JOURNAL OF BACTERIOLOGY, Dec. 1997, p. 7705–7711 Vol. 179, No. 24
0021-9193/97/$04.00+0

Cloning, Functional Analysis, and Transcriptional Regulation of the *Bacillus subtilis araE* Gene Involved in L-Arabinose Utilization

ISABEL SÁ-NOGUEIRA* AND SANDRA S. RAMOS

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

Received 2 September 1997/Accepted 16 October 1997

The *Bacillus subtilis araR* locus (mapped at about 294° on the genetic map) comprises two open reading frames with divergently arranged promoters, the regulatory gene, *araR*, encoding a repressor, and a partially cloned gene, termed *araE* by analogy to the *Escherichia coli* L-arabinose permease gene. Here, we report the cloning and sequencing of the entire *araE* gene encoding a 50.4-kDa polypeptide. The *araE* gene is monocistronic (as determined by Northern blot analysis), and its putative product is very similar to a number of prokaryotic proton-linked monosaccharide transporters (the group I family of membrane transport proteins). Insertional inactivation of the *araE* gene leads to a conditional Ara$^-$ phenotype dependent on the concentration of L-arabinose in the medium. Therefore, we assume that *araE* encodes a permease involved in L-arabinose transport into the cell. The *araE* promoter region contains −10 and −35 regions (as determined by primer extension analysis) very similar to those recognized by RNA polymerase containing the major vegetative-cell sigma factor σ^A, and the −35 region of the transcription start point for *araE* is located 2 bp from the −35 region of the *araR* gene. Transcriptional studies demonstrated that the expression from the *araE* promoter is induced by L-arabinose, repressed by glucose, and negatively regulated by AraR. These observations are consistent with a model according to which in the absence of L-arabinose, AraR binds to a site(s) within the *araE*/*araR* promoter, preventing transcription from the *araE* promoter and simultaneously limiting the frequency of initiation from its own promoter; the addition of L-arabinose will allow transcription from the *araE* promoter and increase the frequency of initiation from the *araR* promoter.

Bacillus subtilis can use L-arabinose as its sole carbon and energy source. The ability to utilize L-arabinose is dependent on three intracellular enzymes, L-arabinose isomerase, L-ribulokinase, and L-ribulose-5-phosphate-4-epimerase, which sequentially convert L-arabinose to L-ribulose, L-ribulose-5-phosphate, and D-xylulose-5-phosphate, respectively (8). D-Xylulose-5-phosphate is further catabolized through the pentose-phosphate pathway. This metabolic pathway is identical to the one found in *Escherichia coli*, and its enzymes were previously described as inducible by L-arabinose (8).

The three metabolic genes, *araA*, *araB*, and *araD*, coding for L-arabinose isomerase, L-ribulokinase, and L-ribulose-5-phosphate-4-epimerase, respectively, are adjacent and constitute the first three open reading frames of a nine-cistron transcriptional unit with a total length of 11 kb (22, 24). This operon, called *araABDLMNPQ-abfA*, is located at about 256° on the *B. subtilis* genetic map (18) and includes 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 a deletion mutation in the region downstream from *araD* revealed that the *araL*, *araM*, *araN*, *araP*, *araQ*, and *abfA* genes are not essential for L-arabinose utilization (24). Transcriptional studies demonstrated that the expression of the *araABDLMNPQ-abfA* operon is directed by a strong σ^A-like promoter identified upstream from the translation start site of the *araA* gene and that expression from the *araABDLMNPQ-abfA* promoter is induced by L-arabinose and repressed by glucose (24). The *araR* locus defined by three additional classes of mutations affecting L-arabinose utilization (17, 18, 25), at about 294° on the *B. subtilis* genetic map, was recently cloned (23). This region comprises two open reading frames with divergently arranged promoters, the regulatory gene, *araR*, and a partially cloned gene, termed *araE*. The predicted amino acid sequence of AraR 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. Insertional inactivation of the *araR* gene and transcriptional studies demonstrated that this 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 (23).

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* (23, 24) and *E. coli ara* (reference 26 and references therein) genes may reflect differences in the mechanisms regulating L-arabinose utilization. In *E. coli* the AraC protein plays a dual role in regulation (reference 26 and references therein). In the absence of L-arabinose, AraC represses its own transcription via a DNA-looping mechanism in the divergently transcribed *araC-araBAD* promoter region. In the presence of L-arabinose, it activates the *araBAD* metabolic operon and the *araFGH* and *araE* transport genes. In *B. subtilis*, the results obtained so far

* Corresponding author. Mailing address: Instituto de Tecnologia Química e Biológica, Av. da República, Apartado 127, 2780 Oeiras Codex, Portugal. Phone: 351-1-4469524. Fax: 351-1-4411277. E-mail: sanoguei@itqb.unl.pt.

