J. Mol. Biol. (1990) **215**, 497–510

Characterization of the *Escherichia coli ara*FGH and *ara*J Promoters

William Hendrickson[1,2]†, Carol Stoner[2] and Robert Schleif[2,3]

[1]*Department of Microbiology and Immunology*
The Pennsylvania State University College of Medicine
Hershey, PA 17033, U.S.A.

[2]*Department of Biochemistry*
Brandeis University, Waltham, MA 02254, U.S.A.

[3]*Biology Department*
Johns Hopkins University
34th and Charles St, Baltimore, MD 21218, U.S.A.

(Received 28 February 1990; accepted 11 June 1990)

The identities of two cloned, arabinose-inducible promoters were tested by hybridizing promoter DNA fragments with restriction digests of chromosomal DNA containing Mud*lac* phage inserted in either *ara*FGH or in *ara*E transport operons. One promoter, thought to be *ara*E, is within 10^3 base-pairs of a Mud*lac* insertion in the *ara*E gene. The second promoter was not found within several thousand base-pairs of either of the known transport genes. This promoter is now named *ara*P_J (*ara*J).

The DNA sequence of the fragment containing the *ara*FGH promoter was determined. The start site of transcription *in vivo* was located to within ± 1 base-pair (bp) by S_1 nuclease mapping. DNase I footprinting revealed that, in comparison with the *ara*BAD and *ara*E promoters, the locations of the AraC and cyclic AMP receptor protein (CRP) binding sites are reversed with CRP lying between AraC and RNA polymerase. The central location of the CRP binding site may explain why the *ara*FGH promoter is more catabolite sensitive than the other *ara* promoters. AraC and CRP were both required for maximal transcription *in vitro*, although a low level of transcription was detected with CRP alone.

S_1 nuclease mapping of mRNA–DNA hybrids from the *ara*J promoter located the transcription start point to within ± 3 bp, and demonstrates that the promoter is dependent upon AraC protein and CRP *in vivo*. DNase footprinting showed that the location of the AraC protein binding site on *ara*J is adjacent to the RNA polymerase site, as seen at the *ara*BAD and *ara*E promoters. Two CRP sites were observed; one is upstream from the AraC site and one is downstream from the transcription start site.

1. Introduction

Three arabinose-inducible operons of *Escherichia coli*, *ara*BAD, *ara*E and *ara*FGH, have been identified on the basis of genetic and physiological evidence (Sheppard & Englesberg, 1967; Schleif, 1969; Clark & Hogg, 1981; Brown & Hogg, 1972; Kolodrubetz & Schleif, 1981*c*). The *ara*BAD genes are required for catabolism of L-arabinose. Two transport systems apear to be present for this sugar. One operon, *ara*FGH, codes for the well-studied arabinose binding protein and additional proteins involved in the high-affinity transport system (Schleif, 1969; Horazdovsky & Hogg, 1989). A second, lower affinity, transport system has been identified for which only one protein product has been observed (Kolodrubetz & Schleif, 1981*c*; MacPherson *et al.*, 1981). The DNA regions controlling transcription of *ara*BAD and *ara*E have been cloned and their regulatory properties have been studied (Smith & Schleif, 1978; Stoner & Schleif, 1983).

In the presence of L-arabinose, transcription of the genes of the *ara*BAD operon is stimulated by

† Author to whom all correspondence should be addressed.

0022–2836/90/200497–14 $03.00/0

© 1990 Academic Press Limited

AraC protein and by cyclic AMP (cAMP†) receptor protein (CRP) (Englesberg *et al.*, 1965; Greenblatt & Schleif, 1971; Wilcox *et al.*, 1974). The binding sites for these proteins have been determined by DNase footprinting of the cloned *ara*BAD and *ara*E promoters (Ogden *et al.*, 1980; Lee *et al.*, 1981; Dunn *et al.*, 1984; Stoner & Schleif, 1983). At both promoters AraC protein binds adjacent to RNA polymerase and CRP binds further upstream next to AraC. In the absence of arabinose, the *ara*BAD operon is repressed by AraC protein bound at two sites whose centers are separated by 211 bp (Dunn *et al.*, 1984; Hahn *et al.*, 1984; Lee & Schleif, 1989; Martin *et al.*, 1986). The two proteins appear to interact with each other, forming a "loop" of the intervening DNA. It is not clear how this structure causes repression.

Maximal expression of the arabinose transport genes, *ara*E and *ara*FGH, is also dependent upon both AraC protein-arabinose and CRP-cAMP (Kolodrubetz & Schleif, 1981*b*,*c*). Studies with Mud*lac* fusions to these promoters on the chromosome have shown that *ara*E and *ara*BAD share similar glucose sensitivity and dependence on CRP-cAMP. In the absence of cAMP the promoters are still capable of substantial induction by arabinose. The *ara*FGH promoter, however, is strongly catabolite sensitive with no promoter activity evident in the absence of cAMP. As seen with the *ara*BAD promoter, the studies *in vivo* suggest that transcription of the transport operons may be repressed as well. Studies of *ara*E transcription *in vitro* confirm the results *in vivo* and show that promoter activity is directly stimulated by AraC and CRP (Stoner & Schleif, 1983).

After the *ara*BAD control region had been cloned, a search for additional arabinose-inducible promoters was initiated by Kosiba & Schleif (1982). One such promoter was cloned from chromosomal DNA and its DNA sequence was determined. The promoter DNA was used as a hybridization probe to identify a set of plasmids from the Clarke–Carbon *E. coli* genomic library that have similar sequences. On the basis of its ability to complement a defect in the *ara*FGH operon and bind arabinose with high affinity, one of the Clarke–Carbon plasmids was suggested to carry the *ara*FGH operon. These data, and the fact that no other *ara* genes had been identified genetically, led to the tentative conclusion that the cloned promoter was *ara*P_{FGH}.

In this paper we describe the characterization *in vitro* of two *ara* promoters. In addition, hybridization studies using chromosomal DNA confirm the identity of the *ara*E promoter. The promoter previously suggested to be *ara*P_{FGH} is shown not to be closely linked with either of the known transport genes. Since there are no other *ara* loci that have been defined genetically, this unidentified promoter is named *ara*P_J and the operon *ara*J. This report also describes the subcloning, DNA sequence and characterization *in vitro* of the authentic *ara*FGH promoter. The requirements for AraC protein and CRP-cAMP have been shown by *in vitro* transcription. The transcriptional start site and the binding sites for the regulatory proteins were also determined.

While this work was in progress, Hogg and coworkers sequenced the entire high affinity *ara* transport operon (Horazdofsky & Hogg, 1987; Scripture *et al.*, 1987) using subclones from the *ara*FGH-containing Clarke–Carbon plasmid identified by Kosiba & Schleif (1982). An additional open reading frame beyond the genetically defined *ara*F and *ara*G genes was found. The proteins encoded by the operon were identified and the new locus named *ara*H. The correspondence between their sequence upstream from the *ara*F gene and our sequence data for the promoter that we cloned confirms that we have characterized *ara*P_{FGH}.

† Abbreviations used: cAMP, cyclic AMP; CRP, cyclic-AMP receptor protein; bp, base-pair(s); kb, 10^3 bases or base-pairs.

2. Materials and Methods

(a) *General materials and methods*

Plasmid DNA was isolated by CsCl/ethidium bromide, density-gradient centrifugation followed by chromatography in agarose A15 (Biorad) (Schleif & Wensink, 1981). For S_1 analysis, cells were grown in M10 medium (Schleif & Wensink, 1981) with 0·2% (w/v) glycerol, 4 μg B1/ml and 0·2% (w/v) arabinose where indicated. Restriction and DNA modifying enzymes were from New England Biolabs, Bethesda Research Labs and Boehringer-Mannheim. DNA rapid preparations, transformations and other general procedures were as described by Maniatis *et al.* (1982) or Schleif & Wensink (1981). Nucleoside triphosphates for *in vitro* transcription were from PL Biochemicals. All chemicals were reagent grade. [γ-^{32}P]ATP was from ICN and Amersham, [α-^{32}P]UTP was from New England Nuclear and [α-^{32}P]dATP was from Amersham. The strains used are *E. coli* K-12 derivatives and are shown in Table 1.

AraC protein, CRP and RNA polymerase were purified as described (Hendrickson & Schleif, 1984) and all were greater than 90% pure. DNA binding activity of AraC protein was 10 to 30%, CRP was 25% and RNA polymerase was 50% as measured by gel-electrophoresis DNA binding assay (Hendrickson & Schleif, 1984). All molar concentrations refer to the amount of active protein present.

(b) *Cloning of the* araFGH *promoter*

Three portions of the plasmid DNA from strain BEK180 that contains the Clarke–Carbon plasmid (Clarke & Carbon, 1976) with high-affinity arabinose transport activity (Kosiba & Schleif, 1982) were each digested with one of *Alu*I, *Hae*III, or *Fnu*DII restriction enzymes, and the resulting DNA fragments were liganded with a 10-fold excess of phosphorylated *Hin*dIII linker. The ligation products were cut with *Hin*dIII and purified on a 2 ml P60 acrylamide column (Biorad), followed by ligation with the 3·6 kb *Hin*dIII–*Eco*RI fragment of the promoter-cloning vector pK01, which contains the *gal*K gene downstream from the *Hin*dIII site (McKenney *et al.*, 1982). The DNA was transformed into the GalK$^-$ strain,

Table 1
Plasmids and strains

Strain	Genotype	Comments
BEK176	pBEK9 ColE1^r, *ara*P_J :: tet/*ara*C^C, *end*A, *thi*, *hsd*R	Kosiba & Schleif (1982)
pWH54	pK01 *ara*P_{FGH} :: *gal*K amp^r	This work
pRFS1595/DLS24	pBR325 *ara*P_E :: tet/*endo*I, *hsd*M, *hsd*A, *thi*1	Schleif & Favreau (1982)
BEK180	ColE1^r, *ara*P_{FGH}/*endo*I, *hsd*M, *hsd*A, *thi*1	Kosiba & Schleif (1982)
SH121	F′ *ara*B53, thr^+, leu^+/Δ*ara*C1022, Δ*lac*74	Hahn & Schleif
SH121/SH84	F′ *ara*B53, thr^+, leu^+/Δ*ara*C1022, Δ*lac*74, Δ*cya*283	Hahn & Schleif
SH313	*thr*, *his*, *ara*$B^-A^+D^+$, *dcm*, *gal*K, *str*	Hahn & Schleif
DJK322	*ara*$C^CB^+A^+D^-$, *thi*1, Δ*lac*74, nal^r, *ara*E, *his*, *thr*	Kolodrubetz & Schleif (1981*a*)
DJK652	*leu*, *thr*, *thi*1, Δ*lac*74, *ara*F :: Mud-*lac*, *ara*E, *his*, *nal*	Kolodrubetz & Schleif (1981*a*)
DJK660	*leu*, *thr*, *thi*1, Δ*lac*74, *ara*(FG)$^-$, *ara*E :: Mud-*lac*, *thy*, *nal*	Kolodrubetz & Schleif (1981*a*)
DJK661	*leu*, *thr*, *thi*1, Δ*lac*74, *ara*(FG)$^-$, *ara*E :: Mud-*lac*, *thy*, *nal*	Kolodrubetz & Schleif (1981*a*)

SH313 and plated on MacConkey, ampicillin, galactose and arabinose medium. Candidates that were positive only in the presence of arabinose were characterized by restriction enzyme digestion.

