JOURNAL OF BACTERIOLOGY, Mar. 1997, p. 1598–1608
0021-9193/97/$04.00+0

Vol. 179, No. 5

Negative Regulation of L-Arabinose Metabolism in *Bacillus subtilis*: Characterization of the *araR* (*araC*) Gene

ISABEL SÁ-NOGUEIRA* AND LUÍS JAIME MOTA

Instituto de Tecnologia Química e Biológica, Universidade Nova de Lisboa, 2780 Oeiras, Portugal

Received 2 October 1996/Accepted 20 December 1996

The *Bacillus subtilis araC* locus, mapped at about 294° on the genetic map, was defined by mutations conferring an Ara⁻ phenotype to strains bearing the metabolic *araA*, *araB*, and *araD* wild-type alleles (located at about 256° on the genetic map) and by mutants showing constitutive expression of the three genes. In previous work, it has been postulated that the gene in which these mutations lie exerts its effect on the *ara* metabolic operon in *trans*, and this locus was named *araC* by analogy to the *Escherichia coli* regulatory gene. Here, we report the cloning and sequencing of the *araC* locus. This region comprises two open reading frames with divergently arranged promoters, the regulatory gene, *araC*, encoding a 41-kDa polypeptide, and a partially cloned gene, termed *araE*, which most probably codes for a permease involved in the transport of L-arabinose. The DNA sequence of *araC* revealed that its putative product is very similar to a number of bacterial negative regulators (the GalR-LacI family). However, a helix-turn-helix motif was identified in the N-terminal region by its identity to the consensus signature sequence of another group of repressors, the GntR family. The lack of similarity between the predicted primary structure of the product encoded by the *B. subtilis* regulatory gene and the AraC regulator from *E. coli* and the apparently different modes of action of these two proteins lead us to propose a new name, *araR*, for this gene. The *araR* gene is monocistronic, and the promoter region contains −10 and −35 regions (as determined by primer extension analysis) similar to those recognized by RNA polymerase containing the major vegetative cell sigma factor σ^A. An insertion-deletion mutation in the *araR* gene leads to constitutive expression of the L-arabinose metabolic operon. We demonstrate that the *araR* gene codes for a negative regulator of the *ara* operon and that the expression of *araR* is repressed by its own product.

Bacillus subtilis is able to grow on L-arabinose as the sole carbon and energy source. The pathway of L-arabinose utilization in *B. subtilis* was described by Lepesant and Dedonder (30) and is identical to the one found in *Escherichia coli* (13). After entering the cell, L-arabinose is sequentially converted to L-ribulose, L-ribulose-5-phosphate, and D-xylulose-5-phosphate by the action of L-arabinose isomerase, L-ribulokinase, and L-ribulose-5-phosphate 4-epimerase, respectively. D-Xylulose-5-phosphate is further catabolized through the pentose phosphate pathway. The synthesis of these enzymes was shown to be inducible by L-arabinose, and the isomerase activity was shown to be subjected to catabolite repression by glucose and glycerol (30).

The three metabolic genes, *araA*, *araB*, and *araD*, coding for L-arabinose isomerase, L-ribulokinase, and L-ribulose-5-phosphate 4-epimerase, respectively, have been cloned, and complementation experiments have shown that the products are functionally homologous to their *E. coli* counterparts (45). These genes are adjacent, with the order A-B-D (45), unlike the B-A-D order found in the *E. coli* operon (15), and constitute the first three open reading frames (ORFs) of a nine-cistron transcriptional unit with a total length of 11 kb (46). This operon, called *ara*, is located at about 256° on the *B. subtilis* genetic map and comprises six other genes, named *araL*, *araM*, *araN*, *araP*, *araQ*, and *abfA*. Analysis of the sequence of the *ara* operon showed that the putative products of *araN*, *araP*, and *araQ* are homologous to bacterial components of binding-protein-dependent transport systems and that the *abfA* gene most probably codes for an α-L-arabinofuranosidase. The function of *araL* and *araM* is unknown, but an in vitro-constructed insertion-deletion mutation in the region downstream from *araD* demonstrated that the *araL*, *araM*, *araN*, *araP*, *araQ*, and *abfA* genes are not essential for L-arabinose utilization (46). Expression of the *araABDLMNPQ-abfA* operon is directed by a strong σ^A-like promoter identified within a 150-bp DNA fragment upstream from the translation start site of the *araA* gene, and studies with strains bearing transcriptional fusions of the operon to the *E. coli lacZ* gene revealed that expression from the *araABDLMNPQ-abfA* promoter is induced by L-arabinose and repressed by glucose (46).

Three additional classes of mutations affecting L-arabinose utilization were identified: (i) mutations conferring an Ara⁻ phenotype to strains bearing the *araA*, *araB*, and *araD* wild-type alleles (39, 40), (ii) mutations leading to constitutive expression of the three genes (47), and (iii) mutations conferring a conditional Ara⁻ phenotype which was dependent on the concentration of sugar (39, 40). These mutations were mapped at about 294° on the *B. subtilis* genetic map and define another *ara* locus. It has been postulated that the gene in which these mutations lie exerts its effect on the *araABDLMNPQ-abfA* operon in *trans*; this locus was named *araC* by analogy to the *E. coli* regulatory gene.

In *E. coli*, arabinose utilization requires the expression of the metabolic operon, *araBAD*, and expression of either the low-affinity transport gene, *araE*, or the high-affinity transport operon, *araFGH*. The AraC protein represses transcription from its own promoter in the absence of arabinose, via a DNA-looping mechanism in the divergently transcribed *araC-araBAD* promoter region, and in the presence of arabinose activates transcription from three other promoters, p_{ABD}, p_E, p_{FGH} (reference (48) and references therein).

* Corresponding author. Mailing address: Instituto de Tecnologia Química e Biológica, Rua da Quinta Grande 6, Apartado 127, 2780 Oeiras Codex, Portugal. Phone: 351-1-4426171. Fax: 351-1-4428766. E-mail: sanoguei@itqb.unl.pt.

TABLE 1. *B. subtilis* strains used

Strain	Genotype	Phenotype	Source or reference
168T$^+$	Prototroph	Ara$^+$	F. E. Young
168T$^+$ derivatives			
IGCg704	*araC4*(*araR4*)*metB10 lys3*	Ara$^-$	40
IQB210	*araR*::pLM1(*ara′R-cat*)	Ara$^+$ Cm^r	pLM1→168T^{+a}
IQB211	*araR′*::pSN2(*araR′-lacZ cat*)	Ara(Con) LacZ$^+$ Cm^r	pSN2→168T$^+$
IQB212	*araR′*::pSN3(*araR′-cat lacZ*)	Ara(Con) LacZ$^-$ Cm^r	pSN3→168T$^+$
IQB213	*amyE′*::(*araA′-′lacZ cat*)	Ara$^+$ LacZ$^+$ Cm^r Amy$^-$	pLM12^b→168T$^+$
IQB214	Δ*araR*::*km amyE′*::(*araA′-′lacZ cat*)	Ara(Con) LacZ$^+$ Cm^r Km^r Amy$^-$	pLM8^b→IQB213
IQB215	Δ*araR*::*km*	Ara(Con) Km^r	pLM8^b→168T$^+$
IQB216	*amyE′*::(*′lacZ cat*)	Ara$^+$ LacZ$^-$ Cm^r Amy$^-$	pAC5^b→168T$^+$
IQB217	Δ*araR*::*km amyE′*::(*′lacZ cat*)	Ara(Con) LacZ$^-$ Cm^r Km^r Amy$^-$	pAC5^b→IQB215
IQB218	*araR*::pLM15(*araR′-lacZ cat*)	Ara$^+$ LacZ$^+$ Cm^r	pLM15→168T$^+$

[a] The arrows indicate transformation and point from donor DNA to recipient strain.
[b] Transformation was carried out with linearized plasmid DNA.

In this communication, we report the cloning and nucleotide sequence of the *B. subtilis araC* locus. This region comprises two ORFs with divergently arranged promoters, the regulatory gene, *araC*, and a partially cloned gene, named *araE*, which most probably codes for a permease similar in function to the *E. coli* AraE transport protein. The lack of similarity between the predicted primary structure of the product encoded by the *B. subtilis* regulatory gene and the regulator AraC from *E. coli* and apparently different modes of action of these two proteins lead us to propose a new name, *araR*, for this gene involved in the regulation of the L-arabinose metabolism in *B. subtilis*. We demonstrate that the *araR* gene codes for a negative regulator of the *araABDLMNPQ-abfA* operon and that the expression from the *araR* promoter is autoregulated in a manner that is independent of the presence of L-arabinose.

MATERIALS AND METHODS

Bacterial strains and growth conditions. The *B. subtilis* strains used in this study are listed in Table 1. *E. coli* DH5α (GIBCO BRL, Life Technologies European Division) was used as a host for all plasmids, and *E. coli* DH5α F′ (GIBCO BRL) was used for the propagation and amplification of recombinant M13 bacteriophages. *E. coli* strains were grown on LB (Luria-Bertani medium [35]). Ampicillin (75 μg ml^{-1}), chloramphenicol (30 μg ml^{-1}), spectinomycin (50 μg ml^{-1}), X-Gal (5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside; 40 μg ml^{-1}), and IPTG (isopropyl-β-D-thiogalactopyranoside; 1 mM) were added as appropriate. *B. subtilis* strains were grown on LB (35), SP (31), or C minimal medium (38). Chloramphenicol (5 μg ml^{-1}), erythromycin (1 μg ml^{-1}), kanamycin (15 μg ml^{-1}), and spectinomycin (50 μg ml^{-1}) were added as appropriate. Solid medium was made with LB, SP, or C medium containing 1.5% (wt/vol) Bacto Agar (Difco). To test for growth of *B. subtilis* integrant strains on L-arabinose as the sole carbon source, strains were plated on minimal C medium containing L-arabinose (0.1% [wt/vol]). The Ara(Con) phenotype was determined on LB plates by a filter paper assay as described by Englesberg et al. (14). The *amyE* phenotype was tested by plating on tryptose blood agar base medium (Difco) containing 1% (wt/vol) potato starch; after overnight incubation, plates were flooded with a solution of 0.5% (wt/vol) I_2–5.0% (wt/vol) KI for detection of starch hydrolysis. For the β-galactosidase assays and RNA preparation, the *B. subtilis* strains were grown in liquid C medium supplemented with casein hydrolysate (1% [wt/vol]) and L-arabinose and glucose were added to the cultures when necessary at a final concentration of 0.4% (wt/vol).