TABLE 1. *B. subtilis* strains used in this study

Strain	Genotype	Relevant phenotype	Source or reference[a]
168T$^+$	Prototroph	Ara$^+$	F. E. Young
IGCg707	*araB7 metB10 lys-3*	Ara$^-$	18
168T$^+$ derivatives			
IQB206	Δ*araL-abfA*::*spc*	Ara$^+$ Sp^r	24
IQB215	Δ*araR*::*km*	Ara (Con) Km^r	23
IQB230	*araE'*::pSR1(*araE'-lacZ cat*)	Ara$^-$ LacZ$^+$ Cm^r	pSR1→168T$^+$
IQB231	*araE'*::pSR2(*araE'-cat lacZ*)	Ara$^-$ LacZ$^-$ Cm^r	pSR2→168T$^+$
IQB232	*araE*::pSN31(*araE'-lacZ cat*)	Ara$^+$ LacZ$^+$ Cm^r	pSN31→168T$^+$
IQB233	Δ*araR*::*km araE*::pSN31(*araE'-lacZ cat*)	Ara (Con) LacZ$^+$ Cm^r Km^r	pSN31→IQB215
IQB234	Δ*araL-abfA*::*spc araE'*::pSR1(*araE'-lacZ cat*)	Ara$^+$ LacZ$^+$ Sp^r Cm^r	pSR1→IQB206

[a] Arrows indicate transformation and point from donor DNA to recipient strain.

indicate that the product of the regulatory gene acts only as a repressor. The proposed model for the action of AraR (23) is that in the absence of L-arabinose, AraR protein binds to a site(s) within the *araABDLMNPQ-abfA* operon promoter and *araE*/*araR* promoter region, preventing transcription, and that in the presence of L-arabinose, a conformational change is induced in AraR such that recognition and binding to DNA is no longer possible, allowing expression of the *ara* genes. Here we report the cloning and complete sequence of the *araE* gene. We demonstrate that the *araE* gene codes for a permease involved in L-arabinose transport into the cell and that the expression from the *araE* promoter is regulated at the transcriptional level by (i) L-arabinose induction, (ii) glucose repression, and (iii) the repressor AraR.

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 Luria-Bertani (LB) medium (12). For cultivation of *E. coli*, ampicillin (75 μg ml^{-1}), chloramphenicol (30 μg ml^{-1}), X-Gal (5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside [40 μg ml^{-1}]), or IPTG (isopropyl-β-D-thiogalactopyranoside [1 mM]) were added as appropriate. *B. subtilis* strains were grown on LB (12), SP (10), or C minimal medium (16) supplemented with auxotrophic requirements (100 μg ml^{-1}). For cultivation of *B. subtilis*, chloramphenicol (5 μg ml^{-1}), kanamycin (15 μg ml^{-1}), or spectinomycin (60 μg ml^{-1}) was added as appropriate. Solid medium was made with LB, SP, or C medium containing 1.5% (wt/vol) Bacto agar (Difco). The Ara phenotype was screened on minimal C medium plates containing 0.1% (wt/vol) L-arabinose after 48 h of incubation at 37°C. *B. subtilis* and *E. coli* DNA transformations were performed as previously described (23).

DNA manipulations and sequencing. DNA manipulations were carried out according to standard methods (20). Sequencing templates were prepared by a combination of subcloning appropriate fragments from pSR5 into the polycloning site of M13mp19 or M13mp18 and sequential deletion of the recombinant M13 derivatives, as previously described (23). The DNA sequences for both strands and across all the restriction sites used for subcloning were resolved by the dideoxy-chain termination method with the Sequenase kit (T7 DNA polymerase; United States Biochemical Corporation). A *lacZ*-specific primer (23) was used to sequence the transcriptional *lacZ* fusions. Sequences were analyzed with the programs DNASIS V2.0, 1991 version (Hitachi Software Engineering Co., Ltd.) and the Genetics Computer Group (Madison, Wis.) package of sequence analysis software.

Plasmid constructions. Plasmids pSN29 and pSN30 were obtained by subcloning a 473-bp *Xmn*I-*Rsr*II (fill-in) DNA fragment (position 1509 to 1982, Fig. 1) from pLM2 (23) at the unique *Sma*I restriction site of the shuttle vector pMK3-1 (29), in both orientations. To construct pSR1 and pSR2, the *Bam*HI-*Eco*RI DNA fragments from pSN29 and pSN30, respectively, were subcloned into the integrational vector pJM783 (19) digested with *Bam*HI and *Eco*RI. Plasmid pSR1 contains the *B. subtilis spoVG* ribosome binding site and the *E. coli lacZ* gene in the same orientation as the *araE* region sequences and pSR2 in the opposite orientation. Plasmid pSR5 was obtained by cutting chromosomal DNA from *B. subtilis* IQB230 (Table 1) with *Nco*I, followed by circularization of this DNA at a low concentration. This ligation mixture was transformed into *E. coli*, and Amp^r transformants were selected as described above. To construct plasmid pSN31, a 372-bp *Hpa*II (fill-in)-*Xmn*I DNA fragment (position 1137 to 1509, Fig. 1) from pLM2 (23) was subcloned at the unique *Sma*I restriction site of the integrational vector pJM783 (19). Plasmid pSN31 contains the *B. subtilis spoVG* ribosome binding site and the *E. coli lacZ* gene in the same orientation as the *araE* region sequences.