(c) *Southern hybridization*

Total cellular DNA was isolated from 50 ml cultures by incubating the cells in 2·2 ml of 20 mM-Tris (pH 8·0), 0·15 M-NaCl, 0·1 M-NaEDTA, 1 mg lysozyme/ml for 30 min at 37°C, followed by addition of 200 μl of 10% (w/v) Sarcosyl and digested with proteinase K (50 μg/ml) for 30 min at 37°C. The samples were gently extracted 3 times with phenol and ether and were dialyzed against 10 mM-Tris (pH 7·4) 10 mM-NaCl, 1 mM-EDTA. After incubation with restriction enzymes, 0·5 μg samples were electrophoresed on 0·5% (w/v) agarose gels. Restriction enzyme digested pBR322 and pBEK180 DNA were run simultaneously as size standards. The DNA was transferred to nitrocellulose and hybridized at 42°C for 24 h (Schleif & Wensink, 1981). DNA probes were labeled by nick translation (Maniatis *et al.*, 1982) with [α-^{32}P]ATP to a specific activity of 2 to 5×10^8 disints/min per μg. Probes used were: *ara*E, 370 bp *Eco*RI fragment isolated from pRFS1595 (Stoner & Schleif, 1983); *ara*J, 600 bp *Eco*RI fragment from pBEK9 (Kosiba & Schleif, 1982); *ara*FGH, 500 bp *Eco*RI–*Hin*dIII fragment from pWH54. Fragments were isolated by electroelution from 5% (w/v) polyacrylamide gels onto NA45 membranes (Schleicher & Schuell) following the manufacturer's directions. After hybridization and washing, the membranes were dried and exposed to Kodak XAR5 film.

(d) *DNA sequencing*

The *ara*FGH control region was sequenced by the dideoxyribonucleotide chain-termination method (Sanger *et al.*, 1977) using as a template pWH54. The template was made single stranded by digesting with either *Hin*dIII or *Eco*RI and digesting away 1 DNA strand with exonuclease III. Sequencing primers were used that hybridize to the pK01 vector in regions adjacent to the *Eco*RI and *Hin*dIII sites and 40 bp upstream from the *Eco*RI site. Sequencing was also performed rightward from a primer at the *Asp*718 site and leftward from a primer at +20 (Fig. 3). Oligonucleotides were synthesized on an Applied Biosystems 380A synthesizer.

(e) In vitro *transcription*

Transcription reactions were performed essentially as described by Hahn *et al.* (1986). Plasmid WH54 DNA was linearized by digestion with restriction enzyme, extracted with phenol, precipitated with ethanol and resuspended in 10 mM-Tris (pH 8·0), 1 mM-EDTA. Transcription reactions were 20 μl and contained a 0·5 nM-DNA template in a buffer of 20 mM-Tris (pH 7·4), 100 mM-KCl, 5 mM-$MgCl_2$, 0·25 mM-EDTA, 1 mM-dithiothreitol, 50 mM-arabinose, 0·2 mM-cAMP and 50 μg bovine serum albumin/ml. AraC protein, CRP and RNA polymerase were added and allowed to bind the template for 10 min at 37°C. A single round of transcription was initiated by addition of nucleotides to a final concentration of 100 μM-CTP, 200 μM each GTP and ATP, 10 μM-UTP and 20 μCi [α-^{32}P]UTP, and heparin to a concentration of 100 μg/ml. After 15 min of incubation at 37°C, reactions were stopped by the addition of an equal volume of 85% (v/v) formamide, 20 mM-EDTA, 0·05% (w/v) bromophenol blue and xylene cyanol. Samples were then heated at 90°C for 3 min and the entire sample immediately loaded onto a 6% denaturing polyacrylamide gel. After electrophoresis the gel was dried and transcripts visualized by autoradiography. The 100 base RNA-1 transcript from the plasmid origin of replication was synthesized in all reactions.

(f) *DNase I footprinting*

DNase I footprinting experiments were performed as described (Stoner & Schleif, 1983). Binding buffer was 20 mM-Tris·HCl (pH 7·4), 3·5 mM-$MgCl_2$, 0·1 mM-$CaCl_2$, 0·1 mM-EDTA, 50 mM-KCl and 25 μg bovine serum albumin/ml. Proteins were added to a 20,000 cts/min DNA template (0·2 to 1×10^7 cts/min per μg; 0·5 to 2 nM) in 100 μl of buffer and allowed to bind for 10 to 15 min. DNase I was added to a final concentration of 40 to 80 ng/ml, and the DNase I reaction was performed at room temperature for 30 s then stopped by the addition of 20 μl of 8 M-ammonium acetate and 5 μg of yeast tRNA. The samples were precipitated and washed with ethanol, lyophilized and resuspended in 5 μl of 80% formamide loading buffer (Maxam & Gilbert, 1980) and electrophoresed on 8% or 20% denaturing polyacrylamide gels (Sanger *et al.*, 1977). Sequencing reactions (Maxam & Gilbert, 1980) of the same DNA were run concurrently to locate the binding sites.

For binding reactions with AraC protein in the absence of arabinose, the 10 mM-sugar present in the protein storage buffer was removed by dialyzing 100 μl of protein for 24 h against 3 changes of 25 ml storage buffer without sugar. The absence of sugar was confirmed by gel electrophoresis DNA binding assays in which the affinity of the protein *ara*I_{BAD} was reduced by 50-fold.

To footprint the bottom (sense) strand of the *ara*FGH promoter, pWH54 was digested with *Hin*dIII or *Taq*I, end-labeled with phage T4 polynucleotide kinase and digested with *Eco*RI. The promoter-containing fragments were isolated following electrophoresis in a 6% polyacrylamide gel. The top (antisense) strand was labeled by digesting the DNA with *Asp*718, labeling at the 5′ ends, digesting with *Hin*dIII and then isolating the fragments as above.

For the isolation of *ara*J fragments labeled at one end of the bottom strand, the 600 bp P_J DNA fragment was isolated from pBEK9, labeled at the 5′ ends, cut 100 bp from the left end with *Taq*I, and the promoter fragments was purified by electrophoresis. To label the top strand, DNA was digested with *Msp*I 150 bp upstream from the transcriptional start point, labeled at the 5′ ends, digested with *Hha*I, which cuts 20 bp from the right end, and the promoter fragment was purified.

(g) S_1 *nuclease mapping*

The isolation of total cellular RNA and S_1 mapping were carried out as described by Stoner & Schleif (1983). The probe for P_{FGH} was the 500 bp *Hin*dIII–*Eco*RI fragment labeled at the 5′ end of the *Hin*dIII site. The probe of P_J was the 500 bp *Taq*I–*Eco*RI fragment labeled at the 5′ end of the *Eco*RI site. Cells were harvested at an A_{550} of 0·6 to 0·7 by pouring 50 ml of culture into a centrifuge bottle containing 50 g of ice, 6 mg of chloramphenicol and 1 mmol NaN_3. The cells were then pelleted, resuspended in 2 ml of 10 mM-Tris·HCl (pH 7·4), 10 mM-KCl, 5 mM-$MgCl_2$ and the RNA was prepared. For each reaction 50,000 cts/min (10 ng) of denatured DNA and 25 μg of RNA were co-precipitated and resuspended in 35 μl hybridization buffer (80% formamide, 0·04 M-Pipes-HCl (pH 6·4), 0·4 M-NaCl, 1 mM-EDTA). For mapping P_{FGH} the samples were heated at 75°C for 10 min, then slowly cooled to 37°C and incubated for 2 h. For P_J the samples were heated to 70°C for 10 min and hybridization was done at 45°C. The samples then were diluted with 315 μl of 5% (v/v) glycerol, 0·25 M-NaCl, 0·03 M-sodium acetate (pH 4·5), 1 mM-$ZnSO_4$ and digested with 5 to 40 units of S_1 nuclease for 30 min at 32, 37 or 45°C as indicated in Figs 2 and 6. Reactions were terminated by the addition of 100 μl of 2·5 M-ammonium acetate, 50 mM-EDTA, 10 μg of tRNA plus 500 μl cold isopropanol. After precipitation 3 times, the samples were resuspended in 3 μl of loading buffer (80% deionized formamide, 50 mM-Tris-borate (pH 8·3), 1 mM-EDTA, 0·1% bromophenol blue and xylene cyanol), heated to 90°C for 3 min and loaded on 8% polyacrylamide denaturing gels (Sanger *et al.*, 1977). Sequencing reactions (Maxam & Gilbert, 1980) of the same fragment were run concurrently to determine the position of the protected bands. The 1·5 base difference in mobility between the S_1-digested band and the sequencing bands (Hentschel *et al.*, 1980) was taken into consideration. The gels were autoradiographed overnight with Kodak XAR5 film and an intensifying screen at −70°C.

3. Results

(a) *Identity of cloned promoters*

The relationship between the cloned promoters of plasmids pRFS1595 (Stoner & Schleif, 1983) and pBEK9 (Kosiba & Schleif, 1982), and the chromosomal genes for the two arabinose transport operons, *ara*E and *ara*FGH, was determined by hybridization of the cloned DNA fragments with chromosomal restriction digests. Chromosomal DNA was isolated from wild-type strains and from strains containing phage Mu*dlac* insertion in the *ara* transport operons (Kolodrubetz & Schleif, 1981*c*). These insertions inactivate the transport gene and also place the *lac* gene under control of the *ara* promoters. Any restriction fragments containing both the promoter and the insertion point of the Mu*dlac* will be altered in size compared to the same fragment from cells without the Mu*dlac* insertion, and this shift in size will be detected if the cloned DNA fragment hybridizes to the appropriate chromosomal location. Figure 1 shows such an experiment in which Southern hybridizations were performed with strains containing Mu*dlac* inserted into either *ara*F or *ara*E. For hybridizations with three different chromosomal restriction digests, *Bam*HI, *Eco*RI and *Hin*dIII, the size of the hybridized chromosomal fragment shifted when strains with insertions in *ara*E were probed with the putative *ara*P_E fragment of pRFS1595 (Fig. 1(a)). P_E appears to reside in a chromosomal region characterized by an 11 kb *Eco*RI fragment, a 1·2 kb *Bam*HI fragment and an 8 kb *Hin*dIII fragment.