Library of *B. subtilis* chromosomal DNA fragments and cloning of the *araR* region. A random library of *B. subtilis* 168 chromosomal DNA fragments, a gift from M. Débarbouillé, obtained by partial digestion with *Sau*3AI was made in the shuttle vector pHT315 Amp^r Em^r (4) digested with *Bam*HI. The total gene bank represents about 3,000 to 5,000 clones, and the bacterial host used was *E. coli* TG1 [K-12 F′ *traD36 proA*$^+$ *B*$^+$ *lacI*^{q} *lacZ M15* Δ(*lac-pro*) *supE thi hsdD*]. Pooled DNA from the original clones of this library was used to transform cells of *B. subtilis* IGCg704 *araC4* (*araR4*) (Table 1) with selection for Ara$^+$ Em^r (erythromycin resistance) on minimal C medium plates containing L-arabinose (0.1% [wt/vol]) and erythromycin (1 μg ml^{-1}). Plasmid DNA isolated from the Ara$^+$ Em^r clones were used to transform *E. coli* with selection for ampicillin resistance (Amp^r). After amplification in *E. coli*, plasmids extracted from the Amp^r clones were used to transform competent cells of *B. subtilis* IGCg704 *araC4* (*araR4*) with selection for Ara$^+$ on minimal C medium plates containing L-arabinose (0.1% [wt/vol]) as the sole carbon source.

DNA manipulations and sequencing. DNA manipulations were carried out as described by Sambrook et al. (43). Enzymes were used according to manufacturers' instructions. DNA sequencing was performed by the method of Sanger et al. (44) with a Sequenase kit (T7 DNA polymerase; United States Biochemical Corporation). Sequencing templates were prepared by a combination of subcloning appropriate fragments from pSN1 and pLM2 into the polycloning site of M13mp19 or M13mp18 (60) and sequential deletion of the recombinant M13 derivatives, by the method of Dale et al. (10), using a Cyclone Biosystem kit (International Biotechnologies, Inc.). The DNA sequence was determined on both strands and across all restriction sites used for subcloning. The primer 5′-CCTCTTCGCTATTACGCC-3′, complementary to the coding sequences of the *lacZ* gene, was used to sequence the transcriptional and translational *lacZ* fusions.

Plasmid constructions. Plasmid pLM1 was constructed by subcloning a 966-bp *Hin*dIII-*Hpa*I DNA fragment (positions 1116 [*Sau*3A1] to 2046 [Fig. 1]) from pSN1 between the *Hin*dIII and *Eco*RV sites of the integrational vector pJH101 (17). Plasmid pLM2 was obtained by cutting chromosomal DNA from *B. subtilis* IQB210 (Table 1) with *Hin*dIII, followed by circularization of this DNA at low concentration. This ligation mixture was transformed into *E. coli*, and Amp^r transformants were selected as described above. To construct plasmid pLM3, we digested pLM2 with *Xmn*I and cloned a purified fragment of 1.4 kb (positions 615 to 2046 [Fig. 1]) at the unique *Sma*I site of the shuttle vector pMK3-1 (53). Plasmid pLM5 was constructed by subcloning a 541-bp *Sac*I DNA fragment (positions 2167 to 2689 [Fig. 1]) from pSN1 into the *Sac*I sites of pAH250 (24). Plasmid pLM6 was obtained by insertion of a 1.5-kb *Xba*I DNA fragment from pAH248 [a pGem7-zf(+) derivative that contains a kanamycin resistance [Km^r] gene (*km*) cloned between its *Xho*I and *Eco*RI sites (25)] that retains the *km* gene into pLM5 partially digested with *Xba*I. Plasmid pLM7 was constructed by subcloning the 378-bp *Xmn*I-*Hpa*II (filled-in) DNA fragment (positions 615 to 993 [Fig. 1]) from pSN1 at the unique *Sma*I site of pBluescript SK(+) (Stratagene, La Jolla, Calif.). Plasmid pLM8 was obtained by subcloning a 395-bp *Eco*RI-*Bam*HI DNA fragment (positions 615 to 993 [Fig. 1]) from pLM7 between the *Eco*RI and *Bam*HI sites of pLM6. To construct pLM12, which carries an *araA-lacZ* in-frame fusion, a 249-bp *Hin*dIII-*Eco*RI (filled-in) DNA fragment from pSNL9 (46), containing the start sites for the *araABDLMNPQ-abfA* operon transcription and translation and the first eight amino acid residues of *araA*, was subcloned between the *Sma*I and *Bam*HI (filled in with dGTP and dATP followed by treatment with S1 nuclease) sites of pAC5 (32). pAC5 contains homology to the *B. subtilis amyE* locus and is derived from ptrpBG1 (50). Plasmid pLM12 carries a translational fusion between codon 8 of *araA* and codon 7 of *E. coli lacZ*. This fusion was confirmed by DNA sequencing. Plasmids pSN2 and pSN3 were obtained by subcloning a 469-bp *Dde*I DNA fragment (positions 1254 to 1723 [Fig. 1]) from pSN1 at the unique *Sma*I restriction site the integrational vector pJM783 (41), in both orientations. pSN2 contains the *B. subtilis spoVG* ribosome binding site and the *E. coli lacZ* gene in the same orientation as the *araR* region sequences, which was confirmed by DNA sequencing, and pSN3 in the opposite orientation. Plasmid pLM15 was constructed by subcloning a 395-bp *Eco*RI-*Bam*HI DNA fragment (positions 615 to 993 [Fig. 1]) from pLM7 between the *Eco*RI and *Bam*HI sites of the integrational vector pJM783 (41). This plasmid contains the *B. subtilis spoVG* ribosome binding site and the *E. coli lacZ* gene in the same orientation as the *araR* region sequences, which was confirmed by DNA sequencing.

Bacterial transformation. *B. subtilis* DNA transformations were performed by the method of Anagnostopoulos and Spizizen (3). *E. coli* transformations were carried out according to standard methods (43).

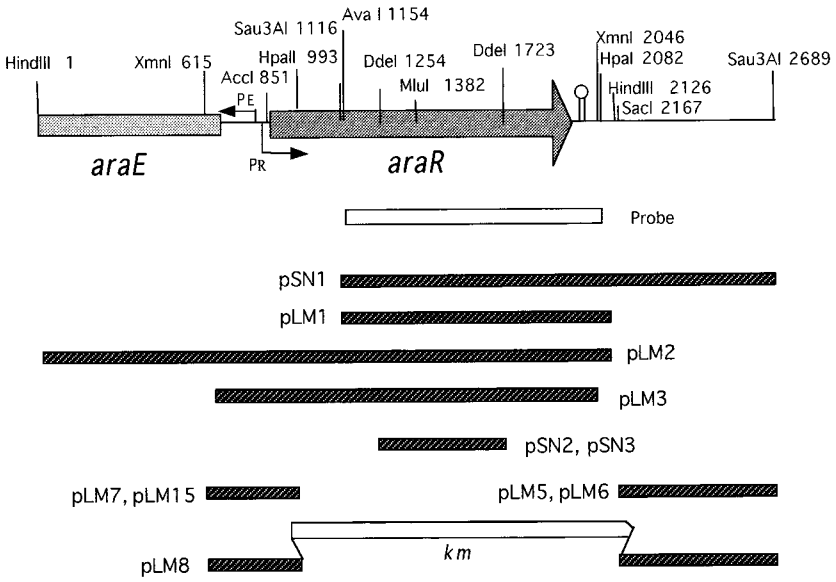

FIG. 1. Physical and genetic map of the *araR* region of the chromosome. The locations of the ORFs (*araR* and *araE*), predicted from the analysis of the nucleotide sequence, are indicated by a rectangle (*araE*) and by an arrow (*araR*) pointing in the direction of transcription. The *araR* promoter (P_R), defined by primer extension, and the region of dyad symmetry that could represent the transcription terminator of the *araR* transcriptional unit are indicated. The putative promoter (P_E) of the incomplete *araE* gene is shown, and the black arrow indicates the direction of transcription. The positions for restriction sites are from the nucleotide sequence. The open box below the physical map denotes the fragment used as probe for Northern analysis of the *araR* transcript, and the striped boxes represent the extension of the inserts in the indicated plasmids. Plasmids pLM1, pSN2, pSN3, and pLM15 were integrated into the host chromosome by means of a single-crossover (Campbell-type) recombinational event that occurred in the region of homology. The site of an insertion-deletion mutation resulting from replacement of wild-type sequences by a double-crossover event, with an in vitro-engineered fragment of the *araR* region, present in plasmid pLM8, is also shown.

