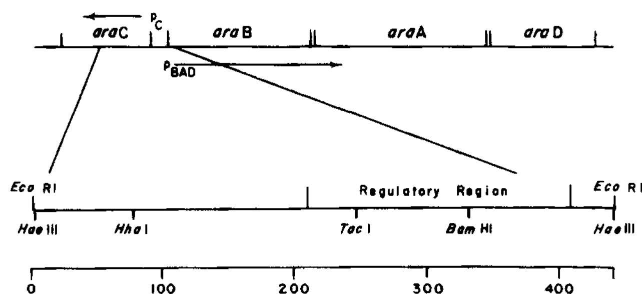

FIG. 1. The L-arabinose operon. Transcription of *ara*B, *ara*A, and *ara*D proceeds rightward from the promoter p_{BAD} and transcription of *ara*C proceeds leftward from the promoter p_C. The fragment derived from *Eco* R1 digestion of pMB9*ara*440 is shown below with the approximate location in base pairs of the *Hae* III, *Hha* I, *Tac* I, and *Bam* HI cleavage sites.

* This work was supported by Public Health Service Research Grant GM-18277 and Training Grant GM-212 from the National Institutes of Health. This publication is No. 1212 of the Department of Biochemistry, Brandeis University. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

[1] Portions of this paper (including additional details of the sequencing and Figs. 3, 4, and 5) are presented in miniprint at the end of this paper. Full size photocopies are available from the Journal of Biological Chemistry, 9650 Rockville Pike, Bethesda, Md. 20014. Request Document No. 78M-417, cite author(s) and include a check or money order for $1.20 per set of photocopies.