The location of P_E with respect to the Mu*dlac* insertions also could be deduced. From the size of the smallest fragment and from the restriction map of Mu*dlac* (O'Connor & Malamy, 1983), the promoter sequence can be located to a maximum distance of 1 kb from the point of the insertions. The *Bam*HI site on Mu is 8·5 kb from the appropriate end of the phage. The chromosomal fragment size shift from 1·2 to 9·5 kb, so the fragment contains about 1 kb of chromosomal DNA. The *Hin*dIII fragment shifts from 8 kb to 1·9 kb. Since the nearest *Hin*dIII site of Mu*dlac* is 250 bp from the end of the phage, the shifted fragment contains 1·7 kb of chromosomal DNA. No hybridization to any additional chromosomal fragments was detected, indicating that the probe sequence is contained within the 9·5 and 1·9 kb fragments. Since the transcriptional start is located 300 bp from the upstream end of the probe, the distance from the *ara*E promoter to the nearest Mu insertion is actually less than the 1 kb maximum estimated from the *Bam*HI data.

The putative P_{FGH} DNA fragment from pBEK9, however, is not located near either of the transport genes. Hybridizations to Clarke–Carbon plasmids containing either *ara*E or *ara*FGH complementing activities show that the promoter is not within 3 to 5 kb flanking the operon that are contained on the plasmids, since no hybridization was observed under stringent conditions. In addition, no shift in the fragment size of the chromosomal region homologous with the probe was seen with Mu inserted at either *ara*E or *ara*FGH (Fig. 1(b)).

(b) *Subcloning of the authentic* ara*FGH promoter*

A Clarke–Carbon plasmid (Clarke & Carbon, 1976) containing an activity that could complement a

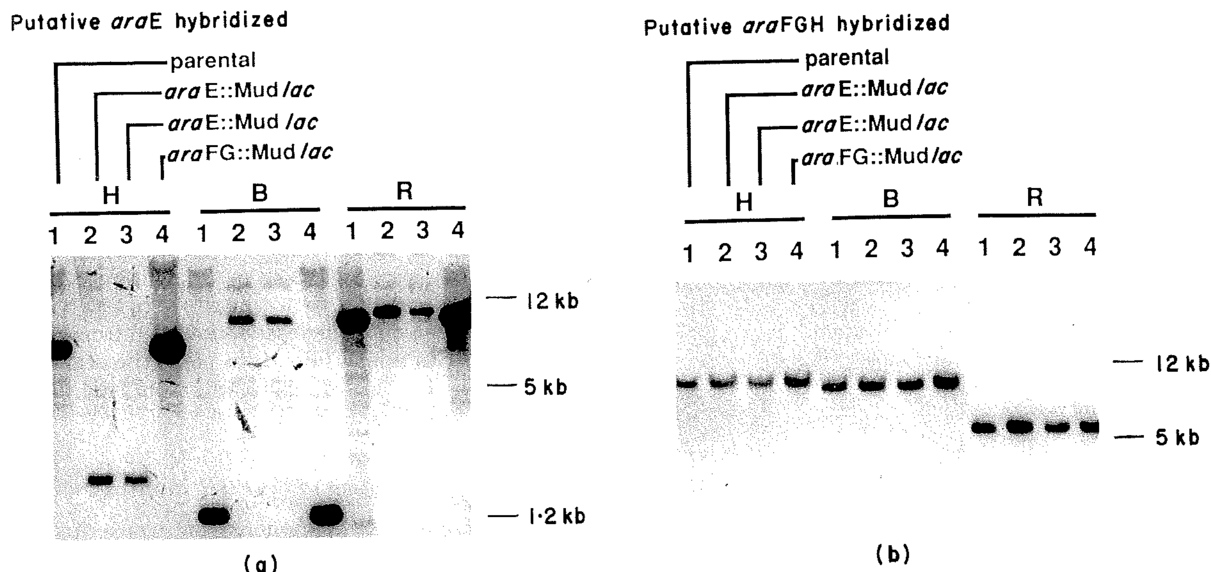


Figure 1. Southern hybridization of DNA from Mud*lac* fusion strains. Chromosomal DNA (0·5 μg) digested with *Eco*RI (R), *Bam*HI (B) or *Hin*dIII (H) was run on a 0·5% (w/v) agarose gel and transferred to a nitrocellulose membrane. (a) The 370 bp fragment from pRFS1595 (putative $araP_E$) and (b) the 600 bp *Eco*RI fragment of pBEK9 (putative $araP_{FG}$) were labeled by nick translation and hybridized. Lane 1, parental strain (DJK322); lane 2, *ara*E :: Mud*lac* (DJK660); lanes 3, *ara*E :: Mud*lac* (DJK661); lane 4, *ara*FG :: Mud*lac* (DJK651). The positions of size markers are indicated at right.

defect in high-affinity L-arabinose-binding activity was previously identified (Kosiba & Schleif, 1982). After finding that the promoter Kosiba & Schleif had identified tentatively as P_{FGH} was not in fact P_{FGH}, we used the Clarke–Carbon plasmid of strain BEK180 as a source of the promoter. DNA from the plasmid was subcloned to isolate a small fragment containing an active *ara*FGH promoter Plasmid BEK180 DNA was cleaved with restriction enzymes *Hae*III, *Fnu*dII or *Alu*I in separate reactions and *Hin*dIII linkers were added. The fragments were ligated to the promoter vector pK01, which places the galactokinase gene under control of the cloned promoter (McKenney *et al.*, 1982). The DNA was transformed into a $GalK^-$ strain and isolates were obtained that were Gal^+ only in the presence of arabinose. A 1·1 kb *Hin*dIII fragment containing an arabinose dependent promoter was found in several candidates. The plasmid was cleaved with *Eco*RI nuclease to remove a 600 bp fragment upstream from the promoter producing the plasmid pWH54, which contains a 500 bp *Hin*dIII–*Eco*RI fragment with the *ara*FGH promoter. Sequence analysis showed that the *Hin*dIII linker had been added at an *Alu*I site (Fig. 2). The sequence of the *ara*FGH operon has been determined by Scripture *et al.* (1987). With few exceptions (see Discussion) our sequence matches the region from 12 to 518 bp upstream from the *ara*F gene, confirming that pWH54 does contain the *ara*FGH promoter.

(c) *Transcriptional start site of* P_{FGH}

The start site of transcription *in vivo* was mapped by isolating cellular RNA and performing S_1 assays using as a probe the 500 bp promoter fragments 5′ labeled at the *Hin*dIII site. This probe will not detect mRNA produced from the chromosomal copy

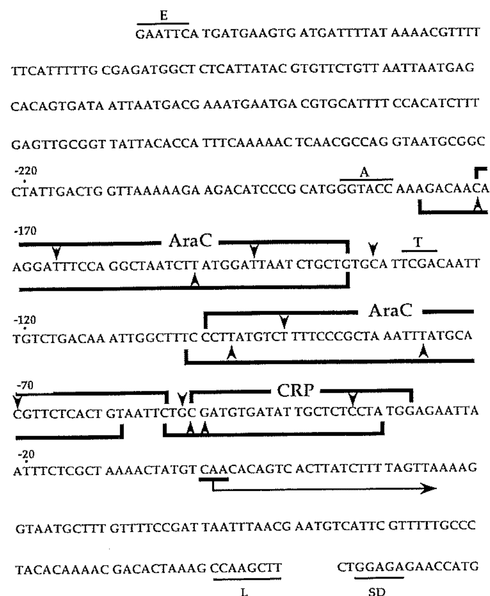


Figure 2. Sequence and location of sites of the *ara*FGH control region. The sequence is numbered from the transcriptional start site identified by S_1 nuclease mapping. The S_1-mapping start site is indicated by the rightward arrow. The sequence extending from this clone to the start of translation (Scripture *et al.*, 1987) is indicated in the lower right. SD, ribosome binding site; L, the *Hin*dIII linker sequence; E, *Eco*RI site; A, *Asp*718 site; T, *Taq*I site. Brackets show regions protected from DNase I digestion by the indicated protein, and arrows show sites of enhanced cleavage.

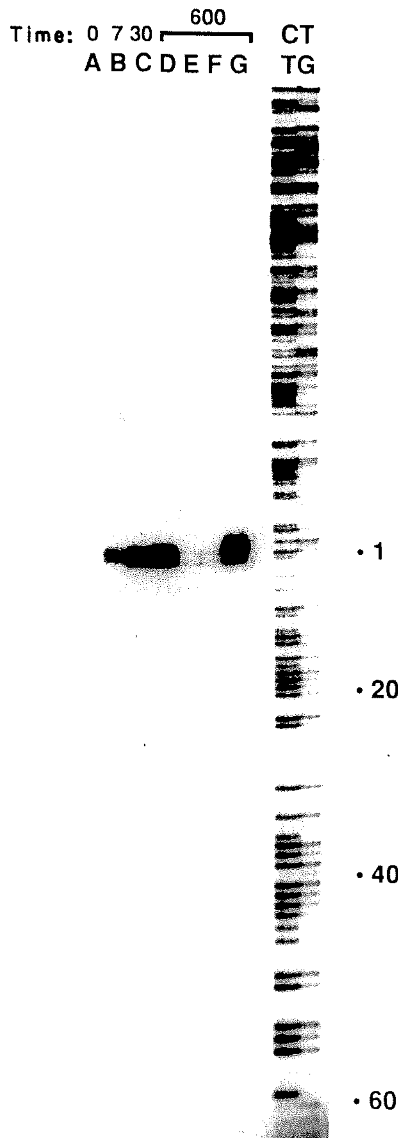


Figure 3. S_1 mapping of the *ara*FGH transcriptional start site. Cellular RNA from strain pWH54/SH313 was hybridized to the 500 bp P_{FGH} probe 5′ end-labeled at the *Hin*dIII site and subjected to S_1 nuclease analysis. Lanes A to D, DNA–RNA hybrids were digested with 5 units S_1 at 37 °C. Cells were grown in M9 salts glycerol medium, and L-arabinose was added before harvesting for time: lane A, 0 min; lane B, 7 min; lane C, 30 min; and lane D, 5 generations. Lanes E to G, cells were grown in the presence of arabinose for 5 generations and DNA–RNA hybrids incubated with: lane E, 40 units S_1, 37 °C; lane F, 5 units S_1, 45 °C; lane G, 5 units S_1 32 °C. Sequencing size markers, G+A and A+C reactions, are shown at right; the sequence of the top strand is indicated.

of *ara*FGH, since the *Hin*dIII end was added as a linker and will not hybridize. Therefore, the kinetics of induction of the operon were studied by growing cells containing the plasmid pWH54 in the presence or in the absence of arabinose. In the absence of arabinose, no signal was observed (Fig. 3). After an induction period of seven minutes very little transcript was produced, while after 30 minutes approximately one-half the maximal level of transcription was seen. When cells were grown for five generations (600 min) in arabinose, the maximum signal was obtained at one region that we have assigned +1. The signal contains a cluster of three adjacent bands, so the central band is proposed as the start site with an errror of ±1 bp. No other apparent sites were observed when the digestion temperature was raised to 45 °C or lowered to 32 °C or when the level of S_1 nuclease was varied from 5 to 40 units per reaction. These data place the start of transcription 111 bases upstream from the first codon of the *ara*F gene.