Construction of strains bearing an *araA′-′lacZ* translational fusion. To create a deletion in the *araR* region, plasmid pLM8 (Km^r Sp^r [spectinomycin resistant] [described above]) was linearized with *Sca*I and used to transform the wild-type strain $168T^+$ to Km^r. The resulting strain IQB215 (Table 1) was Sp^s, which indicated that the Km^r phenotype was the result of a double-crossover event that occurred on both sides of the Km^r cassette inserted between the *araR* 5′-end region and the chromosomal sequences downstream from the *araR* gene (Fig. 1). Plasmid pLM12, carrying an *araA-lacZ* in-frame fusion was linearized and used to transform the *B. subtilis* wild-type strain $168T^+$ to chloramphenicol resistance (Cm^r). The fusion integrated into the chromosome of strain $168T^+$, via double recombination with the *amyE* gene back and front sequences, resulting in an Amy^- Cm^r phenotype in the transformant strain named IQB213 (Table 1). To delete the *araR* gene on the chromosome of IQB213, we used the procedure described above to construct strain IQB215. Competent cells of IQB213 were the recipient of linearized pLM8, and the resulting strain, IQB214 (Table 1), displayed an Sp^s Km^r phenotype, which indicated that the Km^r phenotype was the result of a double-crossover event that occurred on both sides of the Km^r cassette as described above.

Construction of strains bearing an *araR-lacZ* transcriptional fusion. The integrational plasmids pLM15, pSN2, and pSN3 (described above) were used separately to transform *B. subtilis* $168T^+$ to Cm^r and integrated, as single copy, into the chromosome by a Campbell-type recombinational event (confirmed by Southern hybridization [data not shown]). The integration of pSN2 (in which the *lacZ* gene is in the same orientation as the *araR* sequences) and pSN3 (*lacZ* and *araR* in opposite orientations) at the *araR* locus disrupted the *araR* transcriptional unit (Fig. 1) in the resulting strains IQB211 and IQB212 (Table 1), respectively. The integration of pLM15 (in which *lacZ* gene is in the same orientation as the *araR* sequences) at the *araR* locus was not disruptive (Fig. 1) in the resulting strain IQB218 (Table 1).

β-Galactosidase assays. Strains of *B. subtilis* harboring transcriptional *lacZ* fusions were grown on 30 ml of C medium supplemented with casein hydrolysate (1% [wt/vol]). During early logarithmic phase (optical density at 600 nm [OD_{600}] of 0.11 to 0.15), 15 ml of the culture was transferred to a different flask and L-arabinose at a final concentration of 0.4% (wt/vol) was added. Exponential growth of the two cultures was monitored by measuring absorbance until the cultures reached an OD_{600} of 0.7 to 0.8, which corresponds to growth for at least 2.5 generations in the presence of the inducer. At 30-min intervals, 100-μl aliquots of cell culture were collected, harvested, and stored at −70°C overnight. The cells were suspended in 1 ml of Z buffer (35), and 2 drops of chloroform and 1 drop of 0.1% sodium dodecyl sulfate were added and mixed vigorously for 10 s on a tabletop vortex apparatus. β-Galactosidase activity was determined as described by Miller (35), using the substrate *o*-nitrophenyl-β-D-galactopyranoside (ONPG).

RNA preparation, Northern blot, and primer extension analysis. *B. subtilis* $168T^+$ cells were grown in C medium supplemented with 1% (wt/vol) casein hydrolysate in the presence and in the absence of L-arabinose at a final concentration of 0.4% (wt/vol). Cells were harvested during late logarithmic phase of growth (OD_{600} of approximately 0.9), and RNA was prepared essentially as described by Igo and Losick (27). For Northern blot analysis, 5 to 20 μg of total RNA was run in a 1.2% (wt/vol) agarose formaldehyde denaturing gel and transferred to positively charged nylon membranes (Hybond-N^+; Amersham International, Amersham, United Kingdom) according to standard methods (43). A size determination was done by using an RNA ladder (9.5 to 0.24 kb; GIBCO BRL). The DNA probe was labeled by using the Amersham Multiprime random oligonucleotide DNA labeling system and [α-^{32}P]dATP (6,000 Ci/mmol). Primer extension analysis was performed essentially as described by Sambrook et al. (43). Primer A (5′-ACCATACGGCCGACTGCGGC-3′), complementary to nucleotides (nt) 884 to 1003 (Fig. 2), was end labeled with [γ-^{32}P]ATP (3,000 Ci/mmol) by using T4 polynucleotide kinase. A total of 10 ng of primer was used in the labeling reaction and mixed with 25 μg of RNA, denatured by heating to 85°C for 10 min, and annealed by incubation at 42°C overnight. The oligonucleotide primer was extended by using 20 U of avian myeloblastosis virus reverse transcriptase (Amersham) for 2 h at 37°C as described by Sambrook et al. (43). Analysis of the extended products was carried out on 6% polyacrylamide-urea gels.

Computer analysis. Amino acid sequences were deduced from the nucleotide sequence by using DNASIS version 2.0 (Hitachi Software Engineering Co., Ltd.). The GenBank and EMBL databases were accessed by using the Genetics Computer Group (GCG; Madison, Wis.) package of sequence analysis software.

Nucleotide sequence accession number. The DNA sequence reported here has been submitted to the GenBank database and assigned accession number X98354.

RESULTS

Cloning of the *araR* region of the chromosome. Pooled DNA from the original clones of a random library of *B. subtilis* chromosomal DNA fragments made in the shuttle vector pHT315 (see Materials and Methods) was used to transform cells of *B. subtilis* IGCg704 *araC4* (*araR4*) to Ara^+ Em^r. Several Ara^+ Em^r transformants were obtained, but after amplification of plasmid DNA extracted from such clones in *E. coli* DH5α, only one plasmid, named pSN1, retained the ability to

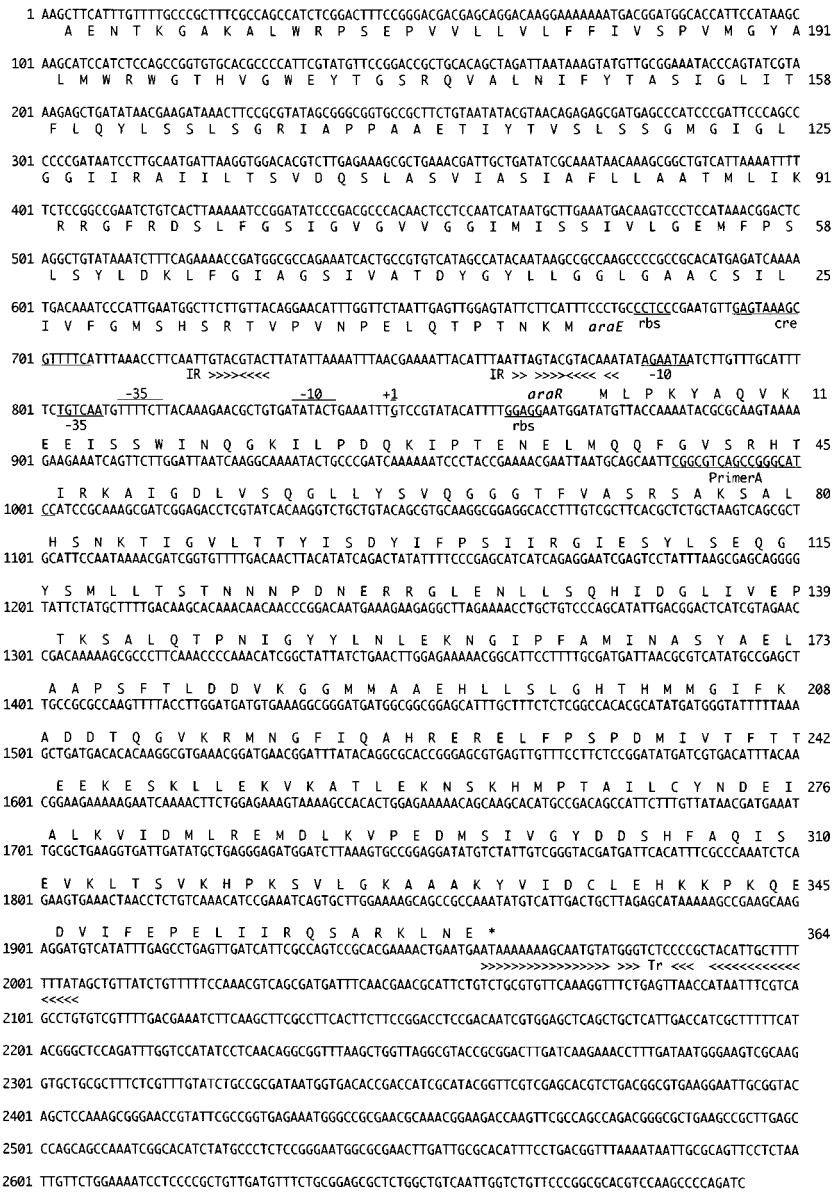

FIG. 2. Nucleotide sequence and translation of the *B. subtilis araR* region. The nucleotide sequence of the *araR* nontranscribed strand and *araE* transcribed strand is shown in the 5′-to-3′ direction. The predicted primary structure of the polypeptides encoded by *araR* (above the nucleotide sequence) and *araE* (below the nucleotide sequence) are given in single-letter code. The *araR* transcription start site (+1) defined by primer extension analysis, the −35 and −10 regions of the promoter, and the putative ribosome binding sites (rbs) are indicated. Stop codons are represented by asterisks. Convergent arrows represent different regions of dyad symmetry (inverted repeats [IR]), and the inverted repeat thought to act as transcriptional terminator for the *araR* gene is indicated by Tr. The complementary sequence of primer A, used in primer extension analysis, is represented below the sequence. The putative −35 and −10 regions of the *araE* promoter and the putative ribosome binding sites are shown below the nucleotide sequence. The putative catabolic repression-associated sequence (cre) is underlined.

