Microbiology (1997), **143**, 957–969 Printed in Great Britain

The *Bacillus subtilis* L-arabinose (*ara*) operon: nucleotide sequence, genetic organization and expression

Isabel Sá-Nogueira,[1] Teresa V. Nogueira,[1]† Sónia Soares[1]‡
and Hermínia de Lencastre[1,2]

Author for correspondence: Isabel Sá-Nogueira. Tel: +351 1 4426171. Fax: +351 1 4428766.
e-mail: sanoguei@itqb.unl.pt

[1] Instituto de Tecnologia
Química e Biológica,
Universidade Nova de
Lisboa. Apartado 127, 2780
Oeiras Codex, Portugal

[2] The Rockefeller University,
Laboratory of
Microbiology, 1230 York
Avenue, New York, NY
10021-6399, USA

The *Bacillus subtilis* L-arabinose metabolic genes *araA*, *araB* and *araD*, encoding L-arabinose isomerase, L-ribulokinase and L-ribulose-5-phosphate 4-epimerase, respectively, have been cloned previously and the products of *araB* and *araD* were shown to be functionally homologous to their *Escherichia coli* counterparts by complementation experiments. Here we report that *araA*, *araB* and *araD*, whose inactivation leads to an Ara phenotype, are the first three ORFs of a nine cistron transcriptional unit with a total length of 11 kb. This operon, called *ara*, is located at about 256° on the *B. subtilis* genetic map and contains six new genes named *araL*, *araM*, *araN*, *araP*, *araQ* and *abfA*. Expression of the *ara* operon is directed by a strong σ^A-like promoter identified within a 150 bp DNA fragment upstream from the translation start site of *araA*. 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 *abfA* most probably encodes an α-L-arabinofuranosidase. The functions of *araL* and *araM* are unknown. An *in vitro*-constructed insertion-deletion mutation in the region downstream from *araD* allowed us to demonstrate that *araL*, *araM*, *araN*, *araP*, *araQ* and *abfA* are not essential for L-arabinose utilization. Studies with strains bearing transcriptional fusions of the operon to the *E. coli lacZ* gene revealed that expression from the *ara* promoter is induced by L-arabinose and repressed by glucose.

Keywords: *Bacillus subtilis*, L-arabinose (*ara*) operon, expression, catabolite repression

INTRODUCTION

Bacillus subtilis, an endospore-forming Gram-positive bacterium, is able to grow on L-arabinose as sole carbon source. L-Arabinose residues are found widely distributed among many heteropolysaccharides of different plant tissues, such as arabinans, arabinogalactans, xylans and arabinoxylans. *Bacillus* species in their natural reservoir, the soil, participate in the early stages of plant material decomposition and *B. subtilis* secretes three enzymes, an endo-arabanase and two arabinosidases, capable of releasing arabinosyl oligomers and L-arabinose from plant cell walls (Kaji & Saheki, 1975; Weinstein & Albersheim, 1979). The pathway of L-arabinose utilization in *B. subtilis* has been described by Lepesant & Dedonder (1967a). 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. Mutants unable to use L-arabinose as sole carbon source, deficient in one of the three enzymes involved in L-arabinose catabolism, have been characterized, as well as constitutive mutants for

---

† **Present address:** Institut de Biologie Physico-Chimique, 13 Rue Pierre et Marie Curie, 75005 Paris, France.

‡ **Present address:** Public Health Research Institute, 455 First Avenue, New York, NY 10016, USA.

Abbreviations: Cm, chloramphenicol; Em, erythromycin; Km, kanamycin; Sp, spectinomycin.

The accession numbers for the nucleotide sequences reported in this paper are X89408 (*araA*, *B* and *D*) and X89810 (*araL*, *M*, *N*, *P*, *Q* and *abfA*).

all three enzymes (Lepesant & Dedonder, 1967a, b). The synthesis of these enzymes was shown to be inducible by L-arabinose and the isomerase activity is subjected to catabolite repression by glucose and glycerol (Lepesant & Dedonder, 1967a).