(d) *Transcription of* ara*FGH* in vitro

To determine the requirements for transcription of this operon, we perform *in vitro* transcription experiments using purified AraC, CRP and RNA polymerase. The plasmid was linearized by cutting at the *Sma*I site at +120, and reactions were carried out on whole plasmid DNA. Only one run-off transcript was found to be dependent upon both AraC and CRP. In the absence of AraC protein a low level of a 125 base transcript was seen (Fig. 4). In the presence of saturating amounts of both activators, the same transcription product was increased fivefold. In the absence of CRP no 125 base product is made even in the presence of maximal AraC protein. The observed transcript runs off the right end of the linearized plasmid, as demonstrated by the reduction in the size of the product when the plasmid is cleaved at the *Hin*dIII site, reducing the size of the right end by 17 bp (Fig. 4, lane 9). A 170 base transcript of unknown origin (marked × in Fig. 4) is produced in the presence of RNA polymerase and only in the absence of CRP. The product originates from a region near the left end of the P_{FGH} fragment and it terminates within the vector (data not shown). It is not clear if this transcript is produced *in vivo*, and it is not always seen with *in vitro* reactions.

(e) *DNase footprint of the* ara*FGH promoter*

The binding sites for the transcriptional activator proteins were located by DNase I protection assay. The *ara*FGH DNA fragment of pWH54 was labeled on the bottom strand at the *Hin*dIII site, and the binding sites for AraC protein and CRP were determined (Fig. 5(a)). AraC protein protected two regons of approximately 40 bp each. One region extends from −60 to −103, while the other is found at −136 to −178. Since the function of these sites is not known, they have been designated *ara*FG_1 for the site centered at −81 and *ara*FG_2 for the site at −157. At a low concentration of AraC protein the *ara*FG2 site was fully occupied, while no protection of *ara*FG1 was observed. This suggests that the affinity of FG_2 for AraC protein is higher than the affinity of FG_1, which is opposite the relative affini-

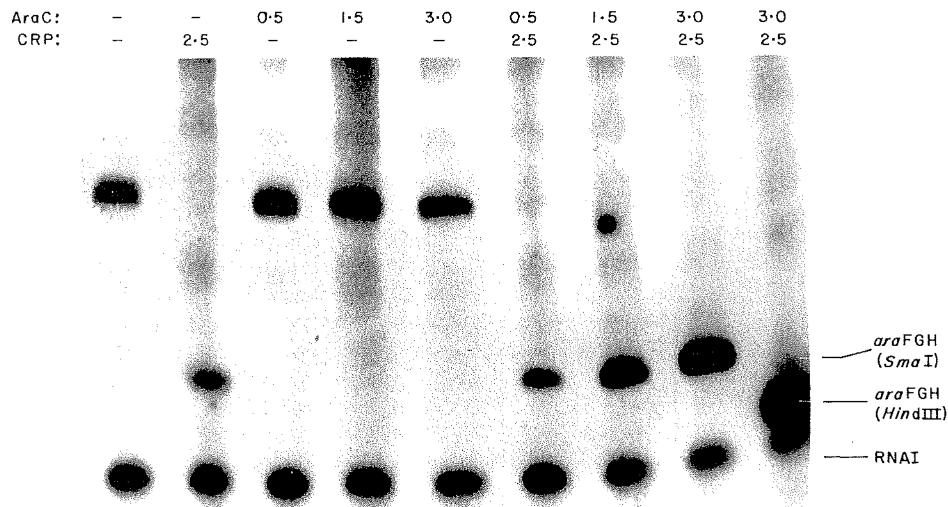


Figure 4. Transcription *in vitro* from P_{FGH}. Plasmid pWH54 was linearized by digestion with *Sma*I (site located in the vector 17 bp to the right of the *Hin*dIII site), lanes 1 to 8, or by digestion with *Hin*dIII, lane 9. Single-round transcription reactions were performed in buffer containing cAMP and L-arabinose. Proteins were bound for 10 min and transcription was allowed to proceed for 15 min at 37 °C. All reactions contained 5 nM-RNA polymerase plus the indicated concentrations, given in nM-active protein, of CRP and AraC protein.

ties of the $araO_2$ and *ara*I sites at P_{BAD} (Hendrickson & Schleif, 1984). A weak enhanced cleavage was seen at −67 and strongly enhanced cleavage occurred at −77 and 98.

Similarly sized footprints were obtained for reactions performed with DNA labeled on the top strand (Fig. 5(b)). AraC protection was from −55 to −100 and from −136 to −173. Enhanced cleavage occurred at −71 and −92 in the FG_1 site and at −133, −146 and −167 in FG_2. The footprint of the FG_2 site was obtained by labeling DNA at the *Asp*718 restriction site just upstream from FG_2, and the reaction products were analyzed on 20% polyacrylamide gels (not shown). Incubation of AraC protein with DNA in the absence of arabinose also produced a footprint identical with that seen in the presence of the sugar, although an increased amount of protein was required (Fig. 5(b), lane 10). The footprinting data are summarized in Figure 2.

CRP-cAMP protected a region of 22 bp of the bottom strand from −32 to −54, and enhanced cleavage occurred at −51 and −52 (Fig. 5(a)). On the top strand 24 bp from −29 to −52 were protected, and cleavage at −53 was strongly enhanced (Fig. 5(b)). When both AraC and CRP were present, the entire region from −29 to −100 was protected. The region protected by CRP contains a sequence that matches the CRP consensus at 14 out of 22 positions including eight out of the ten most highly conserved bases (Berg & von Hippel, 1988). The CRP binding sequence is centered 41·5 bp upstream from the *ara*FGH transcriptional start, which places the right one-half of the site in the −35 region of the promoter.

(f) *Transcription from the* ara*J promoter*

The transcription start point and relative amounts of cellular RNA produced by the promoter identified on plasmid pBEK9 were determined by S_1 nuclease mapping. RNA was extracted from uninduced cells, cells induced by arabinose for various times, cells exposed to the anti-inducer D-fucose and arabinose-induced cells deleted of adenyl cyclase. Figure 6 shows the results of these assays, as well as a "G" sequencing reaction that was done on the same DNA to provide a size standard. Only one arabinose-induced hybrid to the bottom strand is apparent in the range of ±125 bases of the possible transcription start points identified *in vitro* (Kosiba & Schleif, 1982). Hybridization to the other strand showed no evidence of an arabinose-induced transcript.

The DNA fragments of the bottom strand protected by the hybridized RNA varied in size by seven bases. After correction of the 1·5 base migration velocity difference between S_1 digested DNA and the chemically modified size standard DNA (Hentschel *et al.*, 1980), the fragments were seen to range in size from 156 to 163 bases. The various sizes of protected DNA might be thought to arise from incomplete S_1 digestion, but variation of the nuclease concentration from 15 to 40 units, a concentration that almost eliminated the protected

