JOURNAL OF BACTERIOLOGY, Dec. 1991, p. 7765–7771
0021-9193/91/247765-07$02.00/0

Vol. 173, No. 24

Mapping, Sequence, and Apparent Lack of Function of *araJ*, a Gene of the *Escherichia coli* Arabinose Regulon

THADD REEDER[1,2] AND ROBERT SCHLEIF[1]*

Department of Biology, Johns Hopkins University, Baltimore, Maryland 21218,[1] and Department of Biochemistry, Brandeis University, Waltham, Massachusetts 02254[2]

Received 9 May 1991/Accepted 6 October 1991

We report the mapping, sequencing, and study of the physiological role of the fourth arabinose-inducible operon from *Escherichia coli*, *araJ*. It is located at 9 min on the chromosome and codes for a single 42-kDa protein that shows no significant homology to other known proteins. Destruction of the chromosomal *araJ* gene does not detectably affect either of the two arabinose transport systems, the ability of cells to grow on arabinose, or the induction kinetics of the *araBAD* operon, and thus the physiological role of AraJ, if any, remains unknown. We have also found a long open reading frame upstream of *araJ*. The sequence of this upstream open reading frame was found to be identical to the previously reported sequence of the *sbcC* gene (I. S. Naom, S. J. Morton, D. R. F. Leach, and R. G. Lloyd, Nucleic Acids Res. 17:8033–8044, 1989). The carboxyl region of SbcC has an amino acid sequence consistent with this region of SbcC forming an extended alpha-helical coiled-coil.

The enteric bacterium *Escherichia coli* can utilize L-arabinose as a sole carbon source. In the presence of arabinose, expression from the three known *ara* operons is induced up to 300-fold. These are the *araBAD* operon (5, 6), which codes for proteins involved in the catabolism of arabinose, and the *araE* (27) and *araFGH* (2, 17, 31) operons, which code for the low- and high-affinity arabinose transport systems, respectively. In addition to the promoters of these three well-characterized operons, there exists a fourth promoter in *E. coli* that is induced by arabinose. This promoter was originally cloned by Kosiba and Schleif (18) and at that time was misidentified as the *araFGH* promoter. Subsequent work (11, 13) has identified the actual *araFGH* promoter, and the arabinose-inducible promoter cloned by Kosiba and Schleif (18) is now known as p_J. Recent work (11, 18) has shown that arabinose induces substantial transcription from p_J in vivo and in vitro. In vivo, as well as in vitro (11, 18), transcription from p_J is dependent upon the arabinose-specific transcriptional regulator protein AraC and the global catabolite gene activator protein CAP, as is transcription from the *araBAD*, *araE*, and *araFGH* operons (7, 8, 32).

Since all known *ara* mutations are confined to the *araBAD*, *araC*, *araE*, and *araFGH* operons, we were interested to learn whether p_J drives an intact gene and whether its product plays any detectable role in vivo. In particular, we wished to determine whether the *araJ* product was a fourth component in the high-affinity arabinose transport system, since most analogous systems contain four protein components (1, 34).

Here we report the mapping of the *araJ* gene to 9 min on the *E. coli* chromosome and the cloning and sequencing of the *araJ* gene and surrounding regions of the chromosome. We also show that insertion/deletion mutations in *araJ* have no detectable effect on the ability of the bacteria to grow on arabinose and to induce the *araBAD* operon.

MATERIALS AND METHODS

Media, strains, and general methods. Media and general methods have been described (21, 33). Arabinose isomerase assays were performed as described previously (33). Restriction and modifying enzymes were from New England Biolabs or Boehringer Mannheim. All DNA sequencing was performed with the dideoxyribonucleotide chain termination method (30) by using the Sequenase kit (37) from U.S. Biochemical according to the manufacturer's instructions. Synthetic oligodeoxyribonucleotides were made on an Applied Biosystems 381A DNA synthesizer by using chemicals from Applied Biosystems, American Bionetics, or Chruachem. *Taq* polymerase was from Perkin-Elmer Cetus. All chemicals were reagent grade. Strain JM101 (23) was used for all routine transformations and production of single-stranded DNA. Other strains used in this work are listed in Table 1.

Chromosomal mapping of *araJ*. The chromosomal position of *araJ* was initially mapped by whole-genome Southern transfers. *E. coli* genomic DNA was cut with *Sfi*I or *Not*I, fractionated by pulsed-field gel electrophoresis, and transferred to membranes (kindly supplied by Cassandra Smith, University of Berkeley) (35). This was probed with the 735-bp fragment cloned by Kosiba and Schleif (18) containing p_J. The hybridization was performed at 42°C for 10 h in 10 ml of hybridization buffer (600 mM NaCl, 40 mM NaH_2PO_4 [pH 7.4], 5 mM EDTA, 50% formamide, 10% dextran sulfate, 1% sodium dodecyl sulfate, 0.5% dehydrated milk [Carnation], 300 μg of calf thymus DNA per ml, and 500 ng of probe [10^7 cpm/μg]). After hybridization, the blot was sequentially washed at 24°C in 2× SSC (33), 0.5× SSC, and 0.1× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate) and then autoradiographed overnight.

Cloning and sequencing of *araJ*. The cloning of the complete *araJ* gene was done by Robert Hogg's laboratory (12a), where an *E. coli* library in bacteriophage lambda was probed with the 735-bp fragment cloned by Kosiba and Schleif (18) containing the *araJ* promoter. Three positive clones were selected and shown by restriction mapping to contain overlapping copies of the same piece of genomic DNA; these three clones were sent to us for further analysis.