generate Ara^+ transformants in recipient strain IGCg704 *araC4* (*araR4*). Plasmid pSN1, which contained a 1.5-kb chromosomal insert, was used to ascertain the location of the *araR* locus in the cloned DNA segment. The procedure involved transformation of strain IGCg704 *araC4* (*araR4*) to Ara^+, with defined restriction DNA fragments from the pSN1 insert; on the basis of the results, we located the site of the *araR4* mutation at one end of the cloned fragment (between the *Sau*3AI and *Hin*dIII sites, positions 1116 to 2126 [Fig. 1]). This result, together with the DNA sequencing analysis (see below), suggested that only part of the *araR* gene was present in this plasmid. To clone the entire *araR* gene, integrational plasmid pLM1, carrying sequences of *araR* (Fig. 1), was transformed into the *B. subtilis* wild-type strain $168T^+$ with selection for Cm^r. The plasmid integrated into the *B. subtilis* chromosome at the *araR* region of homology by a Campbell-type recombination mechanism, and Southern hybridization (data not shown) showed that the resulting strain (IQB210) contained more than one copy of pLM1 integrated into the *araR* region of the chromosome. Chromosomal DNA from IQB210 digested with *Hin*dIII was used to rescue the entire *araR* gene and its upstream region (see Materials and Methods). This procedure created plasmid pLM2, which carried an additional 1.1 kb of DNA adjacent to the previously cloned fragment in plasmid pLM1 (Fig. 1). The structure of the inserts in pSN1 and pLM2 was compared to that of the corresponding areas of chromo-

TABLE 2. Percentages of amino acid identity between the predicted sequences of the Ara proteins and similar proteins

B. subtilis Ara protein	Homolog (species; reference)[a]	Function	% Identity	Amino acid overlap (no.)
AraR	DegA (*Bacillus subtilis*; 9)	Sporulation-specific enzyme, degradation protein	27.9	315
	CytR (*Escherichia coli*; 55)	Repressor protein	26.8	287
	PurR (*E. coli*; 34)	Repressor protein	25.8	283
	RbsR (*E. coli*; 33)	Repressor protein	26.3	285
	CcpA (*B. megaterium*; 26)	Repressor protein	25.4	290
	CcpA (*B. subtilis*; 23)	Repressor protein	23.8	281
	GalR (*E. coli*; 58)	Repressor protein	22.2	320
	LacI (*E. coli*; 16)	Repressor protein	20.6	290
AraE	AraE (*E. coli*; 51)	Arabinose transport protein	36.5	200
	AraE (*Klebsiella oxytoca*; 49a)	Arabinose transport protein	35.4	212
	XylE (*E. coli*; 11)	Xylose transport protein	42.1	121
	GlcP (*Synechocystis* sp.; 61)	Glucose transport protein	35.2	219
	Glf (*Zymomonas mobilis*; 5)	Glucose-facilitated diffusion protein	32.3	226

somal DNA by Southern blot analysis (data not shown), and the results revealed that no detectable rearrangement occurred during the cloning process.

Nucleotide sequence and organization of the *araR* region. DNA sequence analysis of the *araR* region revealed the presence of two ORFs; the first, by its position in the sequenced fragments of pSN1 and pLM2 and according to our previous results (see above), was identified as the *araR* gene (Fig. 2). The *araR* coding region is preceded by a putative ribosome binding site and could encode a 362-amino-acid product of 40.4 kDa (Fig. 2). One potential hairpin-loop structure, situated next to the UAA stop codon of the *araR* gene (Tr [Fig. 2], with a ΔG value of -27.4 kcal mol^{-1} [54]), probably corresponds to a transcription terminator. Analysis of molecular size of AraR by sodium dodecyl sulfate-polyacrylamide gel electrophoresis after overexpression of *araR* by induction of the T7 promoter in *E. coli* BL21(DE3)(LysS) (52) gave a value of 42.7 kDa, which matches the molecular mass deduced from the *araR* gene (data not shown).

Another ORF was identified beginning 204 nt upstream from the *araR* start codon but transcribed in the opposite direction. This ORF, named *araE*, is preceded by a putative strong ribosome binding site and could encode a polypeptide of at least 223 residues, of which only the amino-terminal part was present in the cloned region (Fig. 1 and 2). Upstream from the presumptive ribosome binding site of *araE* were two sequences, TATTCT and TTGACA, shown in their antisense orientation (Fig. 2), that are similar to the -10 and -35 consensus sequences TATAAT and TTGACA, respectively, recognized by *B. subtilis* RNA polymerase containing the major vegetative cell sigma factor σ^A (36). Preliminary studies with strains bearing *araE-lacZ* transcriptional fusions suggested that the expression of *araE* is subjected to catabolite repression by glucose (37). Between the putative promoter region and the *araE* ribosome binding site is the sequence TGAAAACGCT TTAC (shown in its antisense orientation between positions 694 and 707 [Fig. 2]), which is similar to the consensus sequence (TGWNANCGNTNWCA, W = A or T [57]) for the *cis*-acting sites (CRE) involved in *B. subtilis* catabolite repression.

Comparison of the primary structures of the products predicted to be encoded by the *araR* and *araE* genes with sequences in the GenBank revealed significant similarities with other bacterial proteins of known function (Table 2). The amino acid sequence encoded by *araR* revealed a significant similarity to a group of *E. coli* regulatory proteins (PurR, RbsR, CytR, GalR, and LacI) and to the *B. subtilis* CcpA protein, involved in catabolite repression (23), and DegA, a sporulation-specific protein involved in enzyme degradation (9). These proteins belong to a large family of bacterial regulators homologous to the *E. coli* Gal and Lac repressors (56). The primary structure of the *araR* predicted product could be aligned with the proteins of this family throughout their lengths, with minor gaps (Fig. 3), and the highest degree of identity observed was in some of the regions implicated in effector binding and oligomerization of the *E. coli* proteins belonging to this group (56). The N-terminal region of AraR contains a helix-turn-helix (HTH) motif characteristic of nucleic acid binding proteins (residues 30 to 51 [Fig. 3]). Interestingly, this domain does not display significant similarity to the HTH motif of the regulators that belong to this group (Fig. 3) but is identical to the consensus signature sequence of the members of the *B. subtilis* GntR family of regulatory proteins (21; Fig. 4).

The incomplete predicted product of *araE*, a hydrophobic protein, exhibited a hydropathy profile (29) characteristic of integral membrane proteins: six major regions of high hydrophobicity (hydrophatic index, >1.0), each composed of at least 20 amino acids which could be capable of spanning the membrane (data not shown). The amino-terminal part of AraE showed a high level of identity to the N termini of several integral membrane proteins involved in prokaryotic sugar transport (Table 2), in particular the L-arabinose proton symporter encoded by the *E. coli araE* gene. Interestingly, mutations conferring a conditional Ara$^-$ phenotype dependent on the concentration of L-arabinose in the medium were mapped in this region of the chromosome, and it has been postulated that they could lie in a gene involved in L-arabinose transport (39, 40). These observations suggest that the *B. subtilis araE* gene encodes for a permease similar in function to the *E. coli* AraE protein.

Effect of *araR* deficiency on expression of the L-arabinose (*araABDLMNPQ-abfA*) operon. To characterize the function of the *araR* gene in the catabolism of L-arabinose, we constructed a deletion in the *araR* region. The resulting strain, IQB215 (Table 1), was unable to grow on minimal medium plates with L-arabinose as the sole carbon source and displayed constitutive expression of the L-arabinose metabolic genes, *araA*, *araB*, and *araD* [Ara(Con) phenotype], on rich medium, as determined by the filter paper assay (14). To study the effect of an *araR* null mutation in the expression of the *araABDLMNPQ-abfA* operon, we examined the expression of an *araA-lacZ* in-frame translational fusion in both wild-type and defective *araR* backgrounds during growth in the absence or presence of