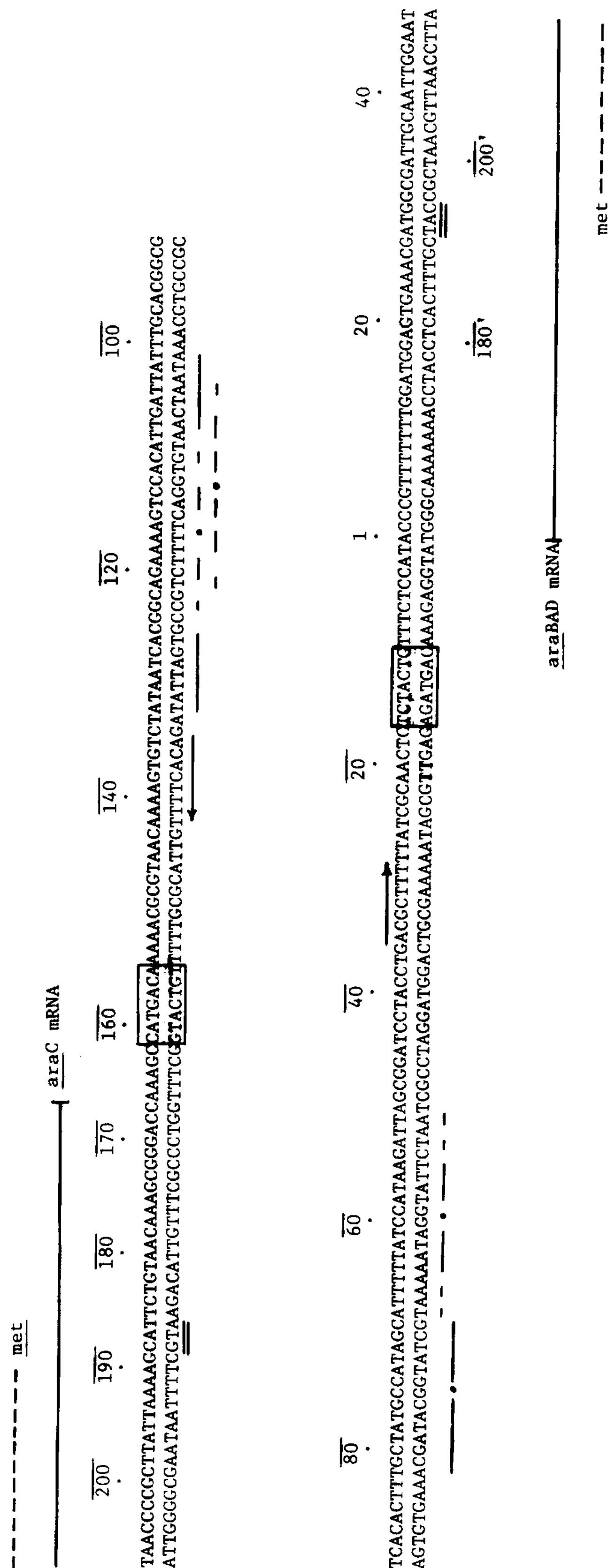

FIG. 2. Nucleotide sequence of the *ara* regulatory region. The start of *ara*BAD messenger is designated position +1. The RNA polymerase recognition sequence ACACTTT is indicated by *arrows* and palindromes are indicated by *lines under the sequence*. The p_C and p_{BAD} Pribnow sequences (14) are *enclosed in a box*. The *ATG* trinucleotide corresponding to the start of *ara*B and the likely start of *ara*C protein synthesis is *underlined twice*.

nucleotides before the start of mRNA in the *lac* and *gal* operons (15, 16) and similar sequences are found in corresponding positions in front of p_C and p_{BAD}. Such sequences could occur by chance in stretches of several hundred nucleotides although it is unlikely that their location should at the same time correspond to that in the other promoters. Thus, we feel these are associated with the promoters p_C and p_{BAD}. There exists another ACACTTT sequence beginning at $\overline{88}$ and a sequence similar to the Pribnow sequence beginning at $\overline{57}$. Any significance of these is as yet unknown.

The *lac* and *gal* CAP binding sites are similar in that they both contain 16-base pair hyphenated palindromes centered about 60 bases before the mRNA start, are A-T-rich, and display sequence homology (15, 16). Four sizeable hyphenated palindromes exist in the *ara* regulatory region. They are centered at 51.5 and 59.5 base pairs ahead of p_C and 59.5 and 75 base pairs ahead of p_{BAD}. They possess 20/32, 10/18, 12/18, and 14/15 palindromic bases, respectively. These sequences are not obviously homologous to the *lac* and *gal* CAP binding site sequences and therefore are not demonstrably CAP binding sites in *ara*.

Although the sequence analysis of the *ara* regulatory region has located parts of RNA polymerase interaction sites and precisely determined the distance between the two parameters, it does not yet provide a clear picture of how the operon is regulated. Comparison of this sequence with others suggests that regulatory proteins and RNA polymerase can interact in a variety of ways and many regulatory region sequences must be analyzed before reliable predictions of mechanisms can be made.

Acknowledgments—We thank Alan Maxam and Forest Fuller for assistance with the sequencing and construction of the plasmid and Steve Kaplan for his early work on sequencing.

REFERENCES

1. Sheppard, D. E., and Englesberg, E. (1967) *J. Mol. Biol.* **25,** 443–454
2. Greenblatt, J., and Schleif, R. (1971) *Nature New Biol.* **233,** 166–170
3. Wilcox, G., Meuris, P., Bass, R., and Englesberg, E. (1974) *J. Biol. Chem.* **249,** 2946–2952
4. Hirsh, J., and Schleif, R. (1973) *J. Mol. Biol.* **80,** 433–444
5. Sheppard, D. G., and Englesberg, E. (1967) *J. Mol. Biol.* **25,** 443–454
6. Englesberg, E., Squire, C., and Meronk, F., (1969) *Proc. Natl. Acad. Sci. U. S. A.* **62,** 1100–1107
7. Wilcox, G., Boulter, J., and Lee, N. (1974) *Proc. Natl. Acad. Sci. U. S. A.* **71,** 3635–3639
8. Hirsh, J., and Schleif, R. (1977) *Cell* **11,** 545–566
9. Hirsh, J., and Schleif, R. (1976) *Proc. Natl. Acad. Sci. U. S. A.* **73,** 1518–1522
10. Lis, J. T., and Schleif, R. (1975) *J. Mol. Biol.* **95,** 395–407
11. Lis, J. T., and Schleif, R. (1975) *J. Mol. Biol.* **95,** 409–416
12. Backman, K., Ptashne, M., and Gilbert, W. (1976) *Proc. Natl. Acad. Sci. U. S. A.* **73,** 4174–4178
13. Maxam, A., and Gilbert, W. (1977) *Proc. Natl. Acad. Sci. U. S. A.* **74,** 560–564
14. Pribnow, D. (1975) *Proc. Natl. Acad. Sci. U. S. A.* **72,** 784–788
15. Dickson, R. C., Abelson, J., Barner, W. M., and Reznikoff, W. S. (1975) *Science* **187,** 27–35
16. Musso, R., DiLauro, R., Rosenberg, M., and DeCrombrugghe, B. (1977) *Proc. Natl. Acad. Sci. U. S. A.* **74,** 106–110
17. Sekiya, T., Gait, M. J., Noris, K., Ramamoorthy, B., and Khorana, H. G. (1976) *J. Biol. Chem.* **251,** 4481–4489
18. Bennet, S. N., Schweingruber, M. E., Brown, K. D., Squires, C., and Yanofsky, C. (1976) *Proc. Natl. Acad. Sci. U. S. A.* **73,** 2351–2355
19. Lee, N., and Carbon, J. (1977) *Proc. Natl. Acad. Sci. U. S. A.* **74,** 49–53