One of the lambda clones was digested with *Pvu*I, and the 1.9- and 2.5-kb fragments that contained p_J and flanking sequences were isolated. The fragments were filled out with DNA polymerase (Klenow fragment) and then ligated into

* Corresponding author.

TABLE 1. *E. coli* strains used in this study

Strain	Genotype	Source or reference
CW2513	K-12, wild type	R. Hogg (17)
CW2549	*araE201*	Derivative of CW2513; R. Hogg (17)
CW2553	*araE201* Δ*araFGH*::Kan^r	Derivative of CW2549; R. Hogg (17)
TRΔJ7	*araE201* Δ*araJ*::Kan^r	CW2549 with *araJ* insertion/deletion (this study)
TRΔJ17	Δ*araJ*::Kan^r	P1 transduction of Δ*araJ*::Kan^r from TRΔJ7 to CW2513 (this study)
TRΔF1	Δ*araFGH*::Kan^r	P1 transduction of Δ*araFGH*::Kan^r from CW2553 to CW2513 (this study)

the *Sma*I site in the polylinker region of pUC19. The 1.9-kb clone was digested at the *Eco*RI and *Hin*dIII sites present on the polylinker and cloned into pES51 (15), which contains the M13 origin of replication. To speed sequencing, the 2.5-kb clone was further divided by cutting at an internal *Eco*RI site as well as the flanking *Eco*RI site in the polylinker; the resulting 1.0- and 1.5-kb fragments were then ligated into the *Eco*RI site in pES51.

Both strands of the cloned DNA were sequenced by using synthetic oligodeoxyribonucleotides as primers. Sequences at the junctions of the subclones were confirmed by using an asymmetric polymerase chain reaction to generate single-stranded DNA from chromosomal DNA and then by using the single-stranded DNA as a template for Sequenase (9).

DNA and protein sequence homology searches were conducted with the FASTA program (28) at a ktup of 2. The data bases searched included GenBank, GenPept, and Swiss-Prot (November 1990).

Construction of a Δ*araJ*::Kan^r strain. We used a gene replacement method (10) to create strain TRΔJ7, which has the kanamycin resistance gene inserted into the middle of the *araJ* gene along with a deletion of the central portion of the *araJ* gene.

The polymerase chain reaction was used to generate two fragments of the *araJ* gene (Fig. 1). Each of the four primers used in the two polymerase chain reactions had a restriction site built into its 5′ terminus. The first fragment spanned from −158 to +454 (with respect to the *araJ* transcriptional start site) and had a *Kpn*I site at the 5′ end and a *Bam*HI site at the 3′ end. The second fragment spanned from +622 to +1257 (i.e., to the end of the *araJ* open reading frame [ORF]) and had a *Bam*HI site at the 5′ end and a *Hin*dIII site at the 3′ end. These two fragments were sequentially cloned into the polylinker region of pMAK705 (a gift of Sidney Kushner, University of Georgia), which has a temperature-sensitive replicon and carries the chloramphenicol resistance gene. The kanamycin resistance gene from pUC-4K (Pharmacia) was then cloned into the *Bam*HI site separating the two fragments of the *araJ* gene to give plasmid pJKJ. Strain CW2549 (*araE201*) was transformed with pJKJ and plated at the restrictive temperature of 44°C. Colonies that were Kan^r Chl^r had undergone a recombination event between the plasmid and the host chromosome to generate a cointegrate. The frequency of this recombination event was 1.5×10^{-4}, very close to the expected frequency of 2.0×10^{-4} for recombination between these lengths of homologous DNA (10).

To resolve the cointegrates, single colonies were picked and grown for two cycles in YT (33) plus kanamycin (100 μg/ml) at 30°C and then for one more cycle in YT plus kanamycin at 44°C before being plated onto kanamycin plates. Of 88 colonies spot tested, 8 were Kan^r Chl^s at 30°C, indicating that the plasmid along with the wild-type *araJ* gene had been excised from the chromosome by a second recombination event and eventually lost from the cell.

To verify that these colonies had undergone gene replacement, we used the polymerase chain reaction directly on colonies (29). The primers (generally 20-mers) were used at a concentration of approximately 7×10^{-7} M (~100 ng per 25 μl of reaction mixture), and the hybridization temperature was set at 63°C. By using various sets of primers and control DNA, we confirmed that the chromosomal *araJ* gene had indeed been replaced with the larger Δ*araJ*::Kan^r insertion and that the plasmid and wild-type *araJ* gene had been lost from the cell (data not shown).

Nucleotide sequence accession number. All nucleotide and protein sequences have been deposited in GenBank under the accession number M64787.

RESULTS AND DISCUSSION

Chromosomal mapping of *araJ*. The 735-bp fragment containing the *araJ* promoter cloned by Kosiba and Schleif (18) hybridized to a single band in each of the chromosomal digests on the Southern transfer supplied by Cassandra Smith (35). Analysis of the hybridized band pattern by C. Smith located p_J to approximately 10 min on the *E. coli* chromosome. To locate *araJ* more precisely, we sent the restriction map of our entire 4.4-kb clone to Yugi Kohara at Nagoya University, who has published a restriction map of the whole *E. coli* chromosome (16). His analysis revealed two *Pvu*I restriction sites overlooked in his original mapping of the genome and definitively located our 4.4-kb clone at 9 min on the *E. coli* chromosome. This is between *aroC* and *phoB* on the genetic map, at 423 kb from the origin, overlapping clones 9G9 and 6A12 on the Kohara map.