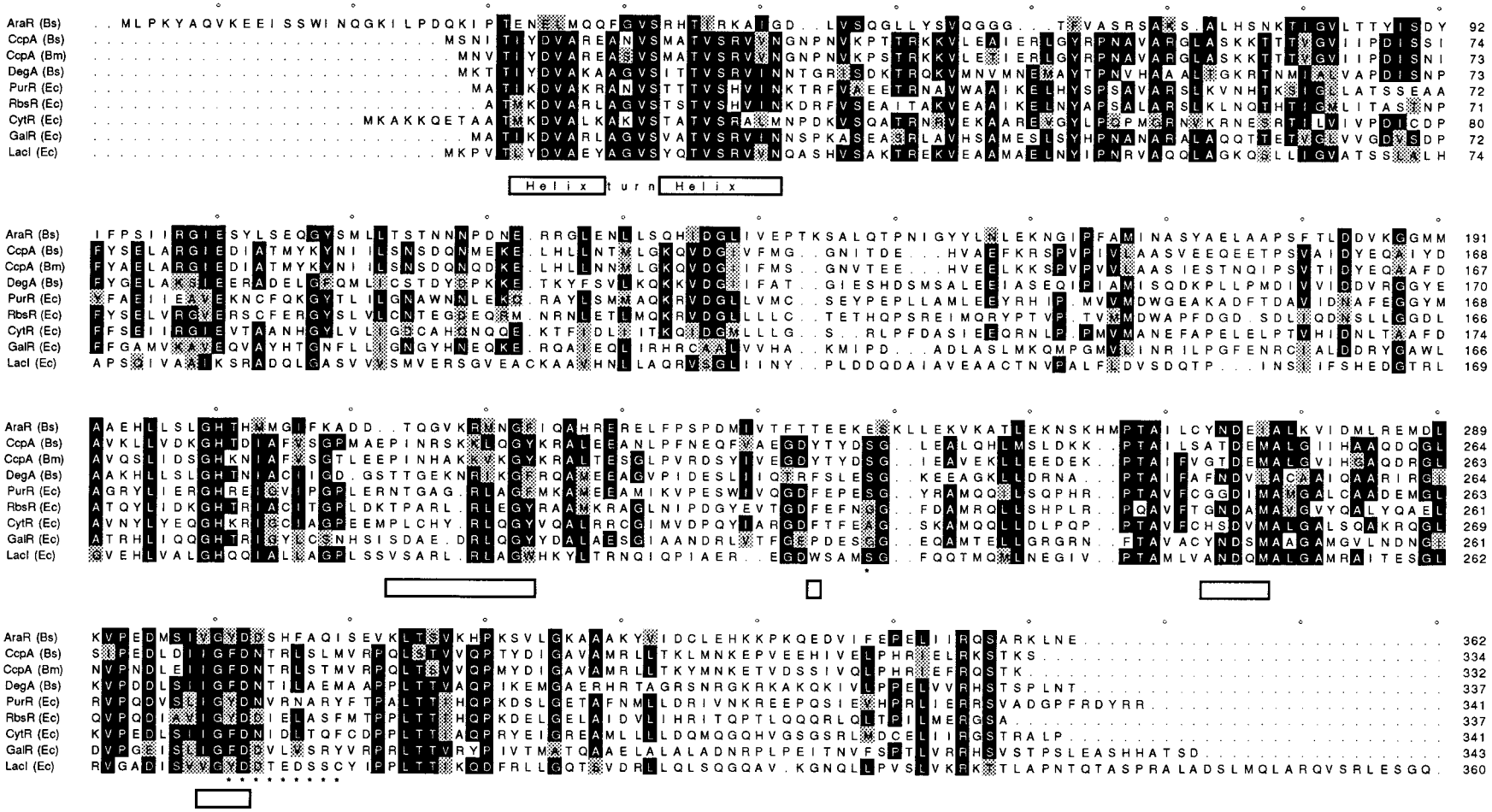

FIG. 3. Alignment of the predicted product of *araR* with members of the *E. coli* GalR-LacI family of regulatory proteins. Alignments were determined by using the PileUp program of the GCG package of sequence analysis software. Black boxes indicate identical residues in at least five of the nine sequences, highlighting in gray indicates conservative replacements, and numbers indicate positions in the amino acid sequences of the respective proteins. The HTH DNA binding region and the amino acids believed to contribute to inducer binding (open box) and dimerization (asterisk) of the proteins in the *E. coli* GalR-LacI family (56) are indicated below the sequence. Proteins indicated: DegA from *B. subtilis* (9); CytR, cytidine repressor from *E. coli* (55); PurR, purine repressor from *E. coli* (34); RbsR, ribose repressor from *E. coli* (33); CcpA, amylase repressor from *Bacillus megaterium* (26); CcpA, amylase repressor from *B. subtilis* (23); GalR, galactose repressor from *E. coli* (58); LacI, lactose repressor from *E. coli* (16).

L-arabinose (Fig. 5A). In the absence of L-arabinose, the level of accumulated β-galactosidase activity detected in the AraR$^-$ background (strain IQB214) was twofold higher than the maximal level observed in the wild-type strain IQB213 in the presence of L-arabinose. The patterns of *lacZ* expression observed with the *araR* null mutant in the absence and presence of L-arabinose were similar; however, addition of L-arabinose to the culture caused immediate cessation of cell growth (data not shown). The minimal L-arabinose concentration (0.02% [wt/vol]) necessary to achieve full induction of the *araABDLMNPQ-abfA* operon promoter was determined in the wild-type strain IQB213 grown in the same medium (data not shown). A study of the behavior of strain IQB215 (*araR*) in liquid minimal C medium supplemented with casein hydrolysate (1% [wt/vol]) after addition of different concentrations of L-arabinose during early-exponential-phase growth allowed us to establish a direct correlation between the cessation of growth of the AraR$^-$ mutant cells and the minimal L-arabinose concentration which corresponds to full induction of the *araABDLMNPQ-abfA* operon promoter in the wild-type strain IQB213 (data not shown; discussed below). These results, together with the similarity observed between the primary structure of the *araR* predicted product and several bacterial regulatory proteins, indicated that the *araR* gene encodes a *trans*-acting factor which is a negative regulator of the *araABDLMNPQ-abfA* operon.

RNA transcript and primer extension analysis of the *araR* gene. Northern blot analysis of RNA from cells grown with or without L-arabinose detected a transcript of 1.4 kb (Fig. 6), which is close to the size predicted for the *araR* transcript (1.2 kb), indicating that *araR* is monocistronic. The *araR* message was detected when the cells were grown in the absence and in the presence of L-arabinose but was much more abundant in L-arabinose grown cells, indicating that transcription of *araR* is stimulated by L-arabinose (Fig. 6).

Primer extension analysis to determine the transcription start site of the *araR* gene gave a single extension product with RNA isolated from cells grown in the presence of L-arabinose; the size of this product suggests that transcription of *araR* starts at a G residue 27 nt upstream from the *araR* start codon (Fig. 7). No extension product was seen with RNA isolated from cells grown in the absence of L-arabinose under the same conditions. Given the lower level of *araR* transcript in cells grown without L-arabinose (see above) and the fivefold increase of the expression of an *araR′-′lacZ* (Ara$^+$) transcriptional fusion in the presence of L-arabinose (see below), the lack of a primer extension product observed with RNA from cells grown without L-arabinose is likely due to the lower level of *araR* transcript in these cells. Centered at −35 and −10 bp upstream from the *araR* transcription start site are two sequences, TTTTCT and TATACT, respectively (Fig. 2), which are similar to the consensus −35 and −10 regions (TTGACA-17 bp-TATAAT), respectively, of promoters recognized by *B. subtilis* σ^A-containing RNA polymerase (36). The −35 region of *araR* is 2 bp from the putative −35 region of *araE* (Fig. 2); in addition there are two inverted repeats in the −10 region of the putative promoter of *araE* (Fig. 2). Interestingly, these two putative operator-like sequences are similar to each other, TT GTACGTAC-33 bp-AAGTACGTAC (shown in their antisense orientation strand, between positions 766 and 775 and positions 723 and 732, respectively [Fig. 2]), and display a high level of identity to two sequences found in the *araABDLMNPQ-abfA* operon promoter region, TTGTTCGTAC-32 bp-TAGTACG TAT (between positions −2 and +8 and positions +41 and +50, respectively, relative to the transcription start site [46]).

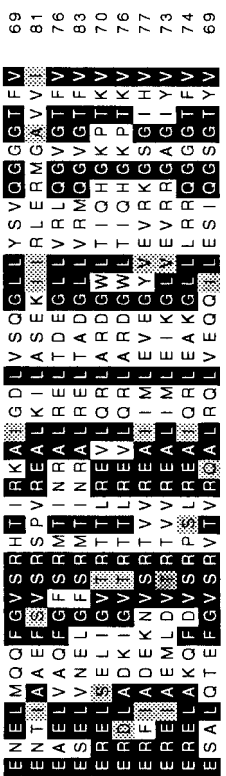

FIG. 4. Alignment of the amino-terminal domain of the predicted product of *araR* with repressor proteins belonging to the *B. subtilis* GntR family of regulatory proteins. Alignments were determined using the PileUp program of the GCG package of sequence analysis software. Black boxes indicate identical residues in at least 5 of the 10 sequences, highlighting in gray indicates conservative replacements, and numbers indicate positions in the amino acid sequences of the respective proteins. The consensus HTH motif, the signature sequence of the *B. subtilis* GntR family of regulatory proteins (21), is indicated. Proteins indicated: GntR, gluconate repressor from *B. subtilis* (19); HutC, histidine repressor from *Klebsiella aerogenes* (49) and *Pseudomonas putida* (2); FadR, repressor of fatty acid metabolism from *E. coli* (12) and *Haemophilus influenzae* (18); ExuR, regulator of the *exu* regulon from *E. coli* (41a); UxuR, regulator of the *uxu* regulon from *E. coli* (8); PdhR, repressor of the pyruvate dehydrogenase complex genes from *E. coli* (22); FarR, regulator of the succinyl synthetase operon from *E. coli* (42).