Additional references are found on page. 6933.

Supplemental Material

to

Nucleotide Sequence of the L-Arabinose Regulatory Region of Escherichia coli K12

by

Brenda R. Smith and Robert Schleif

EXPERIMENTAL PROCEDURES

Materials - Carrier free $^{32}P_i$ was purchased from New England Nuclear, dimethyl sulfate (99%) from Aldrich and hydrazine from Eastman Organic Chemicals. Other chemicals are from commercial sources. E. coli DNA polymerase was obtained from Boehringer Mannheim and T_4 DNA ligase from New England Biolabs. Crude pancreatic deoxyribonuclease (DN-25) was from Sigma Chemical Co. Media, lambda suspension buffer, and bacterial strains F'303 and F'311 ($araB^-D^-$) have been described previously (1-3). Strains RFS1465 and 1485 are reported in Kaplan and Schleif (1975) (5).

Isolation of λparaBAD - Cultures of RFS1368 (Hfr $araD^-_{54}$ Δattλ with λcI857S7 inserted into araC) were induced at 42°C, grown for three hrs, and the supernatants plated on RFS1475 (F⁻ $araB^-_{14}$ trp_{amb} Su_{III} ΔlacZ str $\phi 80y^r$) on tet ara plates (4). Ara^+ lysogens, purified from the centers of the ara^+ plaques, were grown overnight in yeast tryptone broth, heat-induced, grown for three hrs and the lysate used to transduce RFSF'303 and F'311 ($araB^-D^-$) to ara^+. The ara^+ colonies were combined, grown in yeast tryptone, heat induced and the lysate was plated on BRS56 (F'$araC^+$/ΔaraBAD). λparaBAD was obtained from a purified ara^+ plaque.

Preparation of Phage DNA - Isolated plaques of λparaBAD were used to make plate lysates on strain RFS1485 (F⁻ Δara-leu Su_{III}) (5). This lysate was used to infect 50 mls of ligase overproducing strain RFS1465 (F': thi str lig^8 ton) that had been made resistant to an airborne phage contaminant of the laboratory. The cells had been grown in YT medium to 3 x 10^8/ml and $MgCl_2$ was added to 0.01 M before the addition of phage at a multiplicity of infection of five. Three hours after infection the culture was lysed with $CHCl_3$ and debris removed by centrifugation. These phage were used to infect 600 mls of 1465 grown as above to 2 x 10^8/ml in YT + 0.4% maltose at a multiplicity of 15. Cells were harvested and phage were purified as described by Kaplan and Schleif (5). Phage DNA was extracted with phenol and dialyzed against TE buffer.