AraJ sequence. Figure 2 shows the organization and location of our clones and ORFs. Figure 3 shows the nucleotide sequence of the *araJ* promoter and transcript as well as the deduced protein sequence of AraJ. Beginning with the second ATG after transcriptional initiation is a single ORF of 395 codons coding for a protein with a potential molecular size of 41,898 Da. The putative AraJ protein is hydrophobic over the major part of its sequence, as shown by the hydropathicity plot (20) in Fig. 4. The only regions in AraJ that show extended hydrophilicity are between residues 178 and 197 and at the carboxyl terminus. The hydrophobicity profile of AraJ is unlike those of the other arabinose-induced proteins and is given here because it can provide a useful fingerprint for identifying related proteins (34). Searches of both the DNA and protein data bases containing prokaryotic and eukaryotic sequences failed to detect any sequences with significant homology to AraJ or its coding DNA sequence.

The amino terminus of AraJ contains several features that are consistent with this region serving as a signal peptide. These include two positively charged residues (lysines) within the first five amino acids and a stretch of 15 amino acids between the last lysine and the next charged residue (Glu-19) that are almost exclusively hydrophobic. The only

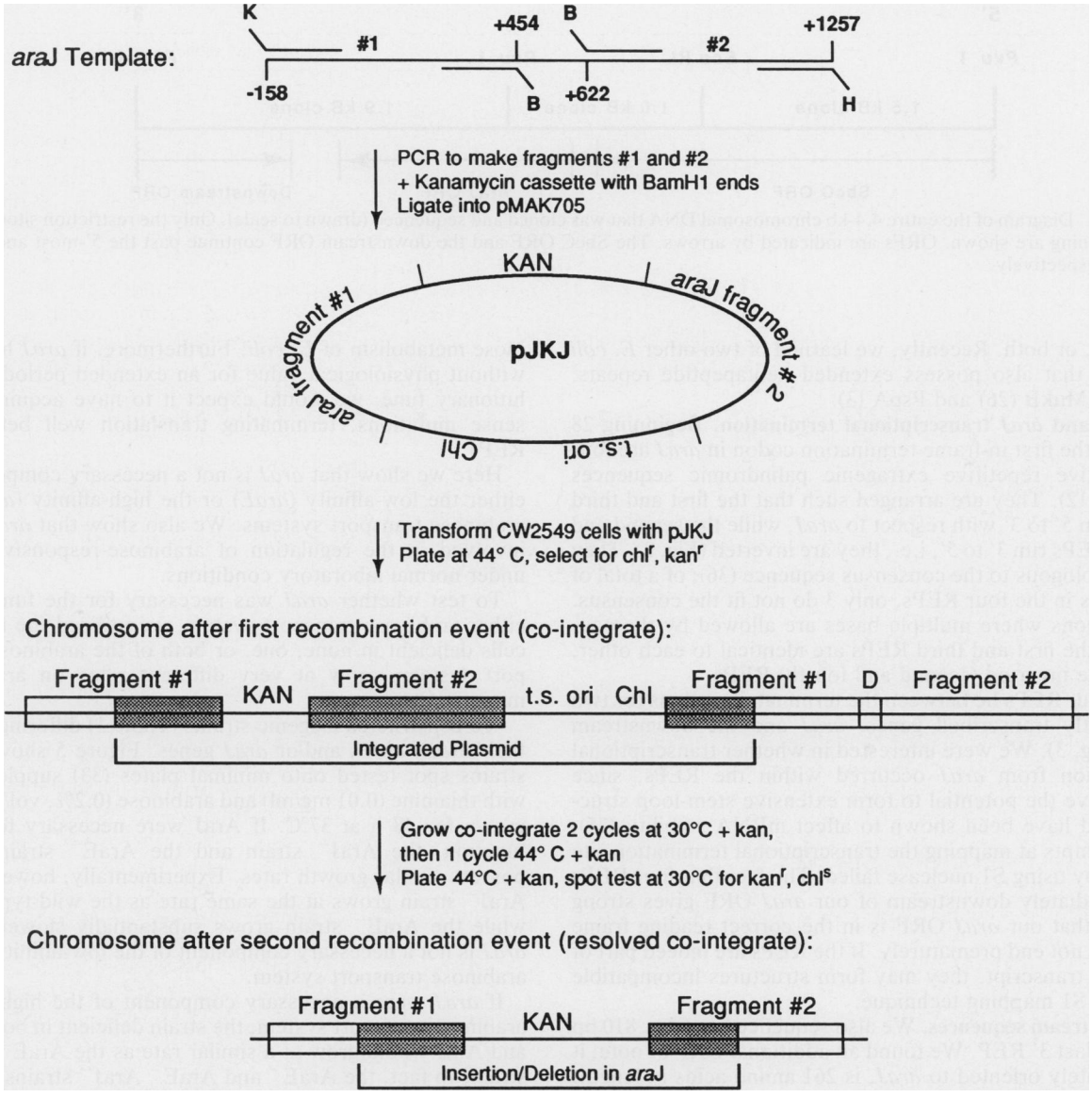


FIG. 1. Schematic of the Δ*ara*::Kan^r strain construction. All numbering is with respect to the major *araJ* transcriptional start site. Abbreviations: K, *Kpn*I site; B, *Bam*HI site; H, *Hin*dIII site. Neither the plasmid nor the chromosomes are drawn to scale. The region of the cointegrate chromosome marked D represents the region of *araJ* (bases 455 to 621) lying between fragment 1 and fragment 2 that was eventually deleted from the chromosome. PCR, polymerase chain reaction.

exceptions to this hydrophobicity are serine, threonine, and glycine residues, and these are located in positions commonly occupied by these residues in the hydrophobic domain of signal peptides (4).