A collection of Ara$^-$ *B. subtilis* mutants was isolated, biochemically characterized and the three metabolic genes, *araA*, *araB* and *araD*, encoding L-arabinose isomerase, L-ribulokinase and L-ribulose-5-phosphate 4-epimerase, respectively, were identified and mapped between *aroG* and *leuA*, at about 256° on the *B. subtilis* genetic map (Paveia & Archer, 1992a, b). Two additional classes of mutations affecting L-arabinose utilization were identified; one included mutations conferring an Ara$^-$ phenotype to strains bearing the *araA*, *araB* and *araD* wild-type alleles (Paveia & Archer, 1992a, b), and another comprised mutants showing constitutive expression of the three genes (Sá-Nogueira *et al.*, 1988). These mutations were mapped between the *cysB* and *hisA* markers, at about 294° on the *B. subtilis* genetic map, and define another *ara* locus named *araC*. Expression of L-arabinose isomerase is severely repressed during growth in media containing L-arabinose plus glucose. Since L-arabinose isomerase expression is still regulated by catabolite repression in strains which contain constitutive mutations (*araC*^{c}), L-arabinose transport does not play a major role in catabolite repression of expression of the metabolic enzymes (Sá-Nogueira *et al.*, 1988). The products of the previously cloned genes *araA*, *araB* and *araD* were shown in complementation experiments to be functionally homologous to their *Escherichia coli* counterparts. Transformation experiments involving defined restriction fragments from the cloned genes showed that they are adjacent and probably constitute an operon with the order *araABD* (Sá-Nogueira & Lencastre, 1989), unlike the *araBAD* order found in the *E. coli* operon (Englesberg *et al.*, 1969).

In this communication we report the cloning of an additional 7·1 kb chromosomal fragment, located downstream from *araD* and the nucleotide sequence of over 11 kb. This region contains a cluster of nine genes: the metabolic genes *araA*, *araB* and *araD*, and six new genes named *araL*, *araM*, *araN*, *araP*, *araQ* and *abfA*. We have demonstrated that all genes comprise a single transcriptional unit, called the *ara* operon, whose expression is directed by a single σ^A-type promoter identified within a 150 bp DNA fragment upstream from the translation start site of *araA*. The *araN*, *araP* and *araQ* gene products are likely components of a binding-protein-dependent transport system and *abfA* most probably encodes an α-L-arabinofuranosidase. In this study we define the promoter region of the *ara* operon and examine its expression and regulation using transcriptional fusions of this operon to the *E. coli lacZ* gene. These results indicate that the *ara* operon is regulated at the transcriptional level because expression from the *ara* promoter is induced by L-arabinose and repressed by glucose.

METHODS

Bacterial strains and growth conditions. The *B. subtilis* strains used in this study are listed on Table 1. *E. coli* DH5α (Gibco/BRL) was used as a host for all plasmids and *E. coli* DH5α F′ (BRL) for the propagation and amplification of recombinant M13 bacteriophages. *E. coli* strains were grown on LB (Luria–Bertani medium; Miller, 1972). Ampicillin (Ap, 75 μg ml^{-1}), chloramphenicol (Cm, 15 μg ml^{-1}), X-gal (40 μg ml^{-1}) or IPTG (1 mM) were added as appropriate. *B. subtilis* strains were grown on LB, SP medium (Martin *et al.*, 1987) or minimal C medium (Pascal *et al.*, 1971). Cm (5 μg ml^{-1}), erythromycin (Em, 1 μg ml^{-1}), kanamycin (Km, 25 μg ml^{-1}) or spectinomycin (Sp, 50 μg ml^{-1}) were added as appropriate. Solid medium was made with LB, SP or minimal C medium containing 1·5 % (w/v) Bacto Agar (Difco). To test for growth of *B. subtilis* integrant strains on L-arabinose as sole carbon source, strains were plated on minimal C medium containing 0·1 % (w/v) L-arabinose. The AraB$^-$ phenotype was determined on minimal C medium plates supplemented with 1 % (w/v) casein hydrolysate, 0·1 % L-arabinose and 1 % (w/v) ribitol. To determine specific growth rates, the *B. subtilis* strains were grown in liquid C medium with 0·4 % L-arabinose as sole carbon source. The cultures were incubated with aeration by shaking (130 r.p.m.) and cell growth was monitored by OD_{600}. For β-galactosidase assays and RNA preparation the *B. subtilis* strains were grown in liquid C medium supplemented with 1 % (w/v) casein hydrolysate, and L-arabinose and glucose were added to the cultures when necessary at a final concentration of 0·4 % (w/v).

DNA manipulations and sequencing. DNA manipulations were carried out according to Sambrook *et al.* (1989). Enzymes were purchased from commercial suppliers and used according to the manufacturers' instructions. DNA sequencing was performed by the method of Sanger *et al.* (1977) with the Sequenase Kit (T7 DNA polymerase; USB). Sequencing templates were prepared by a combination of subcloning appropriate fragments from pSNL1 and pSNL9 into the polycloning site of M13mp19 or M13mp18 (Yanisch-Perron *et al.*, 1985) and sequential deletion of the recombinant M13 derivatives, by the method of Dale *et al.* (1985), using the Cyclone Biosystem Kit (International Biotechnologies Inc.). The DNA sequence was determined on both strands and across all the restriction sites used for subcloning. The primer 5′ CCTCTTCGCTATTACGCC 3′, complementary to the coding sequence of *lacZ*, was used to sequence the transcriptional *lacZ* fusions.