Construction of strains bearing an *araE-lacZ* transcriptional fusion. The integrational plasmids pSR1, pSR2, and pSN31 (described above) were used separately to transform *B. subtilis* 168T$^+$, IQB215 (23), or IQB206 (24) to Cm^r and integrated as a single copy into the chromosome by a Campbell-type recom-

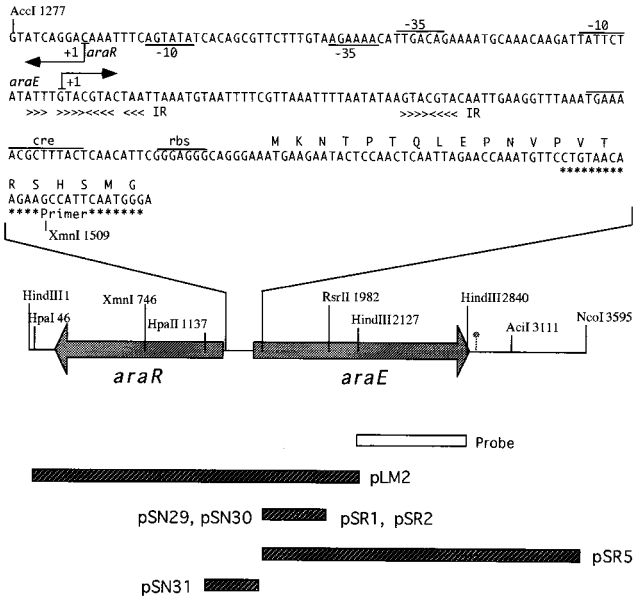

FIG. 1. Physical and genetic map of the *araR*/*araE* region of the chromosome. The location of the *araR* and *araE* coding regions, predicted from the analysis of the nucleotide sequence, are indicated by arrows pointing in the direction of transcription. The positions for relevant restriction sites from the nucleotide sequence and the region of dyad symmetry that could represent the transcription terminator of the *araE* transcriptional unit are indicated. Above the physical map, the nucleotide sequence of the intergenic *araR-araE* region (23), *araE* nontranscribed strand and *araR* transcribed strand, is shown in the 5′-to-3′ direction. The predicted primary structure of the polypeptide encoded by *araE* (above the nucleotide sequence) is given in single-letter code. The *araE* and *araR* (23) 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 above the nucleotide sequence, *araE*, and below the nucleotide sequence, *araR*. Convergent arrows represent different regions of dyad symmetry (IR). The complementary sequence of the primer, used in primer extension analysis, is represented below the sequence. The putative catabolic repression associated sequence (cre) is overscored. The open box below the physical map denotes the fragment used as a probe for Northern analysis of the *araE* transcript, and the striped boxes represent the extension of the inserts in the indicated plasmids. Plasmids pSR1, pSR2, and pSN31 were integrated into the host chromosome by means of a single crossover (Campbell type) recombinational event that occurred in the region of homology.

binational event (confirmed by Southern hybridization [data not shown]). The integration of pSR1 and pSR2 at the *araE* locus disrupted the *araE* transcriptional unit (Fig. 1) in the resulting strains IQB230 or IQB234 and IQB231 (Table 1), respectively. The integration of pSN31 at the *araE* locus was not disruptive (Fig. 1) in the resulting strains IQB232 and IQB233 (Table 1).

β-Galactosidase assays. Strains of *B. subtilis* harboring transcriptional *lacZ* fusions were grown on 45 ml of C medium supplemented with casein hydrolysate (1% [wt/vol]). During the early logarithmic phase (optical density at 600 nm [OD_{600}], 0.11 to 0.15), 15 ml of the culture was transferred to two different flasks, and L-arabinose at a final concentration of 0.4% (wt/vol) or L-arabinose (0.4% [wt/vol]) and D-glucose (0.4% [wt/vol]) were added. Exponential growth of the cultures was followed 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 and treated as described previously (23). β-Galactosidase activity was determined according to the method of Miller (12) with modifications as described before (23). Strain IQB231 ($LacZ^-$ [Table 1]) was used as negative control in the assays, and the levels of accumulated β-galactosidase measured after 2.5 generations in the presence of L-arabinose were less than 7.7 Miller units.

RNA preparation and 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) as described above for the β-galactosidase assays. Cells were harvested when 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, and RNA was prepared as described before (23). For Northern blot analysis, 3.5 and 7.5 μg of total RNA was run in a 1.2% (wt/vol) agarose formaldehyde denaturing gel and transferred as described previously (23). A size determination was done with an RNA ladder (9.5 to 0.24 kb; GIBCO BRL). The DNA probe was labeled with [α-^{32}P]dATP (6,000 Ci/mmol) with the Multiprime random oligonucleotide DNA labeling system, obtained from Amersham. Primer extension analysis was performed as previously described (23). The primer (5′-GGACATTGTTCTTC GGTAAGTTACCC-3′) complementary to nucleotides 1496 to 1521 (Fig. 1) was end labeled with [γ-^{32}P]ATP (6,000 Ci/mmol) with T4 polynucleotide kinase. A total of 5 ng of labeled primer was mixed with 40 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 22 U of avian myeloblastosis virus reverse transcriptase (United States Biochemical Corporation) for 2 h at 37°C, and analysis of the extended products was carried out as described previously (23).

Nucleotide sequence accession number. The DNA sequences reported here (positions 2127 to 3111, Fig. 1) have been submitted to the EMBL database and assigned no. Y12105.

RESULTS AND DISCUSSION

Cloning of the 3′ end of the *araE* gene and sequence analysis. In a previous study we reported the cloning and sequencing of the *araE* 5′-end region (23). Sequences downstream of the pLM2 insert (Fig. 1) were cloned in a chromosomal walking step involving the use of *Nco*I-digested and religated DNA from a Campbell integrant of pSR1, strain IQB230 (see Materials and Methods). This procedure created pSR5, which carried an additional 1.6 kb of DNA adjacent to the previously cloned fragment in pSR1 (Fig. 1). The structure of the insert in pSR5 was compared to that of the corresponding area of chromosomal DNA by Southern blot analysis (data not shown), and the results revealed that no detectable rearrangement occurred during the cloning process.