***sbcC* lies upstream of *araJ* and contains a myosinlike repeat.** Upstream of *araJ* we found a long ORF beginning at our 5′-most clone and consisting of 671 codons. This terminates at position −75 with respect to the *araJ* transcriptional start site (Fig. 2 and 3). When translated, this upstream ORF is identical to the carboxyl-terminal 671 amino acids of the 1,048 amino acids of SbcC, as reported by Naom et al. (24). Thus, *sbcC*, whose complete sequence encodes 1,048 amino acids, lies immediately upstream of *araJ* on the *E. coli* chromosome.

While trying to determine the identity of the upstream SbcC sequence, which was not in the GenPept (28) data base we initially used, we noticed that the amino acid sequence possesses a strong pattern of hydrophobic amino acids, 480 residues long, spaced alternatively three and then four amino acid residues apart. This pattern of hydrophobicity, which is called the heptapeptide repeat, is characteristic of a coiled-coil alpha-helix structure. Such structures are prominent in myosin, tropomyosin, and paramyosin; indeed, the protein with the highest homology to SbcC is myosin. This property of SbcC had not been noticed before, perhaps because its sequence had been compared only with those of prokaryotic proteins.

The extended coiled-coil structure plus the fact that SbcC may also contain a nucleotide binding site (24, 38) warrants speculation that SbcC may have a myosinlike structure,

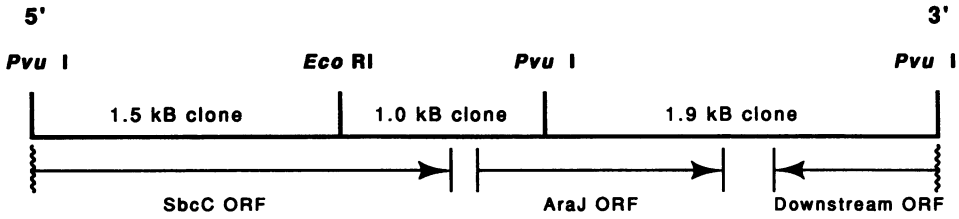


FIG. 2. Diagram of the entire 4.4-kb chromosomal DNA that was cloned and sequenced (drawn to scale). Only the restriction sites relevant to the cloning are shown. ORFs are indicated by arrows. The SbcC ORF and the downstream ORF continue past the 5'-most and 3'-most clones, respectively.

function, or both. Recently, we learned of two other *E. coli* proteins that also possess extended heptapeptide repeats, namely, MukB (26) and PspA (3).

REPs and *araJ* transcriptional termination. Beginning 28 bp after the first in-frame termination codon in *araJ* are four consecutive repetitive extragenic palindromic sequences (REPs) (12). They are arranged such that the first and third REPs run 5' to 3' with respect to *araJ*, while the second and fourth REPs run 3' to 5', i.e., they are inverted (Fig. 3). They are homologous to the consensus sequence (36); of a total of 132 bases in the four REPs, only 3 do not fit the consensus. At positions where multiple bases are allowed by the consensus, the first and third REPs are identical to each other, as are the inverted (second and fourth) REPs.

The four REPs lie between the termination codons for two divergently transcribed genes, *araJ* and the downstream ORF (Fig. 3). We were interested in whether transcriptional termination from *araJ* occurred within the REPs, since REPs have the potential to form extensive stem-loop structures and have been shown to affect mRNA stability (25). Our attempts at mapping the transcriptional termination site of *araJ* by using S1 nuclease failed. The fact that four REPs lie immediately downstream of our *araJ* ORF gives strong support that our *araJ* ORF is in the correct reading frame and does not end prematurely. If the REPs are indeed part of the *araJ* transcript, they may form structures incompatible with the S1 mapping technique.

Downstream sequences. We also sequenced another 810 bp past the last 3' REP. We found an additional ORF of note; it is oppositely oriented to *araJ*, is 261 amino acids in length, and terminates at 25 bp before the end of the fourth REP (Fig. 2 and 3). Searches of the protein data bases with the deduced protein sequence of this downstream ORF revealed significant homology with only one protein, namely, the XylR protein of *Bacillus subtilis*, which is involved in regulation of the xylose operon (19). The identity was 24.7% over 215 amino acids.

***araJ* function.** Although many mutants in the other arabinose-responsive operons have been found, no mutations have ever, to our knowledge, mapped to *araJ*. The facts that arabinose induces the substantial transcription of *araJ* messenger (11) and that this transcription is regulated by CAP and AraC (11, 18) are, however, consistent with *araJ* actually being translated and serving some purpose in the arabinose metabolism of *E. coli*. Furthermore, if *araJ* had been without physiological value for an extended period of evolutionary time, we would expect it to have acquired nonsense mutations, terminating translation well before the REPs.

Here we show that *araJ* is not a necessary component of either the low-affinity (*araE*) or the high-affinity (*araFGH*) arabinose transport systems. We also show that *araJ* is not involved in the regulation of arabinose-responsive genes under normal laboratory conditions.