Isolation of Plasmid pRFSara 440 - Plasmid pmB9 (6) DNA was prepared by isopicnic banding in CsCl ethidium bromide (7) and freed of ethidium bromide by five extractions with equal volumes of n-butanol. It was then passed over a 1.25 x 30 cm agarose A5M column and dialyzed into 0.01 M Tris-HCl, pH 7.9 0.001 M NaEDTA, TE. The ara 440 base pair DNA fragment containing the regulatory region was prepared by Hae III digestion of the ara 1000 base DNA fragment that is produced by nuclease S1 digestion of DNA heteroduplexes between phage λpara 114 and λpara 116 as described previously (8). This DNA was phenol extracted, ethanol precipitated, and resuspended in TE. The plasmid DNA was cut with Eco R1 and the sticky ends filled in with E. coli DNA polymerase I. The DNA, 20 μg, was ethanol precipitated and resuspended in 50 μl of 50 mM Tris-HCl, pH 7.8, 5 mM $MgCl_2$, 1 mM β-mercaptoethanol, 50 μg/ml bovine serum albumin. Then 1 μl each of dATP, dTTP, dGTP, dCTP at 5 mM were added followed by 1 μl of DNA polymerase at three units/μl (9). The solution was incubated 40 min at 15°, then phenol extracted, ether extracted, ethanol precipitated and stored in TE. To join the plasmid to the ara 440 base pair fragment, 200 ng of plasmid and 200 ng of Hae III digested fragment were ethanol precipitated and resuspended in 5 μl of ligase buffer, 20 mM Tris-HCl, pH 7.5, 10 mM $MgCl_2$, 10 mM dithiothreitol, 60 μM ATP, and 500 units of T4 DNA ligase were added in 0.5 μl (10). This was incubated 30 min at 20° and diluted to 30 μl with ligase buffer and incubated another 60 min. Next, the DNA was ethanol precipitated and resuspended in TE. Transformation of strain 294 (11) with this sample yielded about 2000 tetracycline resistant colonies. Six were purified, grown, and their plasmids purified. One of the six plasmids was detectably larger upon electrophoresis on 0.7% agarose gel (12). This plasmid was shown to contain a single copy of the Hae III ara 440 base pair fragment by release of this fragment by Eco R1 digestion and by release of a 700 base pair fragment by cleavage at the Bam HI site within the ara 440 fragment and the Bam HI site within the plasmid.

Preparation of Primary Restriction Fragments - In each of two tubes, λparaBAD DNA (0.5 mg) was digested with Bam HI (45 units) at 37°C in 2 ml of 6 mM Tris-HCl, pH 8, 10 mM $MgCl_2$, 2 mM β-mercaptoethanol. After 19 hrs, 15 units of Bam HI and 20 μl of 200 mM β-mercaptoethanol were added and the reaction continued for 6.5 hrs. The contents of the tubes were combined, ethanol precipitated, and loaded onto a preparative (0.6 x 10.2 x 15 cm) 1% agarose, 40 mM Tris-Ac, pH 7.9, 5 mM NaAc, 1 mM EDTA gel. The fragments containing arabinose operon DNA (0.067 and 0.019 lambda units) (12) were cut out, and the DNA extracted by freezing and squeezing out the liquid from the agarose slabs. The DNA was ethanol precipitated. The ara 440 fragment isolated by Eco R1 digestion and electrophoretic separation from the plasmid as described above (2 μg) was cut with Tac I in 13 μl of 8 mM Tris-HCl, pH 7.4, 0.8 mM EDTA, 8 mM $MgCl_2$ and 0.8 mM dithiothreitol at 60°C for 15 min. The reaction was diluted with 0.3 ml of 10 mM Tris-HCl, pH 7.4, 1 mM EDTA and extracted with an equal volume of saturated phenol three times, and with an equal volume of ether twice, and ethanol precipitated. (Fig. 3).

Dephosphorylation of Primary Restriction Fragments - E. coli alkaline phosphatase from Worthington (100 μl) was diluted with 400 μl of 0.1 M Tris-HCl, pH 8, 0.1 mM NaN_3 and dialyzed against the same buffer at 2°C for 20 hrs. The 5'-terminal phosphates were removed by incubating the DNA (2-10 μg) with the dialyzed enzyme (20 m units) in 40-50 μl of 50 mM Tris-HCl, pH 8, 0.1 mM EDTA for one hr at 37°C. The DNA was diluted with H_2O to 0.3 mls, phenol extracted three times, ether extracted twice, ethanol precipitated and resuspended in 75 μl of kinase denaturation buffer, 10 mM glycine-NaOH, pH 9.5, 1 mM spermidine, 0.1 mM EDTA.