DNA sequence analysis of the *araE* 3′-end region revealed the presence of one potential hairpin-loop structure, situated next to the UGA stop codon of the *araE* gene (positions 2867 to 2901 [Fig. 1]). This structure has a ΔG value of −24.4 kcal mol^{-1} (30) and probably corresponds to a transcription terminator. The *araE* coding region could encode a 464-amino-acid product of 50.4 kDa (Fig. 2). The predicted product of *araE*, a hydrophobic protein (hydropathic index, 0.82 [7]), exhibited a hydropathy profile (not shown) characteristic of integral membrane proteins. The 12 major regions of high hydrophobicity, each composed of 21 amino acids which could be capable of spanning the membrane, and the hydrophilic regions are arranged alternately (Fig. 2).

The *araE* gene product is related to the group I membrane transport proteins. AraE was found to have sequence similarity to a family of membrane transport proteins, group I, which includes sugar transporters from organisms as diverse as bacteria, fungi, algae, plants, and humans (3). The alignment shown in Fig. 2 includes six members of the family that most resemble AraE, with identities that range from 31.5 to 39.9%. All these proteins are prokaryotic proton-linked monosaccharide transporters, and their physiological substrates are L-arabinose (AraE from *E. coli* [28] and *Klebsiella oxytoca* [27]), D-xylose (XylE from *E. coli* [2]), D-galactose (GalP from *E. coli* [5]), and D-glucose (Glf from *Zymomonas mobilis* [1] and GlcP from *Synechocystis* sp. strain PCC 6083 [35]). *B. subtilis* AraE also had sequence similarity (35.7% identity, not shown) with two hypothetical metabolite transport proteins from *B. subtilis*, YfiG (33) and YxbC (34).

It has been suggested that these transport proteins have evolved from the duplication of an ancestral protein with six transmembrane regions. This hypothesis is based on the conservation of the two G-R-[KR] motifs, their similarity in predicted structure, amino acid sequences, and conservation of particular motifs (3, 4, 9). The primary structure of AraE displays all the features of the homologous sugar transport proteins, 12 putative transmembrane domains, a central hydrophilic region, and several motifs which presumably have structural and/or functional significance (Fig. 2). These observations strongly suggest that the *B. subtilis araE* gene encodes a permease involved in L-arabinose transport into the cell.

RNA transcript and primer extension analysis of the *araE* gene. Northern hybridization analysis of total RNA from cells grown in the presence of L-arabinose, with a 0.9-kb DNA probe containing *araE* sequences (probe, Fig. 1), detected a 1.7-kb transcript (Fig. 3), which is close to the size predicted for the *araE* transcript (1.6 kb), indicating that *araE* is monocistronic. No signal was observed with RNA from cells grown under uninduced conditions, indicating that *araE* expression is regulated by L-arabinose as observed with the *araABDLMNPQ-abfA* metabolic operon (24).

Primer extension analysis of total RNA isolated from cells grown in the presence of L-arabinose (Fig. 4) showed that the 5′ end of the *araE* message corresponds to a G residue at position 1360 (Fig. 1). Centered at −10 and −35 bp upstream from this residue are two sequences, TATTCT and TTGACA, respectively (Fig. 1), that closely match the consensus sequences for recognition by *B. subtilis* σ^A-containing RNA polymerase (13, 14). Therefore, we assume that this position corresponds to the *araE* transcription start site. The *araE* and *araR* promoters are arranged divergently, and the distance between the transcriptional start sites of the two genes is 73 bp (Fig. 1). A shorter primer extension product corresponding to an apparent start site, a C residue at position 1363 (Fig. 1), is also detected. Analysis of the DNA sequence upstream of this position did not reveal canonical −10 and −35 sequences, suggesting that this signal may be the result of premature pausing of the reverse transcriptase or of processing of the longer product. No extension product was seen with RNA isolated from cells grown in the absence of L-arabinose under the same conditions, similar to the data obtained by Northern blot analysis.

Expression and regulation of *araE-lacZ* transcriptional fusions. To study the functionality of the *araE* promoter, we constructed a transcriptional *lacZ* fusion at the *araE* locus, resulting in strain IQB232 ($AraE^+$ [see Materials and Methods and Table 1]). β-Galactosidase activity from this fusion increased about 30-fold in cells grown for at least 2.5 generations with L-arabinose, and expression in L-arabinose-grown cells is repressed about 20-fold by glucose (Fig. 5). These data confirm that *araE* expression is induced by L-arabinose and indicate