To test whether *araJ* was necessary for the function of either arabinose transport system, we utilized the fact that cells deficient in none, one, or both of the arabinose transport systems grow at very different rates on arabinose-minimal plates.

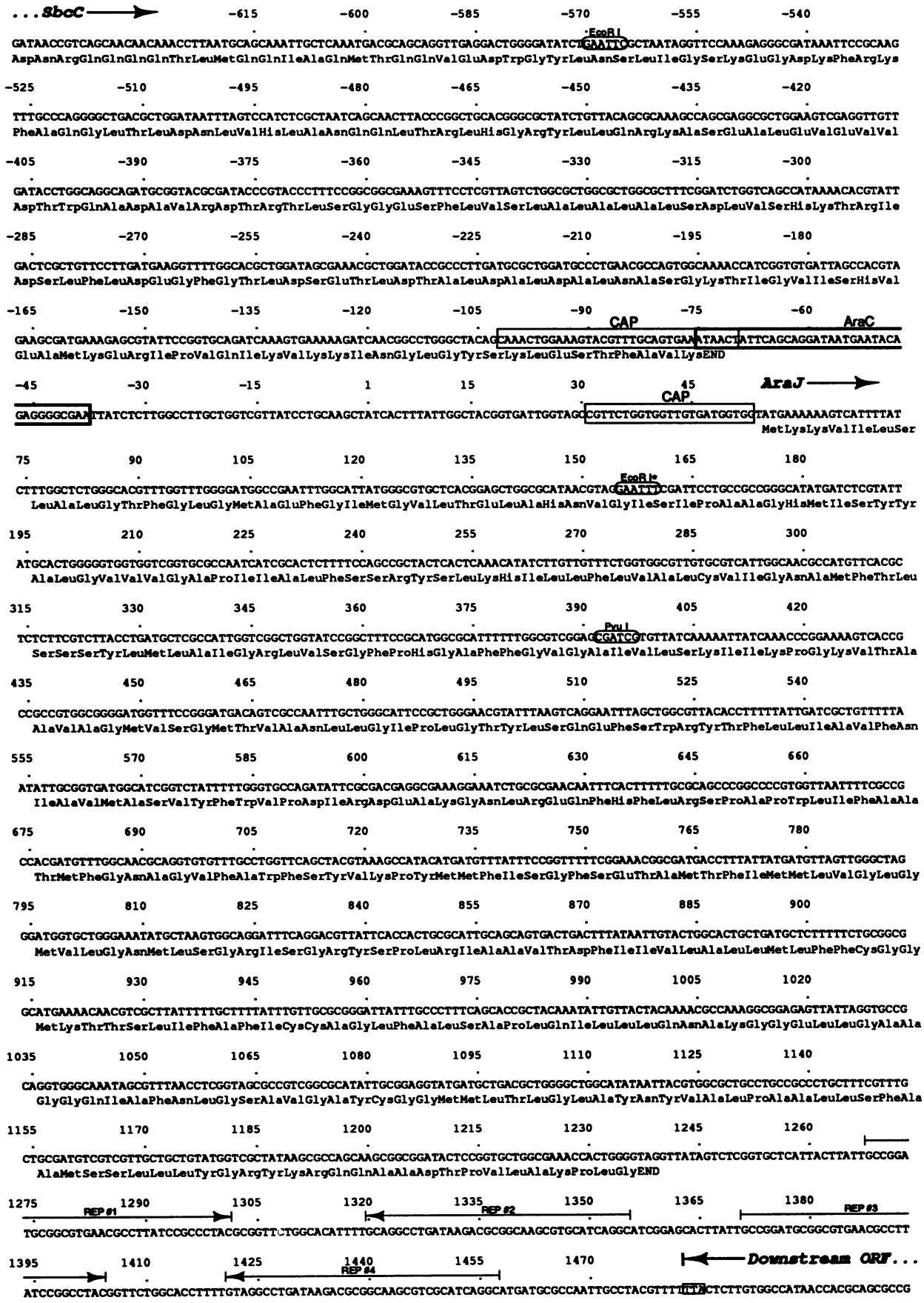
We constructed isogenic strains (Table 1) differing only in their *araE*, *araF*, and/or *araJ* genes. Figure 5 shows these strains spot tested onto minimal plates (33) supplemented with thiamine (0.01 mg/ml) and arabinose (0.2%, vol/vol) and grown for 48 h at 37°C. If AraJ were necessary for AraE function, the AraJ$^-$ strain and the AraE$^-$ strain would possess similar growth rates. Experimentally, however, the AraJ$^-$ strain grows at the same rate as the wild-type strain while the AraE$^-$ strain grows substantially slower. Thus, *araJ* is not a necessary component of the low-affinity (*araE*) arabinose transport system.

If *araJ* were a necessary component of the high-affinity arabinose transport system, the strain deficient in both AraE and AraJ would grow at a similar rate as the AraE$^-$ AraF$^-$ strain. In fact, the AraE$^-$ and AraE$^-$ AraJ$^-$ strains grew at similar rates, whereas the AraE$^-$ AraF$^-$ strain showed no detectable growth. Hence, *araJ* is not a necessary component of the high-affinity (*araFGH*) arabinose transport system.

The fact that *araJ* is not a necessary component of the low-affinity arabinose transport system was not surprising, since all low-affinity mutants have mapped to only the *araE* locus. Originally, in 1988, we had thought that *araJ* might be a component of the arabinose high-affinity transport system, on the basis that at that time all known high-affinity transport systems appeared to possess four components (1, 13), while only three were known for the arabinose system, the products of the *araFGH* operon (34).