Plasmid constructions. pSNL7 was constructed by subcloning a 959 bp *Sma*I–*Pst*I DNA fragment (nt 938–1897, Fig. 1) from pSNL1 (Sá-Nogueira & Lencastre, 1989) between the *Sma*I and *Pst*I sites of the integrational vector pJM783 (Perego, 1983). To construct pSS2, we digested pSNL1 (Sá-Nogueira & Lencastre, 1989) with *Hin*dIII and *Xho*I and cloned a purified fragment of 965 bp (nt 3815–4780, Fig. 1) between the *Hin*dIII and *Sal*I sites of the integrating vector pJH101 (Ferrari *et al.*, 1983). pTN10 was obtained by subcloning a 789 bp *Hin*dIII–*Hinc*II DNA fragment (nt 6545–7334, Fig. 1) from pSS3 between the *Hin*dIII and *Eco*RV sites of the integrational vector pJH101 (Ferrari *et al.*, 1983). pTN14 was constructed by subcloning the 678 bp *Sma*I–*Bgl*II DNA fragment (nt 8242–8920, Fig. 1) from pTN13 between the *Bam*HI and *Sst*I (fill-in) sites of pJM783 (Perego, 1993). pSNL10 was obtained by subcloning a 1·7 kb *Eco*RI–*Hinc*II fragment (nt 2681–4416, Fig. 1) from pSNL1 (Sá-Nogueira & Lencastre, 1989) between the *Eco*RI and *Sma*I sites of pMK4 (Sullivan *et al.*, 1984).

Table 1. *B. subtilis* strains

Strain*	Genotype	Phenotype	Source†
168T^{+}	Prototroph	Ara^{+}	F. E. Young
BR151	*metB10 lys3 trpC2*	Ara^{+}	F. E. Young
IQB100	*araB′*::pSNL7(*araB–cat lacZ*)	Cm^{r} Ara^{-}	pSNL7 → 168T^{+}
IQB101	*araB′*::*lacZ erm*	LacZ^{+} Em^{r} Ara^{-}	pSNL11‡ → 168T^{+}
IQB102	*araB′*::*erm lacZ*	LacZ^{-} Em^{r} Ara^{-}	pSNL12‡ → 168T^{+}
IQB103	*araA′*::pSNL13 (*araA–lacZ cat*)	LacZ^{+} Cm^{r} Ara^{-}	pSNL13 → 168T^{+}
IQB104	*araA′*::pSNL14 (*araA–cat lacZ*)	LacZ^{-} Cm^{r} Ara^{+}	pSNL14 → 168T^{+}
IQB202	*araL′*::pSS2 (*ara*L*–amp cat*)	Cm^{r} Ara^{+}	pSS2 → 168T^{+}
IQB204	*araN′*::pTN10 (*araN–cat amp*)	Cm^{r} Ara^{+}	pTN10 → 168T^{+}
IQB205	*araQ′*::pTN14 (*araQ–lacZ ′cat*)	LacZ^{-} Cm^{r} Ara^{+}	pTN14 → 168T^{+}
IQB206	Δ*ara*L*–abfA*::*spc*	Sp^{r} Ara^{+}	pSN22‡ → 168T^{+}

* All strains are derivatives of *B. subtilis* 168T^{+}.

† The arrows indicate transformation and point from donor DNA to recipient strain. F. E. Young, University of Rochester New York, USA.

‡ Transformation was carried out with linearized plasmid DNA.