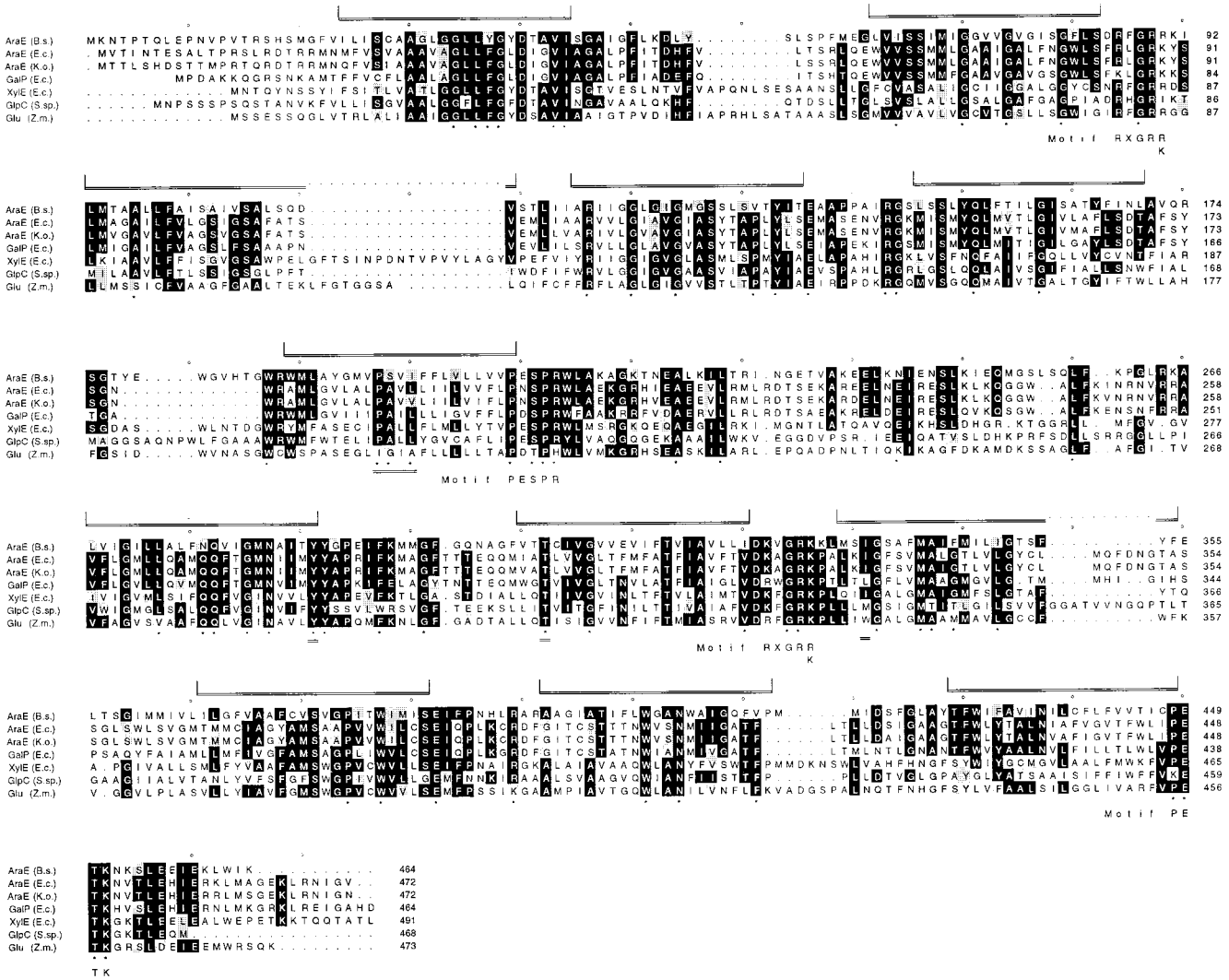

FIG. 2. Alignment of the predicted product of *araE* with proton-linked monosaccharide transport protein members of the group I membrane transport proteins. Alignments were determined with the PileUp program of the Genetics Computer Group package of sequence analysis software. Black boxes indicate identical residues in at least four of the seven sequences, highlighting in gray indicates conservative replacements, and numbers indicate positions in the amino acid sequences of the respective proteins. The duplicated RXGRR motif (9) and the PESPR and PETK motifs (4) are depicted below the sequence. Highly conserved residues in group I membrane transport proteins (4) are denoted by an asterisk, and the residues that may be implicated in sugar binding are marked indicated with a double line below the sequence. The 12 putative membrane-spanning segments of 21 residues are indicated with a double line above the AraE sequence. The sequences shown correspond to the following Swiss-Prot numbers: AraE from *E. coli* (E.c.), P09830; AraE from *K. oxytoca* (K.o.), P45598; GalP from *E. coli*, P37021; XylE from *E. coli*, P09098; GlcP from *Synechocystis* sp. (S.sp.), P15729; Glf from *Z. mobilis* (Z.m.), P21906.

that transcription of the *araE* gene is subjected to glucose repression. In *B. subtilis*, catabolite repression is mediated by CcpA, which binds alone or in conjunction with Hpr(Ser-P) to *cis*-acting catabolite responsive elements (CRE) located in a range of 200 bp upstream or downstream of the transcriptional start sites of genes which are subjected to carbon catabolite repression (references 6 and 21 and references therein). A potential *cis*-acting CRE is located 66 bp downstream of the *araE* transcription start site (Fig. 1), with two deviations (at position 13, A instead of C and at position 14, C instead of A) from the proposed consensus sequence (TGWNANCGNT NWCA, where W is A or T [31]). Interestingly, expression from the *araABDLMNPQ-abfA* operon promoter is also repressed by glucose, and a potential CRE was identified in a location very similar to that of the *araE* promoter (60 bp downstream of the operon transcription start site [24]). Although these putative CREs do not have an optimized consensus sequence for interaction with CcpA, displaying deviation at position 13, which was shown by point mutations to cause significant decrease of carbon catabolite repression of other catabolic genes (11, 31, 32), they could well confer catabolite repression by glucose.