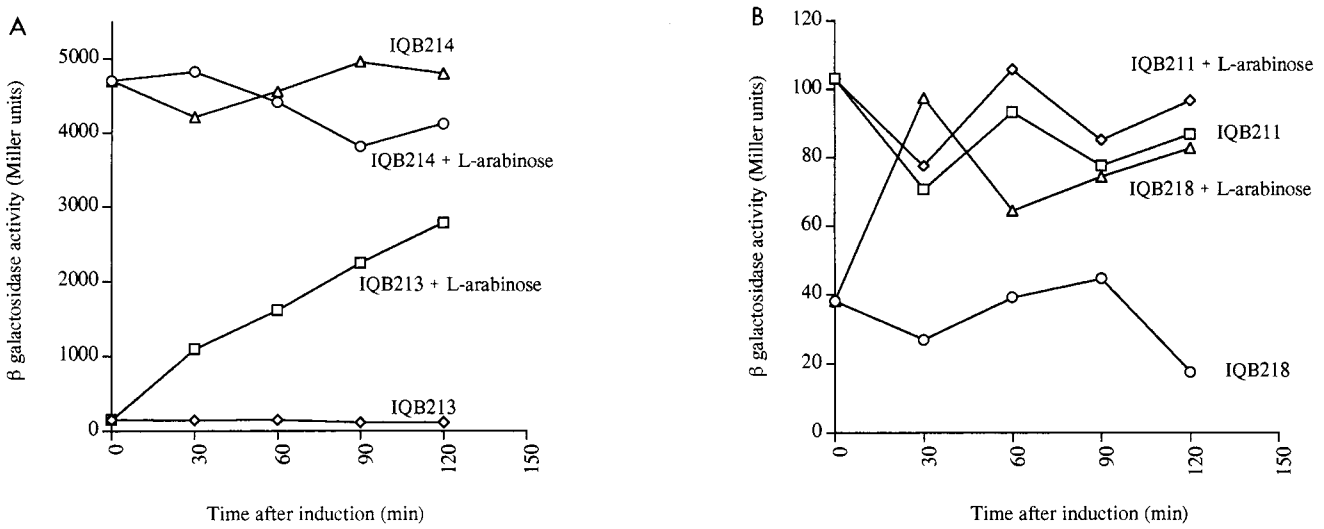

FIG. 5. Expression of the *araA′-lacZ* and *araR-lacZ* fusions in exponentially growing cells with and without L-arabinose. The *B. subtilis* strains were grown on C minimal medium supplemented with casein hydrolysate (1% [wt/vol]) in the presence or absence of L-arabinose (0.4% [wt/vol]) (see Materials and Methods). Time is expressed in minutes after induction; for each strain, the results represent the average from two independent experiments. (A) Expression of the *araABDLMNPQ-abfA* operon promoter in the wild-type and *araR* defective backgrounds. The strains of *B. subtilis* harboring a translational *araA′-lacZ* fusion are IQB213 *amyE′*::(*araA′-'lacZ cat*) (wild type) and IQB214 Δ*araR*::*km amyE′*::(*araA′-'lacZ cat*) AraR$^-$. When the negative control strains IQB216 and IQB217 were grown with or without L-arabinose, the level of accumulated β-galactosidase activity measured at 120 min, which corresponds to growth for at least 2.5 generations in the presence of the inducer, was less than 3.3 U (Miller units [data not shown]). (B) Expression of the *araR* gene promoter in the wild-type and *araR* defective backgrounds. The strains of *B. subtilis* harboring a transcriptional fusion *araR′-lacZ* are IQB218 (wild type) and IQB211 (AraR$^-$). The levels of accumulated β-galactosidase measured in the negative control, IQB212 (LacZ$^-$), in the presence or absence of L-arabinose were less than 5 U (Miller units [data not shown]).

Expression and regulation of *araR-lacZ* transcriptional fusions. To study the functionality of the *araR* promoter, we constructed transcriptional *araR-lacZ* fusions at the *araR* locus. Strain IQB211 (*araR′-lacZ cat* AraR$^-$) (Table 1) harbors a single copy of an *araR-lacZ* transcriptional fusion at the *araR* locus that disrupts the *araR* transcriptional unit. This strain was unable to grow on minimal medium plates with L-arabinose as the sole carbon source and displayed an Ara(Con) phenotype on rich medium, as determined by the filter paper assay (14). In contrast, strain IQB218 (*araR′-lacZ cat* AraR$^+$), which harbors a single copy of a nondisruptive *araR-lacZ* transcriptional fusion at the *araR* locus (see Materials and Methods), dis-

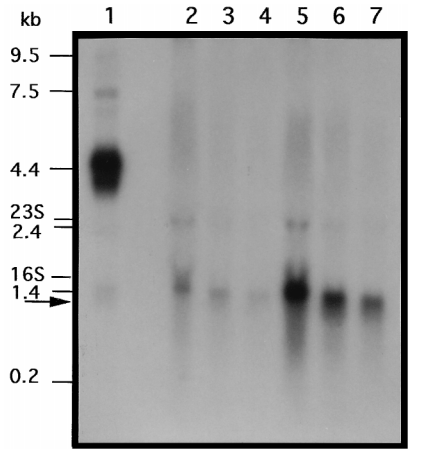

FIG. 6. Northern blot analysis of the *araR*-specific transcript. Lane 1, 4 μg of the RNA ladder (9.5 to 0.24 kb; GIBCO BRL); lanes 2 to 4, 20, 10, and 5 μg, respectively, of total RNA extracted from uninduced wild type strain *B. subtilis* 168T$^+$; lanes 5 to 7, 20, 10, and 5 μg, respectively, of total RNA extracted from the induced wild-type strain *B. subtilis* 168T$^+$ (see Materials and Methods). The samples were run in a 1.2% (wt/vol) agarose formaldehyde denaturing gel. The ^{32}P-labeled DNA probe used was synthesized from the 928-bp *Ava*I-*Hpa*I fragment (positions 1154 to 2082 [Probe in Fig. 1]). The RNA ladder was probed with ^{32}P-labeled λ DNA and also visualized by staining with ethidium bromide. The *araR* transcript of about 1.4 kb is indicated by an arrow. Positions of the rRNA bands are also presented.

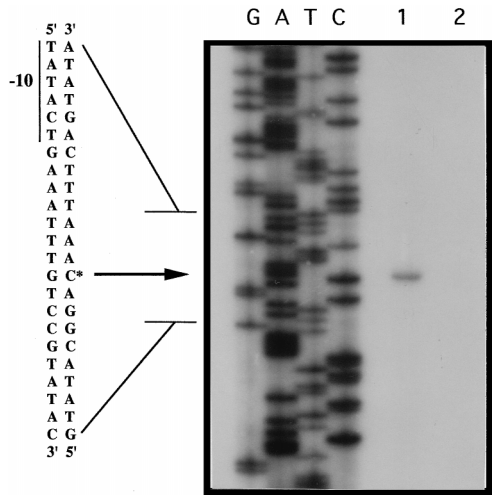

FIG. 7. Mapping of the transcriptional start site of the *araR* gene. A radiolabeled oligonucleotide, primer A (5′-ACCATACGGCCGACTGCGGC-3′), complementary to the 5′ end of the nontranscribed strand of the *araR* cistron was hybridized and used to direct cDNA synthesis from total *B. subtilis* 168T$^+$ RNA isolated from exponentially growing cells in the presence (lane 1) or absence (lane 2) of L-arabinose (see Materials and Methods). After extension, the products were analyzed by gel electrophoresis together with a set of dideoxynucleotide chain termination sequencing reactions performed with the same primer and a single-stranded M13 DNA template which includes the entire *araR* gene and the incomplete *araE* gene (positions 1 to 2126 [Fig. 1]). The arrow indicates the position of the *araR*-specific primer extension product, and the asterisk indicates the deduced start site of transcription, a G residue at position 847 in the *araR* sequence shown in Fig. 2.

played an Ara$^+$ phenotype. The β-galactosidase activities detected in strain IQB211 and IQB218 on rich medium plates as well as on minimal medium with glucose of L-arabinose confirmed that *araR* expression is not dependent on the presence of L-arabinose. Strain IQB212 (*araR'-cat lacZ* LacZ$^-$ AraR$^-$) (Table 1), in which the *lacZ* gene is in the orientation opposite that of the *araR* sequences (see Materials and Methods), displayed a LacZ$^-$ phenotype in the same media.

The β-galactosidase activities in these strains harboring *araR-lacZ* fusions were monitored in the presence or absence of L-arabinose (Fig. 5B). In the absence of L-arabinose, the levels of accumulated β-galactosidase detected in the AraR$^-$ background, strain IQB211, were approximately threefold higher than in wild-type strain IQB218 grown in the absence of L-arabinose and almost identical to the levels measured in wild-type strain IQB218 grown in the presence of L-arabinose. Addition of L-arabinose to the medium did not affect the expression in AraR$^-$ strain IQB211 but caused an immediate cessation of growth as observed in the other *araR* null mutant strains described above (discussed below). These data confirmed the prediction, made on the basis of the intensity of the signals detected in the studies of the *araR* transcript, that *araR* expression is stimulated by L-arabinose and also indicated that transcription of *araR* is autoregulated. Furthermore, these results indicate that the expression of the *araR* gene is driven from a weak promoter, which is in agreement with the results obtained by primer extension analysis.

DISCUSSION

In this report, we describe the nucleotide sequence of a regulatory gene, *araR*, which codes for a repressor of the *araABDLMNPQ-abfA* operon. An insertion-deletion mutation in the *araR* gene leads to constitutive expression of the L-arabinose metabolic operon, indicating that the *araR* gene product is a negatively acting regulatory protein, i.e., a repressor. Indeed, the predicted amino acid sequence of AraR revealed significant similarity to several members of a large family of bacterial regulators homologous to the *E. coli* Gal and Lac repressors (56). Interestingly, however, this similarity does not extend to the N-terminal region of AraR, in which an HTH motif, characteristic of nucleic acid binding proteins, was detected by similarity to the HTH consensus signature sequence of another group of bacterial repressors, the GntR family of regulatory proteins (21).

The product of the *araR* gene acts at the transcriptional level, as demonstrated by studies of expression from an *araA-lacZ* in-frame fusion in the wild-type and AraR$^-$ backgrounds. The classical mode for action of repressors, such as the *E. coli* LacI negative regulator (1), is that they regulate transcription through the binding as dimers or tetramers of identical subunits to specific DNA sequences which exhibit dyad symmetry. All apparently require effectors to modulate regulation, and most of the members of the GalR-LacI family bind carbohydrates effector molecules. By analogy, our results suggest that in the absence of L-arabinose, AraR protein binds to a site(s) within the *araABDLMNPQ-abfA* operon promoter region preventing transcription, and in the presence of L-arabinose a conformational change is induced in AraR such that recognition and binding to DNA is no longer possible and the operon can be expressed. This model predicts that there is a site(s) for AraR binding in the *araABDLMNPQ-abfA* operon promoter region. Candidates for such a sites are two regions of dyad symmetry which display sequence similarity to the two inverted repeats localized within the *araE*/*araR* promoter region. DNA protection studies with purified AraR would permit determination of the precise binding site(s).