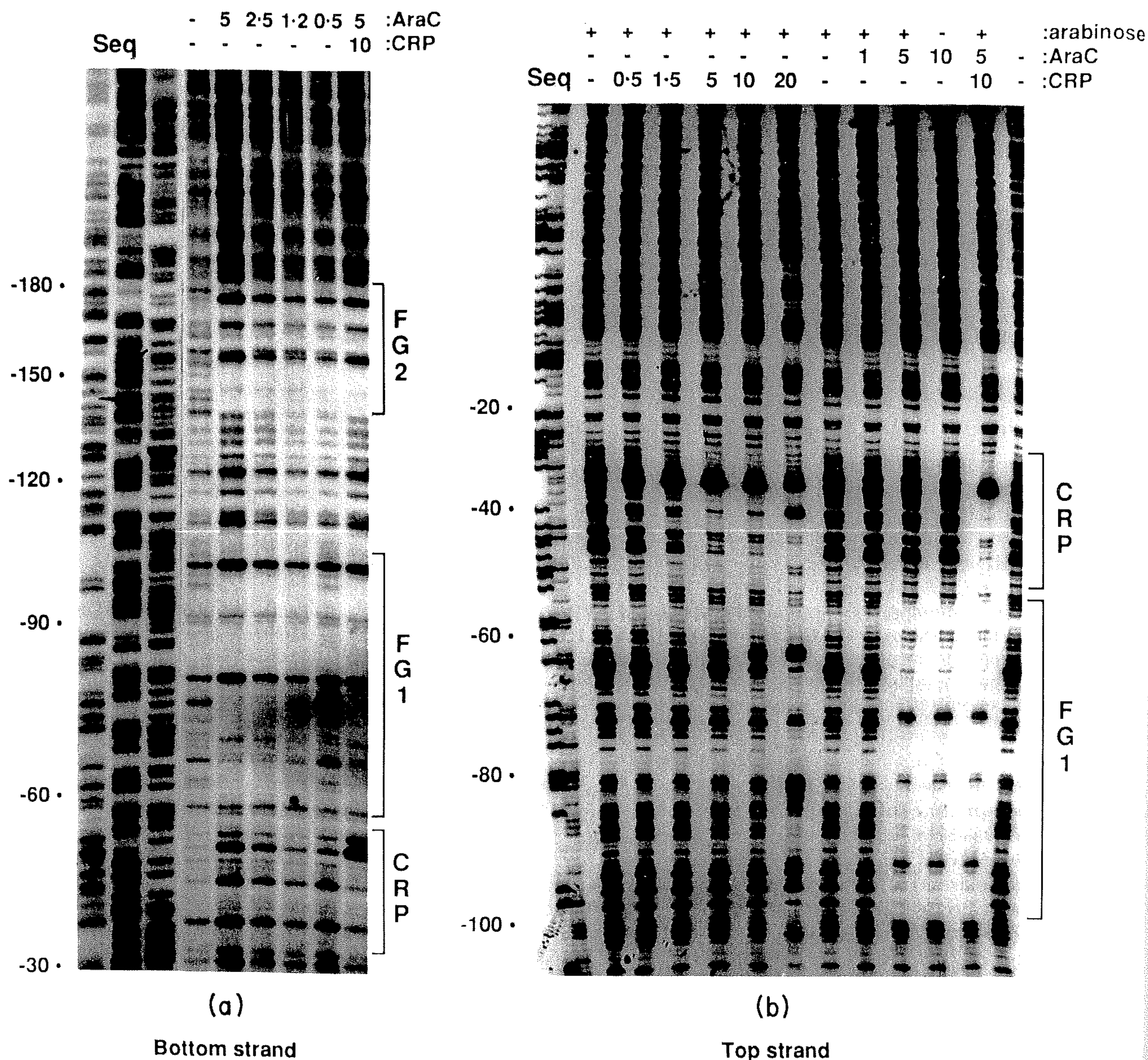


Figure 5. DNase I footprints of the *ara*FGH control region. The *ara*FGH fragment was 5′ end-labeled at the *Hin*dIII site, bottom strand (a) or at the *Asp*718 site, top strand (b). DNA was incubated with proteins for 10 min at 37 °C, digested with DNase I and run on 8% denaturing gels. Seq, sequencing size standards. (a) All reactions contained 50 mM-arabinose and 50 μM-cAMP except (b), lane 10, which had no arabinose or cAMP. Concentrations of AraC protein and CRP in nM active protein are indicated.

DNA, did not alter the range of sizes protected. For numbering the nucleotides of the DNA, we take position 1 of the mRNA to correspond to a transcript of 161 bases. Such an assignment maximizes the similarity in sequence and position of this start site with other characterized start sites (Hawley & McClure, 1983; Harley & Reynolds, 1987). On the basis of the findings that the promoter located on the chromosome produces an arabinose-regulated transcript *in vivo*, and that it is not at either the *ara*E or *ara*FGH loci, we name the promoter *ara*P_J and tentatively name the operon *ara*J.

Figure 6 also shows that P_J is almost fully induced within two minutes of arabinose addition, and its induction increases only slightly over the next 150 minutes. Similar to the other *ara* operons, induction of P_J requires AraC protein, arabinose plus cAMP-CRP, and no induction is observed in cells incubated with the anti-inducer D-fucose.

(g) *Footprints of* araP_J

The binding sites for AraC protein and CRP–cAMP complex were determined by DNase I

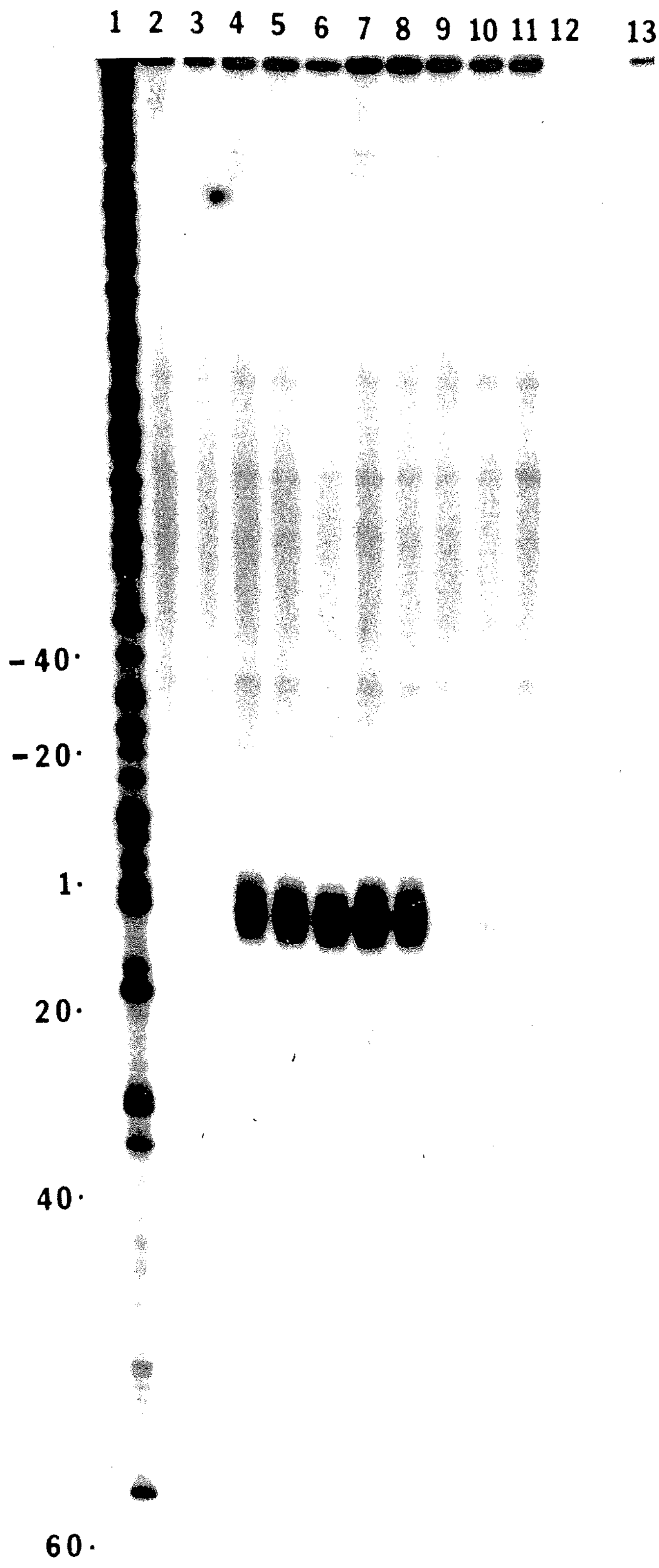


Figure 6. S_1 Mapping of the *ara*J transcription start site. The *Eco*RI–*Taq*I DNA fragment end-labeled on the sense strand was denatured and hybridized to total cellular RNA isolated from cells grown in the absence or presence of arabinose, or grown in the presence of D-fucose. Lane 1, G sequencing reaction size standard; lane 2, RNA from cells grown without arabinose; lane 3, RNA from cells grown in the absence of arabinose that were added to ice containing arabinose, sodium azide and chloramphenicol; lane 4, RNA from cells+arabinose for 2 min; lane 5, RNA from cells+arabinose for 6 min; lane 6, RNA from cells+arabinose for 50 min; lane 8, RNA from cells+arabinose for 150 min; lane 9, RNA from Δ*cya* cells (SH121/SH84), no arabinose; lane 10, RNA from Δ*cya* cells+arabinose for 6 min; lane 11, RNA from cells+fucose for 6 min; lane 12, RNA alone control; lane 13, DNA alone control.