Subsequent work (14) indicates that the *araFGH* gene

FIG. 3. Summary of sequence and footprinting (11) data. The DNA sequence is numbered relative to the *araJ* transcriptional start site. The carboxyl-terminal 191 amino acids of SbcC and the entire deduced protein sequence of AraJ are shown. The four repetitive extragenic elements after *araJ* are identified by small arrows. The termination codon of the downstream ORF is boxed. The *Eco*RI and *Pvu*I sites used in the cloning of *araJ* are shown enclosed in ovals. The original *araJ* promoter cloned by Kosiba and Schleif spanned the region from the *Eco*RI site at −570 to the *Eco*RI (star) site at +160. Regions on the *araJ* promoter that were protected from DNase I digestion by AraC are shown in thick boxes; regions similarly protected by CAP are shown in thin boxes (11).


...SbcC
EcoR I
CAP
AraC
AraJ
CAP
EcoR I*
Pvu I
REP #1
REP #2
REP #3
REP #4
Downstream ORF...

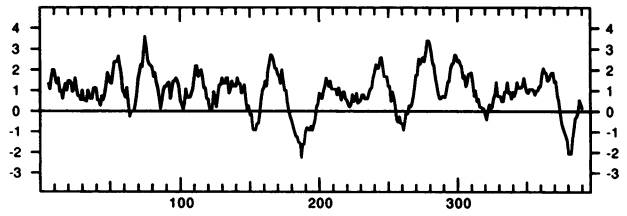


FIG. 4. Kyte-Doolittle hydropathicity plot (20) of AraJ. Analysis was performed with the DNA Stryder program (22) with an averaging window of 11 amino acids. The more positive a value, the greater the average hydrophobicity; the more negative a value, the greater the average hydrophilicity.

products alone are necessary to reconstitute the arabinose high-affinity transport system in vivo. This was done by deleting the chromosomal *araFGH* operon and testing whether high-affinity transport could be rescued by a plasmid containing different combinations of *araF*, *araG*, and *araH* genes driven by a foreign promoter. Arabinose high-affinity transport was restored to wild-type levels only when all three genes were on the plasmid. Nevertheless, there was still the possibility that *araJ* was needed in substoichiometric amounts in the high-affinity transport system, since *araJ* was not disabled in any of the strains used. The experiment shown in Fig. 5, however, clearly shows that *araJ* is not needed, in any amount, for the high-affinity transport system to function. Furthermore, since there are no other known arabinose-responsive genes in *E. coli*, it is likely that the arabinose high-affinity transport system is truly composed of only three proteins.

A second test for *araJ* function was whether it was involved in the regulation of the cell response to arabinose. This was done by examining the kinetics of induction of the *araBAD* operon by using the arabinose isomerase assay, which measures the activity of the product of *araA*.

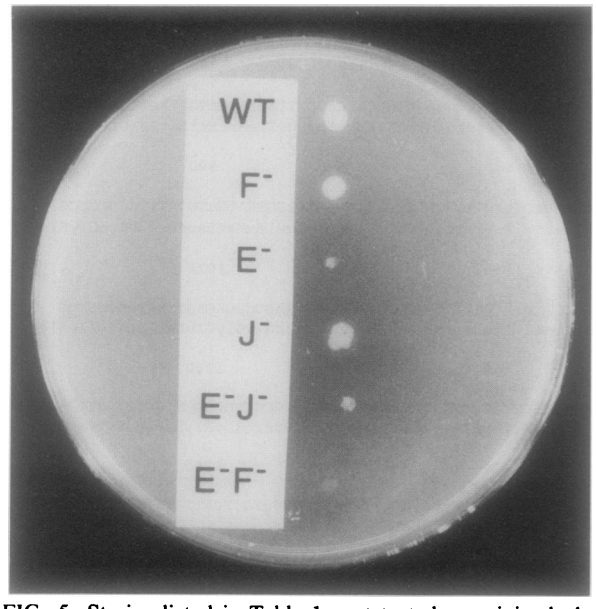


FIG. 5. Strains listed in Table 1 spot tested on minimal plates supplemented with thiamine (0.01 mg/ml) and L-arabinose (0.2%, vol/vol). Strain abbreviations: WT, CW2513; F⁻, TRΔF1; E⁻, CW2549; J⁻, TRΔJ17; E⁻ J⁻, TRΔJ7; E⁻ F⁻, CW2553.

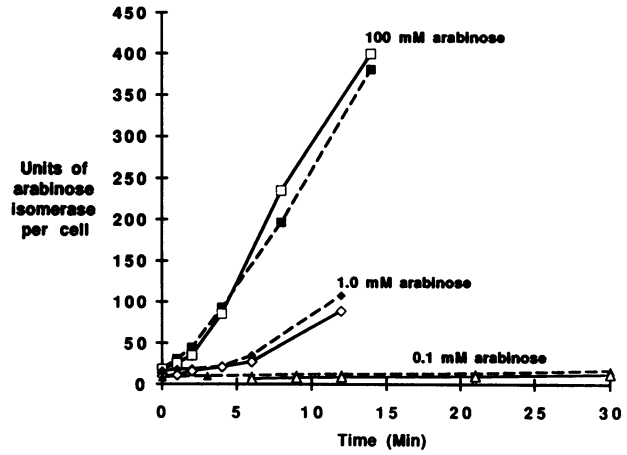


FIG. 6. Arabinose isomerase induction in strains TRΔJ7 (AraJ⁻; open symbols, solid line) and CW2549 (AraJ⁺; solid symbols, markers, dashed line). Cells were grown in M10 media–0.2% glycerol to mid-log phase. L-Arabinose was then added to 100 mM (squares), 1 mM (diamonds), or 0.1 mM (triangles). Aliquots were taken at selected time points, and the amount of arabinose isomerase present was assayed as described (33). Units of arabinose isomerase are per cell and are corrected for cell growth. The first two time points of the assay for TRΔX-7 at 0.1 mM arabinose are missing as a result of experimental error.