In addition, the *araR* gene product negatively regulates its own expression. The *araR* gene is expressed in the absence of L-arabinose; however, addition of L-arabinose increases the expression of an *araR'-lacZ* fusion in the Ara$^+$ background about fivefold. The −35 region of the transcription start point for *araR* is located 2 bp from the putative −35 region of the *araE* gene, whose transcription is divergent from that of *araR* (Fig. 2). Studies with strains harboring *araE-lacZ* transcriptional fusions in wild-type and defective *araR* backgrounds indicate that the expression of *araE* is induced by L-arabinose and negatively regulated by AraR (37). Regions of divergent transcription represent a general type of gene organization (7). In this particular type, where one transcript determines a regulatory molecule and the other determines a nonregulatory polypeptide, it is common that the regulatory protein acts within the divergent transcription unit to control transcription of the nonregulatory polypeptides as observed in the L-arabinose *araC-araBAD* promoter region of *E. coli* (48) or the xylose *xylR-xylAB* promoter region of *B. subtilis* (20). To account for these observations, we propose that AraR binds to a site(s) within the *araE*/*araR* promoter in the absence of L-arabinose that prevents transcription from the *araE* promoter and simultaneously limits the frequency of initiation from its own promoter in order to maintain a low but constant intracellular concentration of its own transcript and product. Addition of L-arabinose will allow transcription from the *araE* promoter and also increase the frequency of initiation from the *araR* promoter.

The *araR* gene, partially present on the high-copy-number plasmid pSN1 (Fig. 1), was cloned by the ability of pSN1 to generate Ara$^+$ transformants in the recipient strain IGCg704 *araC4* (*araR4*) Ara$^-$. During the cloning process, the loss of the *araC4* (*araR4*) marker, from most of the plasmids initially recovered from the Ara$^+$ Em^r transformants, was probably the result of recombination with the chromosome. The *araC4* (*araR4*) mutation, conferring an Ara$^-$ phenotype to a strain bearing the *araA*, *araB*, and *araD* wild-type alleles, was localized in the C-terminal region of AraR, and insertional inactivation of *araR* leads to an Ara(Con) phenotype. According to the model of regulation proposed for the AraR protein, these observations suggest that the *araC4* (*araR4*) mutation lies in a site(s) involved in L-arabinose binding, thus preventing the necessary conformational change induced in AraR by the effector molecule to allow expression of the *ara* metabolic genes.

Apparently *B. subtilis* possesses two alternative L-arabinose transport systems, one encoded by the *araE* gene and the other encoded at least by the genes *araN*, *araP*, and *araQ* (46). Since both systems are under the control of *araR* gene product and deficiency of *araR* induces a twofold increase in the expression from the *araABDLMNPQ-abfA* operon promoter, we may speculate that the absence of AraR could lead to an intracellular increase of L-arabinose and consequently cause an increase in the concentration of the metabolic sugar phosphates intermediates which are toxic to the cell (28). This hypothesis would explain the immediate cessation of growth observed after the addition of L-arabinose to early-exponential-phase growth cultures of the AraR$^-$ mutant strains and is reinforced by the direct correlation established between the cessation of growth of AraR$^-$ mutants and the minimal L-arabinose concentration that corresponds to full induction of the *araABDLMNPQ-abfA* operon promoter in the wild type strain.

B. subtilis is the first gram-positive bacterium in which the L-arabinose system has been characterized, and the differences

found between the organization of the *B. subtilis* and *E. coli ara* genes may reflect differences in the mechanisms regulating L-arabinose utilization. In *E. coli*, the AraC protein plays a dual role in regulation, represses its own transcription, and activates the metabolic and transport operons (reference (48) and references therein). The results obtained so far for *B. subtilis* indicate that the product of the regulatory gene acts only as a repressor. According to the evolutionary history for prokaryotes based on the phylogenetic analysis of molecular sequences (59), *B. subtilis* is closer than *E. coli* to the origin of diversification of the three kingdoms, *Archaea*, *Bacteria*, and *Eucarya* (6); therefore, the *B. subtilis ara* regulon may represent an ancient version of the *ara* regulon of enteric bacteria.

ACKNOWLEDGMENTS

We thank H. de Lencastre, in whose laboratory this work was done, for her constant interest in this work and for critically reading the manuscript. We are grateful to M. Débarbouillé for the generous gift of the *B. subtilis* gene bank, to J. A. Almeida for helpful discussions, and Isabelle Martin-Verstraete, A. O. Henriques, and C. P. Moran, Jr., for gifts of plasmids.

This work was supported by grant 1287/92/Bio from Junta Nacional de Investigação Científica e Tecnológica (JNICT); L. J. Mota was the recipient of fellowships BJI/1666/93, BIC/1426/94, and BD/5689/95 from JNICT.