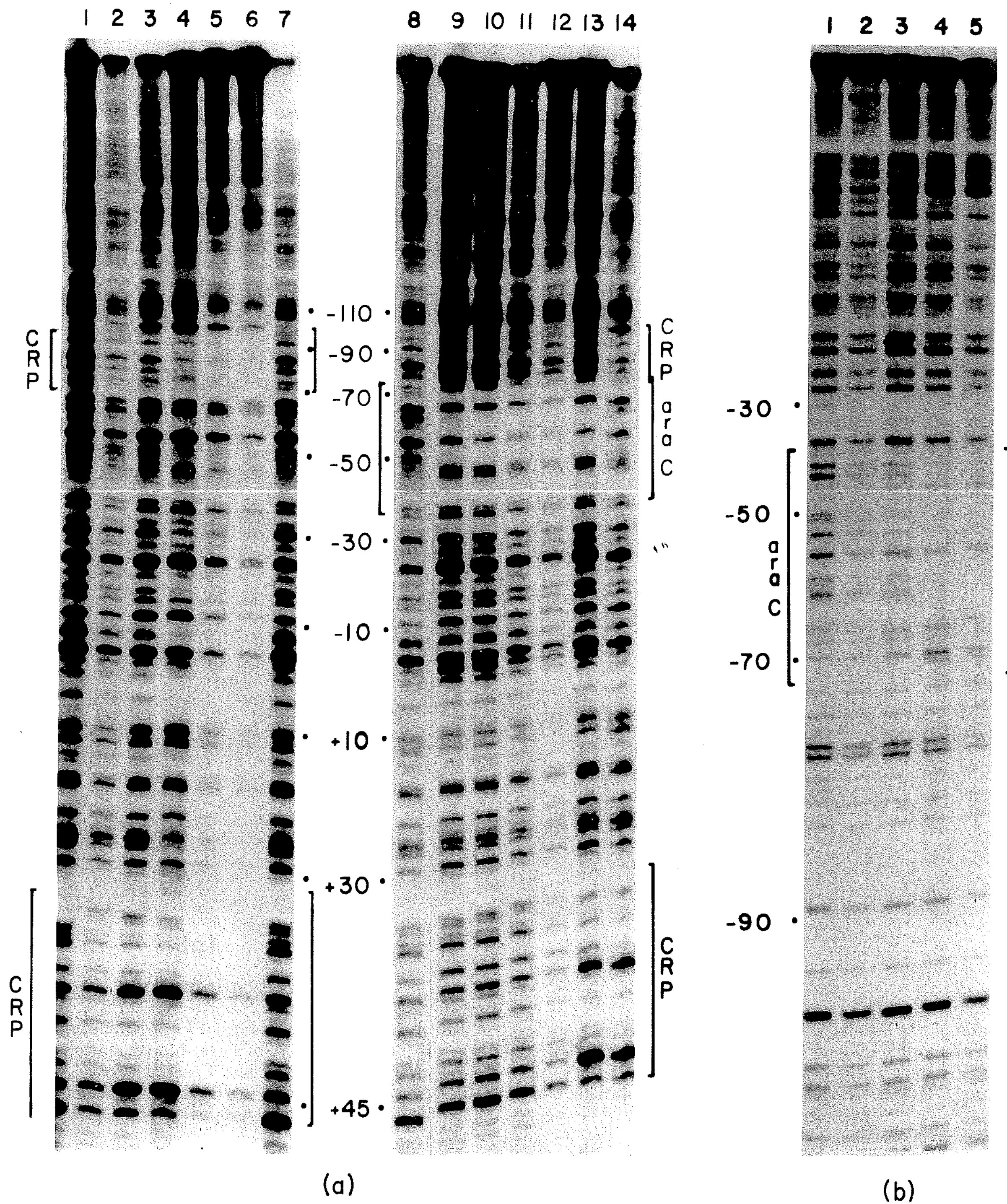


Figure 7. DNase I footprints of the *ara*J regulatory region. The end-labeled *ara*J DNA fragment was incubated with the indicated proteins for 10′ at 37°C, digested with DNase I for 30 s and run on 8% acrylamide sequencing gels. A G sequencing reaction on the same DNA was run in parallel for size standards. Samples were digested with (a) 0·04 μg/ml or (b) 0·08 μg DNase I/ml. All reactions with CRP contained 200 μM-cAMP and sugars were added at 100 mM. (a) The DNA fragment was labeled on the bottom (sense) strand at an *Eco*RI site (+150). Lane 1, no protein, 0·02 μg DNase I/ml; lane 2, 0·5 nM-CRP; lane 3, 2·5 nM-CRP; lane 4, 5·0 nM-CRP; lane 5, 10 nM-CRP; lane 6, 50 nM-CRP; lane 7, no protein, 0·08 μg DNase I/ml; lane 8, no protein, 0·04 μg DNase I/ml; lane 9, 15 nM-AraC protein + fucose; lane 10, 3 nM-AraC protein + arabinose; lane 11, 7·5 nM-AraC protein + arabinose; lane 12, 15 nM-AraC protein + arabinose; lane 13, 5 nM-AraC protein + arabinose, 0·5 nM-CRP; lane 14, 2·5 nM-AraC protein + arabinose, 10 nM-CRP. (b) DNA was end-labeled on the top strand at an *Msp*I site (−148). Lane 1, no protein; lane 2, 1·5 nM-AraC protein + arabinose; lane 3, 15 nM-AraC protein + fucose; lane 4, 3 nM-AraC protein + arabinose; lane 5, 7·5 nM-AraC protein + arabinose.

protection studies. A G sequencing reaction on the same DNA was run in parallel to locate the protected regions. In the presence of arabinose, AraC protein protects a single region of the bottom strand from −37 to −73 (Fig. 7). Protection of the top strand was from −40 to −75 bp. This site is similar to the *ara*I site of P_{BAD} and P_E. AraC protein in the presence of the anti-inducer D-fucose protected the same site, but higher concentrations of protein were required to obtain an equal level of protection.

The site of CRP-cAMP binding is less clear. At a level of 0·6 nM-CRP plus 200 μM-cAMP enhanced cleavage occurred at −100 as well as +44. As the CRP concentration was increased to 6 nM, protection of both regions became more clear from −71 to −100 and from +34 to +46. At levels greater than 6 nM, however, the protein also appeared to bind non-specifically to many sites on the DNA. When cAMP concentration was varied, protection at both sites was reduced equally at levels below 10 μM-cAMP. Salt concentrations of 5 to 150 mM also had no effect on the relative binding at the two sites. When cAMP-CRP and AraC protein plus arabinose were both added, protection extended from −37 tp −100 and the protection in the +44 region could be seen. No distinct RNA polymerase binding site could be detected when RNA polymerase was added in addition to the positive effectors.

4. Discussion

With this work we have completed the cloning and initial characterization of the *ara* promoters of *E. coli*. All the *ara* promoters are maximally active when AraC protein-arabinose and CRP-cAMP are bound to sites near the start of transcription. Among the promoters, however, the relative positions of these sites vary. While the structures of P_{BAD} and P_E are quite similar (Stoner & Schleif, 1983), the locations of the AraC and CRP sites within P_{FGH} were found to be reverse compared with the other *ara* promoters. At P_{FGH} the CRP site is found between AraC protein and RNA polymerase overlapping the −35 region, whereas at P_E and P_{BAD} the AraC site is adjacent to RNA polymerase. CRP and AraC binding sites at P_J were found in the same order as P_{BAD}, but an additional CRP site was found downstream from the transcriptional start. A second AraC binding site was found at P_{FGH} centered at −156. Although AraC binds this site with high affinity, its function is not known. No additional AraC binding sites have been seen at P_J or P_E; however, a site with weak affinity similar to *ara*O_2 may not have been detected in our assays since it might be near the end or even beyond the limit of the cloned DNA fragments.

The central position of CRP, between AraC protein and RNA polymerase, is consistent with the sensitivity of *ara*FGH to the presence of CRP-cAMP both *in vivo* and *in vitro*. The regulation of the *ara* transport operon promoters has been studied previously with *lac* gene fusions (Koldrubetz & Schleif, 1981*b*). In a strain deleted of adenyl cyclase, P_{FGH} activity in the presence of arabinose was reduced 100-fold to a level equal to the uninduced basal activity. Under similar conditions P_E and P_{BAD} activities were reduced only 25-fold and the promoters were still inducible by arabinose. Similarly, the addition of glucose to the cell growth medium reduced activity of P_{FGH} to a greater extent than that of the other promoters. These findings are reflected in the results from transcription *in vitro*. When only one activator was added, low levels of P_{FGH} transcription were obtained in the presence of CRP but not AraC protein, while previous studies have shown that transcription of P_{BAD} proceeds with AraC protein and not with CRP alone (Hahn *et al.*, 1986).

It has been suggested that AraC binds at *ara*I_{BAD} to a sequence with 2-fold symmetry spanning three turns of the DNA helix (Hendrickson & Schleif, 1985). A directly repeated sequence spanning four turns of the helix also is evident when AraC binding sites are compared (Lee *et al.*, 1987; Brunelle & Schleif, 1989). Loss of contact experiments with mutant AraC proteins provide convincing data for the asymmetric binding model (Brunelle & Schleif, 1989). Dominant-negative AraC mutants that fail to activate transcription of *ara*BAD were selected, and their interactions with specific bases of the binding site were probed with chemically modified DNA. Two mutant AraC proteins were isolated that no longer respond to the loss or the methylation of guanines in a pattern that is consistent with a directly repeated recognition sequence, and the pattern is inconsistent with an inverted repeat. The sequences protected by AraC protein at the *ara*FG1 and FG2 sites also are much more consistent with the proposed directly repeated sequence than with a symmetric binding site.

The suggestion that the AraC binding sites are asymmetric implies that they also have an orientation with respect to the direction of transcription of the promoters. Compared with the AraC binding sites at the other *ara* promoters, the orientation of the *ara*FG1 and *ara*FG2 sites is inverted. The inversion of the FG_1 site is consistent with the reversed relative positions of the AraC and CRP sites in the promoter. Thus, the same "end" of the AraC site is always adjacent to the CRP site.

Our results show that on linear DNA the *ara*FG1 and FG2 sites are protected by AraC protein over a region spanning four turns of the DNA helix both in the presence and in the absence of arabinose. Lee *et al.* (1987) have published data on the *ara*I site of P_{BAD}, also obtained using linear DNA, indicating that only a two turn region of this site is occupied in the absence of arabinose, but that a four turn region is occupied in the presence of arabinose. Their correlation of occupancy of the second half of *ara*I with induction leads them to propose that AraC occupancy of this part of the site is necessary and sufficient for induction. Their data does not agree with our own data on AraC binding at *ara*I, and

their hypothesis is also incompatible with our results on the regulation and footprinting of P_{FGH} on linear DNA. It appears unlikely that the slight differences between the buffers that we used for footprinting and for *in vitro* transcription could be the cause for the discrepancy.

The sequence obtained for the *ara*FGH fragment is essentially the same as that obtained by Scripture *et al.* (1987), with a few differences. Adenine residues at their positions 450, 464 and 469 are not found in our sequence (adjacent to position +52, +66 and +70 of Fig. 3). Also, they were not able to locate the *Eco*RI site at the left end of the fragment. We find 11 additional bases, starting at their position 6 (−394 to −404 of Fig. 3), that produce the proper *Eco*RI recognition sequence at the end of the fragment. Since Scripture *et al.* (1987) sequenced the control region primarily to obtain a match with our data, they did not sequence both strands of the entire region. Therefore, we feel that our sequence data are correct.

Horazdovsky & Hogg (1987) found three apparent S_1 mapped start sites for transcription of P_{FGH}. They suggest that the third site was an S_1 artifact due to cleavage at an A+T-rich region. This site is too close to the labeled end of our fragment for us to have detected it. Their site 2 was not detected in our experiments even at various incubation temperatures and levels of S_1 nuclease. The site mapped in this study is 11 bases closer to the right end of the cloned fragment than site 1 of Horazdovsky & Hogg. They were measuring a transcript 730 bases in length, which was too long to be compared directly with a DNA sequencing reaction. It is possible that a slight misalignment of their size standards due, for example, to the 11 bp discrepancy in their sequence data, could account for the difference in the estimated position of the start site.

Genetic data support our proposed transcription start site (Hendrickson & Flaherty, unpublished results). Deletion analysis of the *ara*FGH promoter in plasmid pWH54 has shown that there is no reduction in promoter strength when DNA is deleted between +1 and +97. Also, point mutations within the −10 region suggested by the S_1 data reduce promoter activity by 90%. These data strongly suggested that the site proposed in Figure 3 is, in fact, correct and there are no other *ara*FGH promoters in the region.

The induction of *ara*FGH transcription by arabinose appears to be quite slow. After seven minutes very little message is detected, yet the *ara*E and *ara*J promoters have reached maximum activity by this time (Stoner & Schleif, 1983). Since L-arabinose and AraC protein can directly stimulate transcription *in vitro*, it is unlikely that additional steps, such as metabolism of the sugar or induction of additional activating proteins, are required for induction.