Phosphorylating the 5'-OH with (γ-^{32}P) ATP - (γ-^{32}P) ATP, >1000 Ci/mmole was synthesized as described in Maxam and Gilbert (1977) (13). DNA in kinase denaturation buffer was heated at 100°C for three min and quick chilled in a ice water bath. Next, 10 μl of 500 mM glycine-NaOH, pH 9.5, 100 mM $MgCl_2$, 5 mM dithiothreitol and 5 μl of (γ-^{32}P) ATP (150 pmole was lyophilized and resuspended in 5 μl of H_2O) was added and the reaction mixed. T_4 polynucleotide kinase (six units) was introduced and the reaction incubated for three hrs at 37°C. The phosphorylation was stopped using 100 μl of 4 M ammonium acetate, 10 μg tRNA and the DNA was ethanol precipitated.

Cleavage of 5' (-^{32}P) labeled DNA: Secondary Restriction Cleavage

a) Bam HI/Hae III - The kinased 0.067 Bam HI fragment (2-5 μg) was resuspended in 30-40 μl of 6 mM NaCl, 6 mM Tris-HCl, pH 7.4, 6 mM $MgCl_2$, 6 mM β-mercaptoethanol. Hae III (20 units) was introduced and the reaction was incubated for one hr at 37°C.

b) Bam HI/Hha I - The kinased 0.019 Bam HI fragment was dissolved in 30-40 μl of Hha I buffer (6mM Tris-HCl, pH 7.4, 50 mM NaCl, 6 mM $MgCl_2$, 6 mM β-mercaptoethanol) and Hha I (10 units) was added and the reaction was heated for one hr at 37°C.

c) Eco RI - Tac I/Hha I, Bam HI - The Tac I cut Eco RI fragment (2 μg) was dissolved in 30-40 μl of Hha I buffer and digested with Bam HI (30 units) and Hha I (20 units) for one hr at 37°C.

Following the digestion with the respective restriction enzyme(s), the fragments were separated on a 5% polyacrylamide gel (10 mM Tris-borate, pH 8.3, 1 mM EDTA) and the appropriate bands excised. DNA was recovered as described in Maxam and Gilbert (1977). (Fig. 4).

Dimethyl Sulfate Analysis and Hydrazine Analysis - Fragments labeled with (^{32}P) at the 5' terminus were sequenced and the cleavage products fractionated according to the method of Maxam and Gilbert (1977) except that the concentration of carrier DNA was reduced ten-fold to 1 mg/ml. Each fragment was divided into four reactions (1) guanine cleavage: DNA was partially methylated with dimethyl sulfate and heated at 90°C in the presence of piperidine. (2) strong adenine/weak guanine cleavage: The dimethyl sulfate treated DNA was depurinated in 20 mM HCl at 0°C and then heated to 90°C in alkali. (3) thymine and cytosine cleavage: DNA was degraded by partial hydrazinolysis and heated at 90°C in the presence of piperidine. (4) cystosine cleavage: DNA was degraded by hydrazinolysis in the presence of about 2 M NaCl and heated to 90°C with piperidine. (Fig. 5).

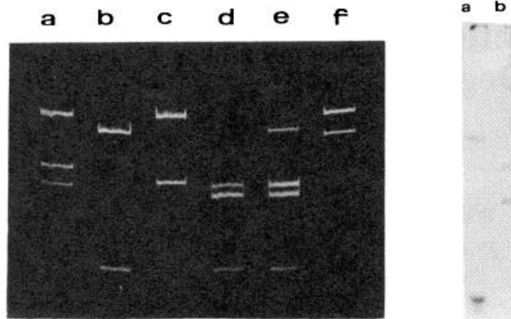

Fig. 3   Fig. 4

Fig. 3. Restriction enzyme cleavage map of ara 440 fragment. The ara 440 fragment was partially digested with Bam HI, Tac I, and/or Hha I in 10 mM Tris-HCl, pH 7.4, 10 mM $MgCl_2$, 1 mM dithiothreitol for one hr at 37°C. a) partial digestion by Tac I producing ara 440, 245 and 195 base pair fragments; b) digestion by Hha I producing 370 and 70 base pair fragments; c) no enzyme control and size standards containing uncut ara 440 and 203 base pair fragment of lac DNA; d) partial digestion by Bam HI, Tac I and Hha I producing 370, 350, 195, 175, 110, 90 and 70 base pair fragments; e) combined Tac I and Hha I digestion producing 370, 350, 195, 175 and 70 base pair fragment; f) partial digestion by Bam HI producing 440, 350 and 90 base pair fragments.