The possibility that the *araE* promoter was regulated by AraR was examined by the construction of a transcriptional *lacZ* fusion at the *araE* locus in an AraR$^-$ background, leading to strain IQB233 (AraE$^+$ AraR$^-$ [see Materials and Methods and Table 1]). In the absence of L-arabinose the levels of accumulated β-galactosidase activity detected in the AraR$^-$ background, strain IQB233, were approximately fivefold higher than in the wild-type strain, IQB232, grown for at least 2.5 generations with L-arabinose (Fig. 5). The addition of L-arabinose to the IQB233 culture caused severe inhibition of growth

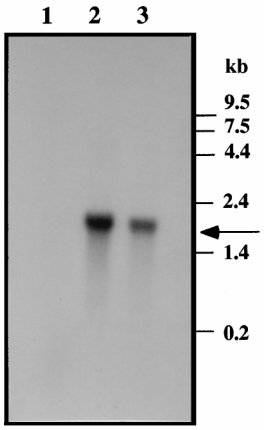

FIG. 3. Northern blot analysis of the *araE*-specific transcript. Lane 1, 7.5 μg of total RNA extracted from uninduced wild-type strain *B. subtilis* 168T$^+$; lanes 2 and 3, 7.5 and 3.5 μg, respectively, of total RNA extracted from the induced wild-type strain *B. subtilis* 168T$^+$ (see Materials and Methods). The RNA ladder (9.5 to 0.24 kb) (GIBCO BRL) was used as molecular size marker. 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 *Hin*dIII fragment (position 1154 to 2082 [probe in Fig. 1]). The RNA ladder was probed with ^{32}P-labeled λ DNA and visualized by staining with ethidium bromide. The *araE* transcript of about 1.7 kb is indicated by an arrow.

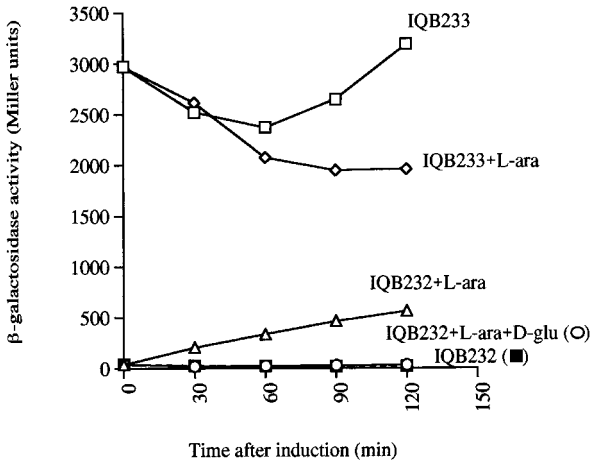

FIG. 5. Expression of the *araE* gene promoter in the wild-type and *araR*-defective backgrounds. The *B. subtilis* strains harboring transcriptional *araE-lacZ* fusions were grown on C minimal medium supplemented with casein hydrolysate 1% (wt/vol) in the absence or presence of L-arabinose (0.4%) or L-arabinose and glucose (0.4% [wt/vol]). Time is expressed in minutes after induction, and for each strain the results represent the average, in Miller units, from two independent experiments. The strains of *B. subtilis* are IQB232 *araE*::pSN31(*araE'-lacZ cat*) (wild type) and IQB233 Δ*araR*::*km araE*::pSN31(*araE'-lacZ cat*) (Ara$^-$).

(data not shown), as previously reported in other *araR* null mutant strains (23). The finding that deficiency of AraR induces a fivefold increase in the expression from the *araE* promoter supports the hypothesis 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, which are toxic to the cell (23).

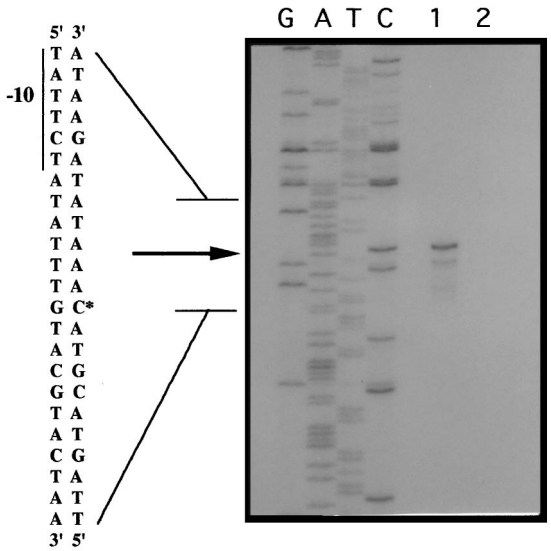

FIG. 4. Mapping of the transcriptional start site of the *araE* gene. A radiolabeled oligonucleotide primer (5′-GGACATTGTTCTTCGGTAAGTTACCC-3′) complementary to the 5′ end of the nontranscribed strand of the *araE* cistron was hybridized and used to direct cDNA synthesis from total *B. subtilis* 168T$^+$ 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 with the same primer and a single-stranded M13 DNA template which includes part of the *araE* gene (positions 1 to 2127, Fig. 1). The arrow indicates the position of the *araE*-specific primer extension product, and the asterisk indicates the deduced start site of transcription, a G residue at position 1360 in the *araR*/*araE* intergenic sequence shown in Fig. 1.