Cells were grown for at least five generations in minimal M10 media (33) supplemented with thiamine (0.01 mg/ml) and glycerol (0.2%, vol/vol). Arabinose was then added to a variable final concentration, and, at a series of times thereafter, aliquots of cells were assayed for the amount of *araA* activity. Figure 6 shows that the induction kinetics as well as the basal and induced levels of isomerase are identical in the AraJ⁺ (CW2549) and the AraJ⁻ (TRΔJ7) strains at all of the arabinose concentrations tested. Hence, *araJ* is not involved in arabinose regulation under the conditions used for this assay.

This leaves *araJ* a gene without an obvious function. It is unlikely that *araJ* is involved in chemotaxis since *E. coli* does not detectably respond to arabinose by chemotaxis. Neither is it likely that *araJ* is involved in arabinose catabolism; there is no evidence that a catabolic pathway other than the known one exists in *E. coli*. Perhaps the most attractive hypothesis for *araJ* function is that it is involved in either the transport or processing of arabinose polymers, which are widespread in nature. The probable signal peptide on AraJ is consistent with this hypothesis, since both transport (outer membrane) or processing (periplasmic space) would require passage of AraJ across the inner membrane.

Our major results are summarized as follows. (i) We have mapped and sequenced an arabinose-responsive gene, *araJ*, that is located at 9 min on the *E. coli* chromosome. (ii) The potential protein product of *araJ* is 395 amino acids in length and largely hydrophobic in composition. AraJ likely contains a signal peptide and does not have significant homology with other known proteins. (iii) Destruction of the chromosomal *araJ* gene has no detectable phenotype under the conditions assayed in this laboratory. (iv) *sbcC* lies immediately upstream of *araJ*. At least 480 of the amino acids in the carboxyl region of SbcC are organized into heptapeptide repeats. This organization is consistent with the carboxyl region of SbcC forming an alpha-helical coiled-coil. (v) An unidentified ORF lies immediately downstream of *araJ*; its

putative protein product has significant homology with the XylR protein of *B. subtilis*.

ACKNOWLEDGMENTS

We thank Bob Hogg for the lambda clones and strains, Cassandra Smith for the Southern transfer and analysis of the resulting Southern blot, Sidney Kushner and Brian Washburn for plasmid pMAK 705, Yuji Kohara and Ken Rudd for helping to map *araJ*, and Sota Hiraga for personal communications. We also thank John Carra, Guy Duffaud, and Sudha Iyer for discussions and suggestions and Saul Roseman and Robert Maier for reviewing this manuscript.

REFERENCES

1. **Ames, G. F.-L.** 1988. Structure and mechanism of bacterial periplasmic transport systems. J. Bioenerg. Biomembr. **20**:1–18.
2. **Brown, L. E., and R. W. Hogg.** 1972. A second transport system of L-arabinose in *Escherichia coli* controlled by the *araC* gene. J. Bacteriol. **111**:606–613.
3. **Brissette, J. L., L. Weiner, T. L. Ripmaster, and P. Model.** 1991. Characterization and sequence of the *Escherichia coli* stress-induced *psp* operon. J. Mol. Biol. **220**:35–48.
4. **Duffaud, G. D., S. K. Lehnhardt, P. E. March, and M. Inouye.** 1985. Structure and function of the signal peptide. Curr. Top. Membr. Transp. **24**:65–103.
5. **Englesberg, E.** 1961. Enzymatic characterization of 17 L-arabinose negative mutants of *Escherichia coli*. J. Bacteriol. **81**:996–1006.
6. **Englesberg, E., R. L. Anderson, R. Weinberg, N. Lee, P. Hoffee, G. Huttenhauer, and H. Boyer.** 1962. L-Arabinose-sensitive L-ribulose 5-phosphate 4-epimerase-deficient mutants of *Escherichia coli*. J. Bacteriol. **84**:137–146.
7. **Englesberg, E., J. Irr, and N. Lee.** 1965. Positive control of enzyme synthesis by gene C in the L-arabinose system. J. Bacteriol. **90**:946–957.
8. **Greenblatt, J., and R. Schleif.** 1971. Arabinose C protein: regulation of the arabinose operon in vitro. Nature (London) New Biol. **233**:166–170.
9. **Gyllensten, U.** 1989. Direct sequencing of in vitro amplified DNA, p. 45–60. *In* H. A. Erlich (ed.), PCR technology. Stockton Press, New York.
10. **Hamilton, C. M., M. Aldea, B. K. Washburn, P. Babitze, and S. Kushner.** 1985. New method for generating deletions and gene replacements in *Escherichia coli*. J. Bacteriol. **171**:4617–4622.
11. **Hendrickson, W., C. Stoner, and R. Schleif.** 1990. Characterization of the *Escherichia coli araFGH* and *araJ* promoters. J. Mol. Biol. **215**:497–510.
12. **Higgens, C. F., G. F.-L. Ames, W. L. Barnes, J. M. Clement, and M. Hofnung.** 1982. A novel intercistronic regulatory element of prokaryotic operons. Nature (London) **298**:760–762.