Fig. 4. Autoradiograph of DNA labeled at single 5' terminus with ^{32}Phosphate. Dephosphorylated Bam HI fragments from λparaBAD were kinased using (γ-^{32}P) ATP (>1000 Ci/mmole), digested with the appropriate restriction enzyme, and the fragments resolved on a 5% polyacrylamide as described in experimental procedures. a) the 0.019 Bam HI fragment was digested by Hha I to produce 280 and 7 base pair labeled cleavage products; b) the 0.067 fragment was digested with Hae III to yield labeled 150 and 90 base pair products. The ara regulatory region DNA is contained in the Hha 1270 and Hae III 90 base pair fragments.

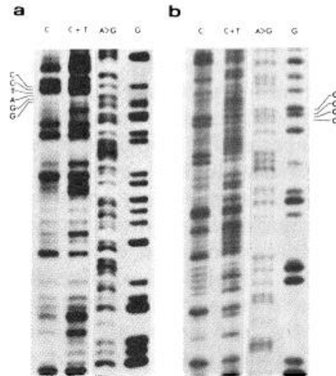

Fig. 5. Hydrazine and dimethyl sulfate analysis of the Bam HI and Tac I cleavage sites in the ara regulatory region DNA. In accordance with the DNA sequencing method developed by Maxam and Gilbert (13), hydrazinolysis was performed in the presence of salt for cytosine residues, 'C', and in the absence of salt for pyrimidine, 'C + T'. Dimethyl sulfate treated DNA was depurinated under conditions allowing fast release of adenine and slow release of guanine, 'A > G', and in the presence of base, specific for guanine 'G'. Products were fractionated by polyacrylamide gel electrophoresis as described in the reference; a) ara 440 DNA was kinased with (γ-^{32}P) ATP, cleaved with Hha I, and the purified 370 bp fragment sequenced. The region presented extends from +6 (bottom) through the Bam HI site, 'GGATCC', at -47 (top); b) 5'-3nd of the 0.019 Bam HI fragment was labeled with (γ-^{32}P) ATP by T_4 polynuclease kinase, digested with Hha I and the 280 base pair fragment obtained was sequenced. The bottom of the gel starts at -95 and extends through the Tac I site, CGCG, at -150 near the top.

REFERENCES

1. Schleif, R. (1969) J. Mol. Biol. 46, 185-196.
2. Lis, J.T. and Schleif, R. (1973) J. Mol. Biol. 79, 149-162.
3. Schleif, R. (1972) Proc. Nat. Acad. Sci. 69, 3479-3484.
4. Lis, J.T. and Schleif, R. (1975) J. Mol. Biol. 95, 395-407.
5. Kaplan, S. and Schleif, R. (1975) J. Bact. 124, 1403-1410.
6. Maniatis, T., Kee, S.G., Efstratiadis, A. and Kafatos, F.C. (1976) Cell 8, 163-182.
7. Tanaka, T. and Weisblum, B. (1975) J. Bact. 121, 354-362.
8. Lis, J.T. and Schleif, R. (1975) J. Mol. Biol. 95, 409-416.
9. Richardson, C.C., Schildkraut, C.L., Aposhian, H.V. and Kornberg, A. (1964) J. Biol. Chem. 239, 222-232.
10. Weiss, B. Jacquemin-Sablon, A., Live, T.R., Fareed, B. and Richardson, C., (1968) J. Biol. Chem. 243, 4543-4555.
11. Backman, K., Ptashne, M. and Gilbert, W. (1976) Proc. Nat. Acad. Sci., 73, 4174-4178.
12. Haggerty, D. and Schleif, R. (1976) J. of Virology 18, 659-663.
13. Maxam, A. and Gilbert, W. (1977) Proc. Nat. Acad. Sci. 74, 560-564.