Kosiba & Schleif (1982) cloned, sequenced and performed preliminary transcription studies on P_J. The S_1 nuclease mapping reported here locates the transcription start site on P_J to their position 226,

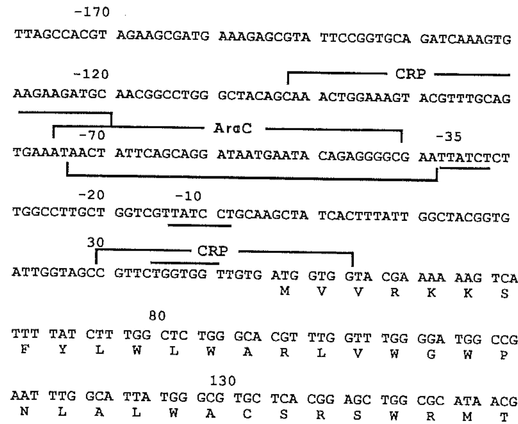


Figure 8. Transcription start point and protein binding sites of the *ara*J regulatory region. The nucleotide sequence of the *ara*J region is shown (Kosiba & Schleif, 1982). The putative −10 and −35 sequences are underlined. The potential ribosome binding site is indicated with a line drawn above at +37. The AUG start codon and the 35 amino acid residues derived from an open reading frame extending to the end of the fragment are shown beneath the nucleotide sequence. AraC protein protection and CRP–cAMP complex protection are shown by brackets.

which we now renumber as +1. A possible RNA polymerase recognition site (Siebenlist *et al.*, 1980) at the −10 and −35 regions, which are underlined in Figure 8, can be identified. Mapping the transcription start point with S_1 nuclease digestion yielded several start points in a 15 bp stretch of this area. This could result from a multiplicity of transcription start sites as seen in other promoters (Aiba *et al.*, 1981; Baker *et al.*, 1979; Carpousis *et al.*, 1982; Moran *et al.*, 1981), or it could be due to imprecise digestion by S_1 nuclease at the end of the transcript.

Although we have not identified a protein product or even a complete open reading frame downstream from the *ara*J promoter, it seems likely that such a product exists. Downstream 35 bases from the transcription start is a potential ribosome binding site (Shine & Dalgarno, 1974) and an AUG with an open reading frame for 35 amino acid residues at the end of the cloned DNA fragment. Scripture *et al.* (1987) originally suggested, on the basis of a comparison with other sugar transport systems, that an additional membrane protein might be found for the *ara*FGH system. However, overproduction of *ara*FGH increased arabinose transport function leading them to conclude that there are no other components of the system. It is possible that *ara*J codes for a component of the high-affinity system that is not produced in quantities that are limiting for transport. Other possibilities for *ara*J include involvement in the low-affinity transport system for which only *ara*E has been implicated, or that P_J is simply a remnant of evolutionary changes in the *E. coli* genome. The presence of a reasonably stable message, the open reading

frame and potential ribosome binding site suggests that a protein may be produced, but further work will be necessary to confirm this point. Although no unidentified, arabinose-inducible proteins have been observed in two-dimensional gel electrophoresis, the *ara*H product was not seen with this method (Kolodrubetz & Schleif, 1981*a*) so other arabinose regulated proteins also may exist.

We also performed DNase footprinting to determine the binding locations of the P_J regulatory proteins. As with P_{BAD} and P_E, a CRP binding site, albeit one showing weak binding, was found upstream from and adjacent to an AraC protein binding site. Downstream from the AraC site is an RNA polymerase binding site consensus sequence. A second CRP site was found downstream from the transcriptional start point, but more detailed studies will be required to determine whether this site plays any role *in vivo*. It is interesting to note that a CRP binding site has been bound in a similar position in the CRP operon (Aiba, 1983).

The locations of the CRP binding sites in the four *ara* promoters raise questions about the function of this protein in assisting initiation by RNA polymerase. In *ara*FGH and several other systems, the CRP binding site appears to be positioned so that CRP could make direct contact with RNA polymerase. Thus, if CRP and RNA polymerase are to make direct contact in *ara*E and *ara*BAD, the DNA must be tightly bent or the conformation of the protein–DNA complex must be unusual. On the other hand AraC protein could easily make direct contact with polymerase at *ara*E and *ara*BAD, but CRP might obstruct contact in the *ara*FGH promoter. Even if the DNA can be bent to accommodate direct contacts of both proteins with polymerase, it is difficult to envision structures in which the same surface regions of each protein make identical contacts with polymerase in the different promoters. Results from both *in vitro* and *in vivo* assays of *ara*BAD, *ara*E and *ara*FGH show that the activator protein adjacent to RNA polymerase appears to play a more direct or essential role in assisting its function. One possibility is that the closest activator protein makes direct contact with polymerase, and the more distal protein acts indirectly by contacting the first activator or by bending the DNA. Alternatively, DNA bending or protein contacts by the centrally placed activator may be required before the second activator may exert its effects on polymerase.

One speculative model that is consistent with the present data is that both DNA bending and protein–protein contacts are required for stimulation of RNA polymerase function. In addition, the *ara*FGH and *ara*BAD operons may have evolved to use the two activators in somewhat different ways. At P_{FGH}, CRP bends the DNA and, based on the overlap in binding sites, direct contact with RNA polymerase may occur within or near to the DNA-binding domain. Without this DNA bend, AraC protein alone cannot contact polymerase; however, the combined bending caused by CRP and AraC protein (unpublished results) may allow AraC directly to contact a second region of polymerase and further stimulate its activity. The combined bending may also help to destabilize the DNA, thus allowing strand opening during the isomerization step of initiation. At P_{BAD} the roles of the activators may be reversed. AraC contacts polymerase at a position similar to that of CRP at P_{FGH}, and a DNA bend is produced that allows CRP to contact polymerase at the second region to provide additional stimulation. At both promoters, direct contact between AraC and CRP may stabilize AraC in an inducing conformation, although such interactions are not required for protein binding to the DNA (Hendrickson & Schleif, 1984).

We thank S. Hahn for helpful discussions. This work was supported by grants from the National Institutes of Health and National Science Foundation to R.S. and by grant GM35917 from the National Institutes of Health to W.H.

References

Aiba, H. (1983). *Cell*, **32**, 141–149.

Aiba, H., Adhya, S. & de Crombrugghe, B. (1981). *J. Biol. Chem.* **256**, 11905–11910.

Baker, C. C., Herisse, J., Courtois, G., Galibert, F. & Ziff, E. (1979). *Cell*, **18**, 569–580.

Berg, O. G. & von Hippel, P. H. (1988). *J. Mol. Biol.* **200**, 709–723.

Brown, C. E. & Hogg, R. W. (1972). *J. Bacteriol.* **111**, 606–613.

Brunelle, A. & Schleif, R. (1989). *J. Mol. Biol.* **209**, 607–622.

Carpousis, A., Stefano, J. & Gralla, J. (1982). *J. Mol. Biol.* **157**, 619–633.

Clark, A. F. & Hogg, R. W. (1981). *J. Bacteriol.* **147**, 920–924.

Clarke, M. J. & Carbon, J. (1976). *Cell*, **9**, 91–99.

Dunn, T., Hahn, S., Ogden, S. & Schleif, R. (1984). *Proc. Nat. Acad. Sci., U.S.A.* **81**, 5107–5020.

Englesberg, E., Irr, J., Powers, J. & Lee, N. (1965). *J. Bacteriol.* **90**, 946–957.

Greenblatt, J. & Schleif, R. (1971). *Nature New Biol.* **233**, 166–170.

Hahn, S., Dunn, T. & Schleif, R. (1984). *J. Mol. Biol.* **180**, 61–72.

Hahn, S., Hendrickson, W. & Schleif, R. (1986). *J. Mol. Biol.* **188**, 355–367.

Harley, C. B. & Reynolds, R. P. (1987). *Nucl. Acids Res.* **15**, 2343–2361.

Hawley, D. & McClure, W. (1983). *Nucl. Acids Res.* **11**, 2237–2255.

Hendrickson, W. & Schleif, R. (1984). *J. Mol. Biol.* **178**, 611–628.

Hendrickson, W. & Schleif, R. (1985). *Proc. Nat. Acad. Sci., U.S.A.* **82**, 3192–3133.

Hentschel, C., Irminger, J., Bucher, P. & Birnstiel, M. (1980). *Nature (London)*, **285**, 147–151.

Horazdovsky, B. F. & Hogg, R. W. (1987). *J. Mol. Biol.* **197**, 27–35.

Horazdovsky, B. F. & Hogg, R. W. (1989). *J. Bacteriol.* **171**, 3053–3059.

Kolodrubetz, D. & Schleif, R. (1981*a*). *J. Mol. Biol.* **149**, 133–139.

Kolodrubetz, D. & Schleif, R. (1981*b*). *J. Mol. Biol.* **151**, 215–227.

Kolodrubetz, D. & Schleif, R. (1981*c*). *J. Bacteriol.* **148**, 472–479.
Kosiba, B. & Schleif, R. (1982). *J. Mol. Biol.* **156**, 53–66.
Lee, D. H. & Schleif, R. (1989). *Proc. Nat. Acad. Sci., U.S.A.* **86**, 476–480.
Lee, N. L., Gielow, W. O. & Wallace, R. G. (1981). *Proc. Nat. Acad. Sci., U.S.A.* **78**, 752–756.
Lee, N., Francklyn, C. & Hamilton, E. P. (1987). *Proc. Nat. Acad. Sci., U.S.A.* **84**, 8814–8818.
MacPherson, A. J. S., Jones-Mortimer, M. C. & Henderson, P. J. F. (1981). *Biochem. J.* **196**, 269–287.
Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982). *Molecular Cloning: A Laboratory Manual,* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Martin, K., Huo, L. & Schleif, R. (1986). *Proc. Nat. Acad. Sci., U.S.A.* **83**, 3654–3658.
Maxam, A. M. & Gilbert, W. (1980). *Methods Enzymol.* **65**, 499–559.
McKenney, K., Shimatake, H., Court, D., Schmeissner, U., Brady, C. & Rosenberg, M. (1982). In *Gene Amplification and Analysis: Analysis of Nucleic Acids* (Chirkjian, J. & Papas, T., eds), pp. 383–415, Elseviar North-Holland Press, Netherlands.
Moran, C., Jr, Lang, N., Banner, C., Haldenwang, W. & Losick, R. (1981). *Cell,* **25**, 783–791.
O'Connor, M. B. & Malamy, M. H. (1983). *J. Bacteriol.* **156**, 669–679.
Ogden, S., Haggerty, D., Stoner, C. M., Kolodrubetz, D. & Schleif, R. (1980). *Proc. Nat. Acad. Sci., U.S.A.* **77**, 3346–3350.
Sanger, F., Nicklen, S. & Coulson, A. R. (1977). *Proc. Nat. Acad. Sci., U.S.A.* **74**, 5463–5467.
Schleif, R. (1969). *J. Mol. Biol.* **61**, 275–279.
Schleif, R. & Favreau, A. (1982). *Biochemistry,* **21**, 778–782.
Schleif, R. & Wensink, P. (1981). *Practical Methods in Molecular Biology,* Springer-Verlag, New York.
Scripture, J. B., Voelker, C., Miller, S., O'Donnell, R. T., Polgar, L., Rade, J., Horazdovsky, B. F. & Hogg, R. W. (1987). *J. Mol. Biol.* **197**, 37–46.
Seibenlist, U., Simpson, R. B. & Gilbert, W. (1980). *Cell,* **20**, 269–281.
Sheppard, D. E. & Englesberg, E. (1967). *J. Mol. Biol.* **25**, 443–454.
Shine, J. & Dalgarno, L. (1974). *Proc. Nat. Acad. Sci., U.S.A.* **71**, 1342–1346.
Smith, B. & Schleif, R. (1978). *J. Biol. Chem.* **253**, 6931–6933.
Stoner, C. & Schleif, R. (1983). *J. Mol. Biol.* **171**, 369–381.
Wilcox, G., Meuris, P., Bass, R., Englesberg, E. (1974). *J. Biol. Chem.* **249**, 2946–2952.