REFERENCES

1. **Adler, K., K. Beyreuther, E. Fanning, N. Geisler, B. Gronenborn, A. Klemm, B. Muller-Hill, M. Pfahl, and A. Schmitz.** 1972. How *lac* repressor binds to DNA. Nature (London) **237:**322–327.
2. **Allison, S. L., and A. T. Phillips.** 1990. Nucleotide sequence of the gene encoding the repressor for the histidine utilization genes of *Pseudomonas putida*. J. Bacteriol. **172:**5470–5476.
3. **Anagnostopoulos, C., and J. Spizizen.** 1961. Requirements for transformation in *Bacillus subtilis*. J. Bacteriol. **81:**741–746.
4. **Arantes, O., and D. Lereclus.** 1991. Construction of cloning vectors for *Bacillus thuringiensis*. Gene **108:**115–119.
5. **Barnell, W. O., K. C. Yi, and T. Conway.** 1990. Sequence and genetic organization of a *Zymomonas mobilis* gene cluster that encodes several enzymes of glucose metabolism. J. Bacteriol. **172:**7227–7240.
6. **Barns, S. M., C. F. Delwiche, J. D. Palmer, and N. R. Pace.** 1996. Perspectives on archeal diversity, thermophily and monophyly from environmental rRNA sequences. Proc. Natl. Acad. Sci. USA **93:**9188–9193.
7. **Beck, C. F., and R. A. J. Warren.** 1988. Divergent promoters, a common form of gene organization. Microbiol. Rev. **52:**318–326.
8. **Burland, V., G. Plunkett III, H. J. Sofia, D. L. Daniels, and F. R. Blattner.** 1995. Analysis of the *Escherichia coli* genome. VI. DNA sequence of the region from 92.8 through 100 minutes. Nucleic Acids Res. **23:**2105–2119.
9. **Bussey, L. B., and R. L. Switzer.** 1993. The *degA* gene product accelerates degradation of *Bacillus subtilis* phosphoribosylpyrophosphate amidotransferase in *Escherichia coli*. J. Bacteriol. **175:**6348–6353.
10. **Dale, R. M. K., B. A. MaClure, and J. P. Houchins.** 1985. A rapid single-stranded cloning strategy for producing a sequential series of overlapping clones for use in DNA sequencing: application to sequencing the corn mitochondrial 18S RDNA. Plasmid **13:**31–40.
11. **Davis, E. O., and P. J. Henderson.** 1987. The cloning and DNA sequence of the gene *xylE* for xylose-proton symport in *Escherichia coli* K12. J. Biol. Chem. **262:**13928–13932.
12. **DiRusso, C. C.** 1988. Nucleotide sequence of the *fadR* gene, a multifunctional regulator of fatty acid metabolism in *Escherichia coli*. Nucleic Acids Res. **16:**7995–8009.
13. **Englesberg, E., R. 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.
14. **Englesberg, E., J. Irr, J. Power, and N. Lee.** 1965. Positive control of enzyme synthesis by gene C in the L-arabinose operon. J. Bacteriol. **90:**946–957.
15. **Englesberg, E., C. Squires, and F. Meronk.** 1969. The arabinose operon in *Escherichia coli* B/r: a genetic demonstration of two functional states of the product of a regulatory gene. Proc. Natl. Acad. Sci. USA **80:**6790–6794.
16. **Farabaugh, P. J.** 1978. Sequence of the *lacI* gene. Nature **274:**765–769.
17. **Ferrari, E., A. Nguyen, D. Lang, and J. Hoch.** 1983. Construction and properties of an integrable plasmid for *Bacillus subtilis*. J. Bacteriol. **154:** 1513–1515.
18. **Fleischmann, R. D., M. D. Adams, O. White, R. A. Clayton, E. F. Kirkness, A. R. Kerlavage, C. J. Bult, J.-F. Tomb, B. A. Dougherty, J. M. Merrick, K. McKenney, G. Sutton, W. Fitzhugh, C. A. Fields, J. D. Gocayne, J. D. Scott, R. Shirley, L.-I. Liu, A. Glodek, J. M. Kelley, J. F. Weidman, C. A. Phillips, T. Spriggs, E. Hedblom, M. D. Cotton, T. R. Utterback, M. C. Hanna, D. T. Nguyen, D. M. Saudek, R. C. Brandon, L. D. Fine, J. L. Fritchman, J. L. Fuhrmann, N. S. M. Geoghagen, C. L. Gnehm, L. A. McDonald, K. V. Small, C. M. Fraser, H. O. Smith, and J. C. Venter.** 1995. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. Science **269:**496–512.
19. **Fujita, Y., T. Fujita, Y. Miwa, J. Nihashi, and Y. Aratani.** 1986. Organization and transcription of the gluconate operon, *gnt*, of *Bacillus subtilis*. J. Biol. Chem. **261:**13744–13753.
20. **Gärtner, D., J. Degenkolb, J. Rippberger, R. Allmansberger, and W. Hillen.** 1992. Regulation of *Bacillus subtilis* W23 xylose utilization operon: interaction of Xyl repressor with *xyl* operator and the inducer xylose. Mol. Gen. Genet. **232:**415–422.
21. **Haydon, D. J., and J. R. Guest.** 1991. A new family of bacterial regulatory proteins. FEMS Microbiol. Lett. **79:**291–296.
22. **Haydon, D. J., M. A. Quail, and J. R. Guest.** 1993. A mutation causing constitutive synthesis of the pyruvate dehydrogenase complex in *Escherichia coli* is located within the *pdhR* gene. FEBS Lett. **336:**43–47.
23. **Henkin, T. M., F. J. Grundy, W. L. Nicholson, and G. H. Chambliss.** 1991. Catabolite repression of α-amylase gene expression in *Bacillus subtilis* involves a *trans*-acting gene product homologous to the *Escherichia coli lacI* and *galR* repressors. Mol. Microbiol. **5:**575–584.
24. **Henriques, A. O., B. W. Beall, and C. P. Moran, Jr.** CotM of *Bacillus subtilis*, a member of the α-crystallin family of stress proteins, is induced during development and participates in spore outer coat formation. *J. Bacteriol.*, in press.
25. **Henriques, A. O., and C. P. Moran, Jr.** Unpublished results.
26. **Hueck, C., A. Kraus, and W. Hillen.** 1994. Sequences of *ccpA* and two downstream *Bacillus megaterium* genes with homology to the *motAB* operon from *Bacillus subtilis*. Gene **143:**147–148.
27. **Igo, M. M., and R. Losick.** 1986. Regulation of a promoter that is utilized by minor forms of RNA polymerase holoenzyme in *Bacillus subtilis*. J. Mol. Biol. **191:**615–624.
28. **Kadner, R. J., G. P. Murphy, and C. M. Stephens.** 1992. Two mechanisms for growth inhibition by elevated transport of sugar phosphates in *Escherichia coli*. J. Gen. Microbiol. **138:**2007–2014.
29. **Kyte, J., and R. F. Doolittle.** 1982. A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. **157:**105–132.
30. **Lepesant, J. A., and R. Dedonder.** 1967. Metabolism du L-arabinose chez *Bacillus subtilis*. Marburg Ind. 168. C. R. Acad. Sci. Ser. D **264:**2683–2686.
31. **Martin, I., M. Débarbouillé, E. Ferrari, A. Klier, and G. Rapoport.** 1987. Characterization of the levanase gene of *Bacillus subtilis* which shows homology to yeast invertase. Mol. Gen. Genet. **208:**177–184.
32. **Martin-Verstraete, I., M. Débarbouillé, A. Klier, and G. Rapoport.** 1992. Mutagenesis of the *Bacillus subtilis* "−12, −24" promoter of the levanase operon and evidence of existence of an upstream activating sequence. J. Mol. Biol. **226:**85–99.
33. **Mauzy, C. A., and M. A. Hermodson.** 1992. Structural homology between *rbs* repressor and ribose binding protein implies functional similarity. Protein Sci. **1:**843–849.
34. **Meng, L. M., M. Kilstrup, and P. Nygaard.** 1990. Autoregulation of PurR repressor synthesis and involvement of *purR* in the regulation of *purB*, *purC*, *purL*, *purMN* and *guaBA* expression in *Escherichia coli*. Eur. J. Biochem. **187:**373–379.
35. **Miller, J. H.** 1972. Experiments in molecular genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
36. **Moran, C. P., Jr., N. Lang, S. F. J. LeGrice, G. Lee, M. Stephens, A. L. Sonensheim, J. Pero, and R. Losick.** 1982. Nucleotide sequences that signal the initiation of transcription in *Bacillus subtilis*. Mol. Gen. Genet. **186:**339–346.
37. **Mota, L. J., and I. Sá-Nogueira.** Unpublished results.
38. **Pascal, M., F. Kunst, J. A. Lepesant, and R. Dedonder.** 1971. Characterization of two sucrase activities in *Bacillus subtilis* Marburg. Biochemistry **53:** 1059–1066.
39. **Paveia, H., and L. Archer.** 1992. Genes for L-arabinose utilization in *Bacillus subtilis*. Broteria Genet. Lisboa **13:**149–159.
40. **Paveia, H., and L. Archer.** 1992. Mapping of *ara* genes in *Bacillus subtilis*. Broteria Genet. Lisboa **13:**161–167.
41. **Perego, M.** 1993. Integrational vectors for genetic manipulation in *Bacillus subtilis*, p. 615–624. *In* A. L. Sonensheim, J. A. Hoch, and R. Losick (ed.), *Bacillus subtilis* and other gram-positive bacteria: biochemistry, physiology, and molecular genetics. American Society for Microbiology, Washington, D.C.

41a. **Plunkett, G., III.** Unpublished data; GenBank accession number P42608.

42. **Quail, M. A., C. E. Dempsey, and J. R. Guest.** 1994. Identification of a fatty acyl responsive regulator (FarR) in *Escherichia coli*. FEBS Lett. **356:**183–187.
43. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
44. **Sanger, F. S., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with

chain-terminating inhibition. Proc. Natl. Acad. Sci. USA **74:**140–144.
45. **Sá-Nogueira, I., and H. Lencastre.** 1989. Cloning and characterization of *araA*, *araB* and *araD*, the structural genes for L-arabinose utilization in *Bacillus subtilis*. J. Bacteriol. **171:**4088–4091.
46. **Sá-Nogueira, I., T. V. Nogueira, S. Soares, and H. Lencastre.** 1997. The L-arabinose (*ara*) operon of *Bacillus subtilis*: nucleotide sequence, genetic organization and expression. Microbiology **143:** 957–969.
47. **Sá-Nogueira, I., H. Paveia, and H. Lencastre.** 1988. Isolation of constitutive mutants for L-arabinose utilization in *Bacillus subtilis*. J. Bacteriol. **170:**2855–2857.
48. **Schleif, R.** 1996. Two positively regulated systems, *ara* and *mal*, p. 1300–1309. *In* F. C. Neidhardt, R. Curtiss III, J. L. Ingraham, E. C. C. Lin, K. Low, B. Magasanik, W. S. Reznikoff, M. Riley, M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella*: cellular and molecular biology, 2nd ed. American Society for Microbiology, Washington, D.C.
49. **Schwacha, A., and R. A. Bender.** 1990. Nucleotide sequence of the gene encoding the repressor for the histidine utilization genes of *Klebsiella aerogenes*. J. Bacteriol. **172:**5477–5481.
49a. **Shatwell, K. P., B. M. Charalambous, T. P. McDonald, and P. J. F. Henderson.** 1995. Cloning, sequencing, and expression of the *araE* gene of *Klebsiella oxytoca* S017 which encodes arabinose-H^+ symport activity. *J. Bacteriol.* **177:** 5379–5380.
50. **Shimotsu, H., and D. J. Henner.** 1986. Construction of a single-copy integration vector and its use in analysis of regulation of the *trp* operon of *Bacillus subtilis*. Gene **43:**85–94.
51. **Stoner, C., and R. Schleif.** 1983. The *araE* low affinity L-arabinose transport promoter. Cloning, sequence, transcription start site and DNA binding sites of regulatory proteins. J. Mol. Biol. **171:**369–381.
52. **Studier, F. W., A. H. Rosenberg, J. J. Dunn, and J. W. Dubendorff.** 1990. Use of T7 polymerase to direct expression of cloned genes. Methods Enzymol. **185:**60–89.
53. **Sullivan, M. A., R. E. Yasbin, and F. E. Young.** 1984. New shuttle vectors for *Bacillus subtilis* and *Escherichia coli* which allow rapid detection of inserted fragments. Gene **29:**21–26.
54. **Tinoco, I., P. N. Borer, B. Dengler, M. D. Levine, O. C. Uhlenbeck, D. M. Crothers, and J. Gralla.** 1973. Improved estimation of secondary structure in ribonucleic acids. Nature New Biol. **246:**40–41.
55. **Valentin-Hansen, P., J. E. Larsen, P. Hojrup, S. A. Short, and C. S. Barbier.** 1986. Nucleotide sequence of the *cytR* regulatory gene of *E. coli* K-12. Nucleic Acids Res. **14:**2215–2228.
56. **Weickert, M. J., and S. Adhya.** 1992. A family of bacterial regulators homologous to Gal and Lac repressors. J. Biol. Chem. **267:**15869–15874.
57. **Weickert, M. J., and G. H. Chambliss.** 1990. Site-directed mutagenesis of a catabolic repression operator sequence in *Bacillus subtilis*. Proc. Natl. Acad. Sci. USA **87:**6238–6242.
58. **von Wilcken-Bergmann, B., and B. Muller-Hill.** 1982. Sequence of *galR* gene indicates a common evolutionary origin of *lac* and *gal* repressor in *Escherichia coli*. Proc. Natl. Acad. Sci. USA **79:**2427–2431.
59. **Woese, C. R.** 1987. Bacterial evolution. Microbiol. Rev. **51:**221–271.
60. **Yanisch-Perron, C., J. Vieira, and J. Messing.** 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene **33:**103–119.
61. **Zhang, C. C., M. C. Durand, R. Jeanjean, and F. Joset.** 1989. Molecular and genetical analysis of the fructose-glucose transport system in the cyanobacterium *Synechocystis*. Mol. Microbiol. **3:**1221–1229.