These results corroborate the proposed model for the mode of action of AraR (23); in the absence of L-arabinose, AraR binds to site(s) within the *araE*/*araR* promoter, preventing transcription from the *araE* promoter and simultaneously limiting the frequency of initiation from its own promoter in order to maintain a low but constant intracellular concentration of its own transcript and product. The addition of L-arabinose will allow transcription from the *araE* promoter and increase the frequency of initiation from the *araR* promoter. This model predicts that there is a site(s) for AraR binding within the *araE*/*araR* promoter region and that candidates for such a site(s) are two regions of dyad symmetry (Fig. 1) which display sequence similarity to the two inverted repeats localized in the *araABDLMNPQ-abfA* operon promoter region (24). Preliminary DNA protection studies with purified AraR-MBP fusion protein indicate that these regions are indeed the binding sites for AraR (15).

Functional analysis of the *araE* gene. To assess the function of the *araE* gene, strain IQB230, a Campbell integrant of pSR1 which disrupts the *araE* transcriptional unit and establishes a transcriptional *lacZ* fusion (see Materials and Methods and Table 1) was plated on minimal C medium containing L-arabinose (0.1% [wt/vol]) as the sole carbon source. The *araE* null mutant showed an Ara$^-$ phenotype in these conditions. However, increasing the amount of L-arabinose in the medium up to 4% (wt/vol) led to an Ara$^+$ phenotype. The effect of increasing the external concentration of L-arabinose in the expression of the *araE* promoter was examined by monitoring β-galactosidase activity from *araE-lacZ* fusions in both wild-type and defective *araE* backgrounds (Table 2). Different concentrations of L-arabinose did not affect the levels of accumulated β-galactosidase activity in the wild-type strain (IQB232 [Table 1]), but a direct correlation between the external concentration of L-arabinose and expression from the *araE* promoter was observed in the *araE* null mutant (IQB230). Furthermore, the maximal level of β-galactosidase activity in the wild-type strain was 4.7-fold higher than in the *araE* null mu-

TABLE 2. Analysis of an *araE-lacZ* transcriptional fusion in both wild-type and defective *araE* backgrounds[a]

Strain	Genotype and relevant phenotype	β-Galactosidase activity (Miller units) with:			
		No sugar added	0.4% (wt/vol) L-arabinose	1% (wt/vol) L-arabinose	4% (wt/vol) L-arabinose
IQB232	*araE*::pSN31(*araE′-lacZ cat*); $AraE^+$	19.6 ± 4.4	549.8 ± 21.5	562.3 ± 15.8	457.9 ± 3.3
IQB230	*araE′*::pSR1(*araE′-lacZ cat*); $AraE^-$	8.2 ± 0.5	34.9 ± 0.1	59.3 ± 1.2	119.3 ± 1.8

[a] Cultures were grown on C minimal medium supplemented with casein hydrolysate 1% (wt/vol) in the absence or presence of L-arabinose at the concentrations indicated. The levels of accumulated β-galactosidase activity were measured 2 h after induction, which corresponds to growth for at least 2.5 generations in the presence of the inducer (see Materials and Methods).

tant grown in the presence of 4% (wt/vol) L-arabinose, suggesting that the internal concentration of L-arabinose in the *araE* null mutant is lower than the minimal concentration necessary to achieve full induction of the *araE* promoter.

The Ara^- conditional phenotype of the *araE* null mutant, which is typical of a transport mutant when the cell has alternative transport systems for the same substrate, together with the similarity observed between the primary structure of the *araE* predicted product and several prokaryotic proton-linked sugar transporters, indicates that the *araE* gene encodes a permease involved in L-arabinose transport into the cell. The transport of L-arabinose across the *E. coli* cytoplasmic membrane requires the expression of either the low-affinity high-capacity arabinose transport gene, *araE*, a proton symporter, or the high-affinity but low-capacity arabinose transport operon, *araFGH*, a binding-protein-dependent system (reference 26 and references therein). Since these systems function in parallel, only if both are inactivated will the cells show an Ara^- phenotype. In *B. subtilis* the putative products of the *araN*, *araP*, and *araQ* genes, belonging to the *araABDLMNPQ-abfA* operon, showed weak similarity with AraF, AraG, and AraH from *E. coli* but are homologous to bacterial components of binding-protein-dependent transport systems involved in the high-affinity transport of malto-oligosaccharides and multiple sugars (24). 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 the utilization of L-arabinose. However, this deletion caused a 1.8-fold increase in the doubling time of the mutant strain (24). Whether these genes are involved in L-arabinose and/or L-arabinose oligomer transport is unknown. In order to determine if these two systems are the only ones involved in L-arabinose transport, we constructed a double *araE* Δ*araL-abfA* mutant, strain IQB234 (Table 1). This strain showed an Ara phenotype similar to that seen with the *araE* null mutant, IQB230; however, the growth on minimal medium plates with L-arabinose as the sole carbon source was slightly slower than that observed with strain IQB230. Assuming that free sugar cannot traverse the permeability barrier, the Ara^- conditional phenotype displayed by the double *araE* Δ*araL-abfA* mutant indicates that *B. subtilis* should have another transporter for L-arabinose, and it is anticipated that such a system will have a low affinity and/or low capacity for L-arabinose.

ACKNOWLEDGMENTS

We thank H. de Lencastre, in whose laboratory this work was done, for her constant interest in this work, J. Almeida for critically reading the manuscript, and L. J. Mota for participating in an early phase of this project.