Edited by P. von Hippel

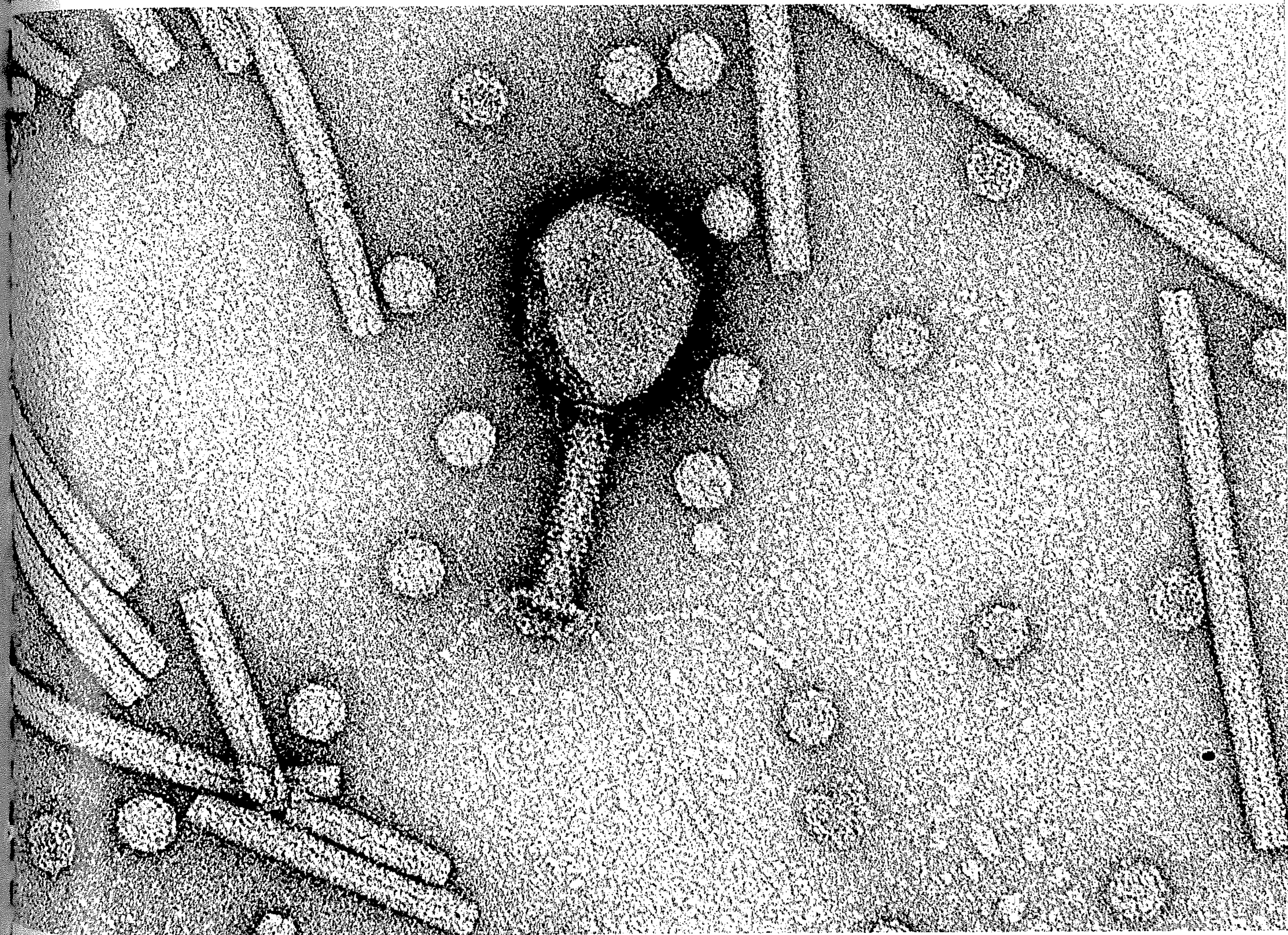
This issue completes Volume 215

Volume 215 Number 4 20 October 1990

BIO QC 173 .J59

PHYSICS LIBRARY
GILMER HALL, UNIV. OF VA
CHARLOTTESVILLE, VA. 22903

Journal of MOLECULAR BIOLOGY



ACADEMIC PRESS

Harcourt Brace Jovanovich, Publishers

JMOBAK 215 (4) 477–666 ISSN 0022-2836

For Reference
Not to be taken from this room

Journal of Molecular Biology

Editor-in-Chief
P. Wright
Department of Molecular Biology, Research Institute of Scripps Clinic
10666 N. Torrey Pines Road, La Jolla, CA 920327, U.S.A.

Assistant Editor
J. Karn
MRC Laboratory of Molecular Biology
Hills Road, Cambridge CB2 2QH, U.K.

Founding Editor
Sir John Kendrew

Consulting Editor
Sydney Brenner

Editors

A. R. Fersht, University Chemical Laboratory, Cambridge University, Lensfield Road, Cambridge CB2 1EW, U.K.
M. Gottesman, Institute of Cancer Research, College of Physicians & Surgeons of Columbia University, 701 W. 168th Street, New York, NY 10032, U.S.A.
P. von Hippel, Institute of Molecular Biology, University of Oregon, Eugene, OR 97403–1229, U.S.A.
R. Huber, Max-Planck-Institut für Biochemie, 8033 Martinsried bei München, Germany.
A. Klug, MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, U.K.

Associate Editors

C. R. Cantor, Human Genome Center, Donner Laboratory, Lawrence Berkeley Laboratory, University of California, Berkeley, CA 94720, U.S.A.
N.-H. Chua, The Rockefeller University, 1230 York Avenue, New York, NY 10021, U.S.A.
F. E. Cohen, Department of Pharmaceutical Chemistry, School of Pharmacy, University of California, San Francisco, CA 94143–0446, U.S.A.
D. J. DeRosier, Rosenstiel Basic Medical Sciences Research Center, Brandeis University, Waltham, MA 02254, U.S.A.
W. A. Hendrickson, Department of Biochemistry & Molecular Biophysics, College of Physicians & Surgeons of Columbia University, 630 West 168th Street, New York, NY 10032, U.S.A.
I.B. Holland, Institut de Génétique et Microbiologie, Bâtiment 409, Université de Paris XI, 91405 Orsay Cedex 05, France.
B. Honig, Department of Biochemistry & Molecular Biophysics, College of Physicians & Surgeons of Colombia University, 630 West 168th Street, New York, NY 10032, U.S.A.
H. E. Huxley, Rosenstiel Basic Medical Sciences Research Center, Brandeis University, Waltham, MA 02254, U.S.A.
V. Luzzati, Centre de Génétique Moléculaire, Centre National de la Recherche Scientifique, 91 Gif-sur-Yvette, France.
J. L. Mandel, Laboratoire de Génétique Moléculaire des Eucaryotes du CNRS, Institut de Chimie Biologique, Faculté de Médecine, 11 Rue Humann, 67085 Strasbourg Cedex, France.
B. Matthews, Institute of Molecular Biology, University of Oregon, Eugene, OR 97403–1229, U.S.A.
J. H. Miller, Department of Microbiology, University of California, 405 Hilgard Avenue, Los Angeles, CA 90024, U.S.A.
M. F. Moody, School of Pharmacy, University of London, 29/39 Brunswick Square, London WC1N 1AX, U.K.
T. Richmond, Institut für Molekularbiologie und Biophysik, Eidgenössische Technische Hochschule, Hönggerberg, CH 8093 Zurich, Switzerland.
R. Schleif, Biology Department, Johns Hopkins University, Charles & 34th Streets, Baltimore, MD 21218, U.S.A.
N. L. Sternberg, Central Research & Development Department, E. I. du Pont Nemours & Company, Wilmington, DE 19898, U.S.A.
K. R. Yamamoto, Department of Biochemistry and Biophysics, School of Medicine, University of California, San Francisco, CA 94143–0448, U.S.A.
M. Yanagida, Department of Biophysics, Faculty of Science, Kyoto University, Sakyo-Ku, Kyoto 606, Japan.

Editorial Office

G. Harris, Journal of Molecular Biology, 10d St Edwards Passage, Cambridge CB2 3PJ, U.K.

JOURNAL OF MOLECULAR BIOLOGY: ISSN 0022–2836. Volumes 211–216, 1990, published twice a month on the 5th and 20th by Academic Press at 24–28 Oval Road, London NW1 7DX, England. Annual subscription price including postage: £768 U.K. and U.S.$1464 overseas. Personal subscription rate: £233 U.K. and U.S. $350 overseas. Subscription orders should be sent to Academic Press Limited, Foots Cray, Sidcup, Kent DA14 5HP, U.K. (Tel: 081–300 3322). Send notices of changes of address to the publisher at least 6–8 weeks in advance, including both old and new addresses.

Second class postage rate paid at Jamaica, NY 11431, U.S.A.

Air freight and mailing in the U.S.A. by Publications Expediting Inc., 200 Meacham Avenue, Elmont, NY 11003, U.S.A.

U.S.A. POSTMASTERS: send change of addresses to JOURNAL OF MOLECULAR BIOLOGY, c/o Publications Expediting, Inc., 200 Meacham Avenue, Elmont, NY 11003, U.S.A.
Printed in U.K.