12a. **Hogg, R., et al.** Personal communication.

13. **Horazdovsky, B. F., and R. W. Hogg.** 1987. High-affinity L-arabinose transport operon. Gene product and mRNAs. J. Mol. Biol. **197**:27–35.
14. **Horazdovsky, B. F., and R. W. Hogg.** 1989. Genetic reconstitution of the high-affinity L-arabinose transport system. J. Bacteriol. **171**:3053–3059.
15. **Huo, L., K. Martin, and R. Schleif.** 1988. Alternative DNA loops regulate the arabinose operon in *Escherichia coli*. Proc. Natl. Acad. Sci. USA **85**:5444–5448.
16. **Kohara, Y., K. Akiyama, and K. Isono.** 1987. The physical map of the whole *E. coli* chromosome: application of a new strategy for rapid analysis and sorting of a large genomic library. Cell **50**:495–508.
17. **Kolodrubetz, D., and R. Schleif.** 1981. L-Arabinose transport systems in *Escherichia coli* K-12. J. Bacteriol. **148**:472–479.
18. **Kosiba, B. E., and R. Schleif.** 1982. Arabinose inducible promoter from *Escherichia coli*, its cloning from chromosomal DNA, identification as the *araFG* promoter and sequence. J. Mol. Biol. **156**:53–66.
19. **Kreuzer, P., D. Gartner, R. Allmansburger, and W. Hillen.** 1989. Identification and sequence analysis of the *Bacillus subtillus* W23 *xylR* gene and *xyl* operator. J. Bacteriol. **171**:3840–3845.
20. **Kyte, J., and R. F. Doolittle.** 1982. A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. **157**:105–132.
21. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
22. **Marck, C.** 1988. 'DNA Stryder': a 'C' program for the fast analysis of DNA and protein sequences on the Apple Macintosh family of computers. Nucleic Acids Res. **16**:1829–1836.
23. **Messing, J.** 1979. A multipurpose cloning system based on single-stranded DNA bacteriophage M13. Recomb. DNA Tech. Bull. **2**:43.
24. **Naom, I. S., S. J. Morton, D. R. F. Leach, and R. G. Lloyd.** 1989. Molecular organization of *sbcC*, a gene that affects genetic recombination and the viability of DNA palindromes in *Escherichia coli* K-12. Nucleic Acids Res. **17**:8033–8044.
25. **Newbury, S. F., N. H. Smith, E. C. Robinson, I. D. Hiles, and C. F. Higgens.** 1987. Stabilization of transcriptionally active mRNA by prokaryotic REP sequences. Cell **48**:297–310.
26. **Niki, H., A. Jaffe, R. Imamura, T. Ogura, and S. Hiraga.** 1991. The new gene *mukB* codes for a 177 kd protein with coiled-coil domains involved in chromosome partioning of *E. coli*. EMBO J. **10**:183–193.
27. **Novotny, C., and E. Englesberg.** 1966. The L-arabinose permease system in *Escherichia coli* B/r. Biochim. Biophys. Acta **117**:217–230.
28. **Pearson, W. R., and P. J. Lipman.** 1982. Improved tools for biological sequence comparison. Proc. Natl. Acad. Sci. USA **85**:2444–2448.
29. **Sandhu, G., J. W. Precup, and B. C. Kline.** 1989. The polymerase chain reaction colony miniprep. Biotechnology **7**:689–690.
30. **Sanger, F., S. Miklen, and A. R. Coulson.** 1977. DNA sequencing using chain termination inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
31. **Schleif, R.** 1969. L-arabinose binding protein and arabinose permeation in *Escherichia coli*. J. Mol. Biol. **46**:185–196.
32. **Schleif, R.** 1987. The L-arabinose operon, p. 1473–1481. *In* F. C. Neidhardt, J. L. Ingraham, K. B. Low, B. Magasanik, M. Schaecter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella typhimurium*: cellular and molecular biology, vol. 2. American Society for Microbiology, Washington, D.C.
33. **Schleif, R., and P. Wensink.** 1981. Practical methods in molecular biology. Springer-Verlag, New York.
34. **Scripture, J. B., C. Voelker, S. Miller, R. T. O'Donnell, L. Polgar, J. Rade, B. F. Horazdovsky, and R. W. Hogg.** 1987. High affinity L-arabinose transport operon: nucleotide sequence and analysis of gene products. J. Mol. Biol. **197**:37–46.
35. **Smith, C., J. Econome, A. Schuff, S. Kico, and C. Cantor.** 1987. A physical map of the *Escherichia coli* K12 genome. Science **236**:1448–1453.
36. **Stern, M. J., G. F.-L. Ames, N. H. Smith, E. C. Robinson, and C. F. Higgens.** 1984. Repetitive extragenic palindromic sequences: a major component of the bacterial genome. Cell **37**:1015–1026.
37. **Tabor, S., and C. C. Richardson.** 1987. DNA sequence analysis with a modified bacteriophage T7 DNA polymerase. Proc. Natl. Acad. Sci. USA **84**:4767–4771.
38. **Walker, J. E., M. Saraste, M. J. Runswick, and N. J. Gay.** 1982. Distantly related sequences in the alpha and beta subunits of ATP synthase, myosin, kinases, and other ATP-requiring enzymes and a common nucleotide binding fold. EMBO J. **1**:945–951.