This work was supported in part by grant no. Praxis/PCNA/P/Bio/33/96, from Junta Nacional de Investigação Científica e Tecnológica (JNICT).

REFERENCES

1. **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.
2. **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.
3. **Griffith, J. K., M. E. Baker, D. A. Rouch, M. G. P. Page, R. A. Skurray, I. T. Paulsen, K. F. Chater, S. A. Baldwin, and P. J. F. Henderson.** 1992. Membrane transport proteins: implications of sequence comparisons. Curr. Opin. Cell Biol. **4:**684–695.
4. **Henderson, P. J. F., S. A. Baldwin, M. T. Cairns, B. M. Charalambous, H. C. Dent, F. Gunn, W.-J. Liang, V. A. Lucas, G. E. Martin, T. P. McDonald, B. J. McKeown, J. A. R. Muiry, K. R. Petro, P. E. Roberts, K. P. Shatwell, G. Smith, and G. T. Tate.** 1992. Sugar-cation symport systems in bacteria. Int. Rev. Cytol. **137A:**149–208.
5. **Henderson, P. J. F., R. A. Giddens, and M. C. Jones-Mortimer.** 1977. Transport of galactose, glucose and their molecular analogues by *Escherichia coli* K12. Biochem. J. **162:**309–320.
6. **Hueck, C., and W. Hillen.** 1995. Catabolite repression in *Bacillus subtilis*: a global regulatory mechanism for the Gram-positive bacteria? Mol. Microbiol. **15:**395–401.
7. **Kyte, J., and R. F. Doolittle.** 1982. A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. **157:**105–132.
8. **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.
9. **Maiden, M. C. J., E. O. Davis, S. A. Baldwin, D. C. M. Moore, and P. J. F. Henderson.** 1987. Mammalian and bacterial sugar transport proteins are homologous. Nature (London) **325:**641–643.
10. **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.
11. **Martin-Verstraete, I., J. Stülke, A. Klier, and G. Rapoport.** 1995. Two different mechanisms mediate catabolite repression of the *Bacillus subtilis* levanase operon. J. Bacteriol. **177:**6919–6927.
12. **Miller, J. H.** 1972. Experiments in molecular genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
13. **Moran, C. P., Jr.** 1993. RNA polymerase and transcription factors, p. 653–667. *In* A. L. Sonenshein, 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.
14. **Moran, C. P., Jr., N. Lang, S. F. J. LeGrice, G. Lee, M. Stephens, A. L. Sonenshein, J. Pero, and R. Losick.** 1982. Nucleotide sequences that signal the initiation of transcription in *Bacillus subtilis*. Mol. Gen. Genet. **186:**339–346.
15. **Mota, L. J., P. Tavares, and I. Sá-Nogueira.** Unpublished results.
16. **Pascal, M., F. Kunst, J. A. Lepesant, and R. Dedonder.** 1971. Characterization of two sucrase activities in *Bacillus subtilis* Marburg. Biochimie **53:**1059–1066.
17. **Paveia, H., and L. Archer.** 1992. Genes for L-Arabinose utilization in *Bacillus subtilis*. Broteria Genet. Lisboa **13:**149–159.
18. **Paveia, H., and L. Archer.** 1992. Mapping of *ara* genes in *Bacillus subtilis*. Broteria Genet. Lisboa **13:**161–167.
19. **Perego, M.** 1993. Integrational vectors for genetic manipulation in *Bacillus subtilis*, p. 615–624. *In* A. L. Sonenshein, 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.
20. **Saier, M. H., S. Chauvaux, G. M. Cook, J. Deutscher, I. T. Paulsen, J. Reizer, and J.-J. Ye.** 1996. Catabolite repression and inducer control in Gram-positive bacteria. Microbiology **142:**217–230.
21. **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.
22. **Sá-Nogueira, I., and H. de 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.
23. **Sá-Nogueira, I., and L. J. Mota.** 1997. Negative regulation of L-Arabinose metabolism in *Bacillus subtilis*: characterization of the *araR* (*araC*) gene. J. Bacteriol. **179:**1598–1608.
24. **Sá-Nogueira, I., T. V. Nogueira, S. Soares, and H. Lencastre.** 1996. The L-arabinose (*ara*) operon of *Bacillus subtilis*: nucleotide sequence, genetic organization and expression. Microbiology. **143:**957–969.
25. **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.
26. **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. B. 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.
27. **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.
28. **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.
29. **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.
30. **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. Nat. New Biol. **246:**40–41.
31. **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.
32. **Wray, L. W., Jr., F. K. Pettengil, and S. H. Fisher.** 1994. Catabolite repression of the *Bacillus subtilis hut* operon requires a *cis*-acting site located downstream of the transcription initiation site. J. Bacteriol. **176:**1894–1902.
33. **Yamomoto, H., S. Uchiyama, A. N. Fajar, N. Ogasawara, and J. Sekiguchi.** 1996. Determination of a 12 kb nucleotide sequence around the 76° of the *Bacillus subtilis* chromosome. Microbiology **142:**1417–1421.
34. **Yoshida, K., S. Seki, M. Fujimura, Y. Miwa, and Y. Fujita.** 1995. Cloning and sequencing of a 36-kb region of the *Bacillus subtilis* genome between the *gnt* and *iol* operons. DNA Res. **2:**61–69.
35. **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.