pSNL11 and pSNL12 were obtained as follows. A 4·5 kb *Bam*HI–*Hin*dIII (fill-in) fragment extracted from pMC11 (Debarbouillé *et al.*, 1990), containing *lacZ* and *erm* from pTV32 (Perkins & Youngman, 1986), was subcloned in both orientations at the unique *Eco*RV restriction site (nt 3214, Fig. 1) of pSNL10. pSNL13 and pSNL14 were obtained by subcloning a 470 bp *Dra*I–*Eco*RV DNA fragment (nt 82–552, Fig. 1) from pSNL9 at the unique *Sma*I site of the integrational vector pJM783 (Perego, 1993) in both orientations. pSNL13 contains *lacZ* in the same orientation as the *araA* region sequences and pSNL14 contains *lacZ* in the opposite orientation. pSN20 was constructed by cloning the 1·2 kb *Eco*RV–*Hin*cII fragment (nt 3214–4416, Fig. 1) from pSS3 into the *Sma*I site of pAH248 [a pGem-7Zf(+) (Promega) derivative that contains a Km^{r} gene cloned between its *Xho*I and *Eco*RI sites (A. O. Henriques & C. P. Moran Jr, Emory University School of Medicine, Atlanta, GA, USA, personal communication)]. To obtain pSN21 a 1·7 kb *Eco*RV fragment from pSN5 (nt 10632–about 12332, Fig. 1) was inserted into the *Hin*cII site of pAH250 [a pBluescript SK+ (Stratagene) derivative that contains a Sp^{r} gene (*spc*) cloned into the *Eco*RV site (A. O. Henriques, B. W. Beall & C. P. Moran Jr, personal communication)]. To construct pSN22, we digested pSN20 with *Pst*I and *Nsi*I and cloned a purified fragment of about 2790 bp, which contains the Km^{r} gene, in the *Sma*I site of pSN21. pSNL9, pSS3, pTN13 and pSN5 were obtained by cutting chromosomal DNA from *B. subtilis* strains IQB100, IQB202, IQB204 and IQB205 (Table 1) with *Hin*dIII, *Eco*RI, *Nco*I and *Sma*I, respectively, followed by circularization of the DNA fragments at low concentration.

Bacterial transformation. *B. subtilis* DNA transformations were performed according to the method of Anagnostopoulos & Spizizen (1961). *E. coli* transformations were carried out according to standard methods (Sambrook *et al.*, 1989).

***β*-Galactosidase assays.** Strains of *B. subtilis* harbouring transcriptional *lacZ* fusions were grown in 75 ml C medium supplemented with 1% casein hydrolysate. During early exponential phase (OD$_{600}$ = 0·11–0·15) 25 ml of the culture was transferred to two different flasks and L-arabinose at a final concentration of 0·4% or both L-arabinose and glucose each at a final concentration of 0·4% were added. At this time, t_0, 100 µl aliquots of cell culture were collected, harvested and stored at −70 °C overnight. Exponential growth of the three cultures was followed by measuring OD$_{600}$ and at 30 min intervals, 100 µl of cell culture samples was removed and stored at −70 °C until the cultures reached an OD$_{600}$ = 0·7–0·8, which corresponds to growth for at least 2·5 generations in the presence of the inducer. The cells were resuspended in 1 ml Z buffer (Miller, 1972) and two drops of chloroform plus one drop of 0·1% SDS were added and mixed vigorously for 10 s on a table top vortex apparatus. *β*-Galactosidase activity was determined as described by Miller (1972) using the substrate ONPG.

RNA preparation, Northern blotting and primer extension analysis. *B. subtilis* 151 or 168T^{+} cells were grown in C medium supplemented with 1% casein hydrolysate in the presence and absence of L-arabinose at a final concentration of 0·4%. Cells were harvested during late exponential phase (OD$_{600}$ ∼ 0·9) and RNA prepared as described by Igo & Losick (1986). For Northern blot analysis, 2·5–10 µg total RNA was run in 1·0–1·2% (w/v) agarose/formaldehyde and transferred to positively charged nylon membranes (Hybond-N+, Amersham) according to standard methods (Sambrook *et al.*, 1989). Size determination was done using an RNA ladder (0·24–9·5 kb; Gibco/BRL). The probes were labelled using the Multiprime random-prime DNA labelling system from Amersham and [α-^{32}P]dATP [3000 Ci mmol^{-1} (111 TBq mmol^{-1})]. Primer extension analysis was performed as described by Sambrook *et al.* (1989). The two synthetic oligonucleotides used in primer extension experiments were primer A (5′ GAAGCATGTAAACTGCCCC 3′), complementary to nt 216–234 (Fig. 1), and primer B (5′ CCAGCG-TCTCTTCCCCG 3′), complementary to nt 283–300 (Fig. 1). The two oligonucleotides were used in separate experiments to rule out the possibility of primer-specific artifacts. A total of 10 ng of primer was used in the labelling reaction mixed with 25 µg RNA, denatured by heating to 85 °C for 10 min and annealed by incubation at 42 °C for 3 h. The oligonucleotide primer was extended using 15 units of avian myeloblastosis virus reverse transcriptase for 2 h at 37 °C, as described by

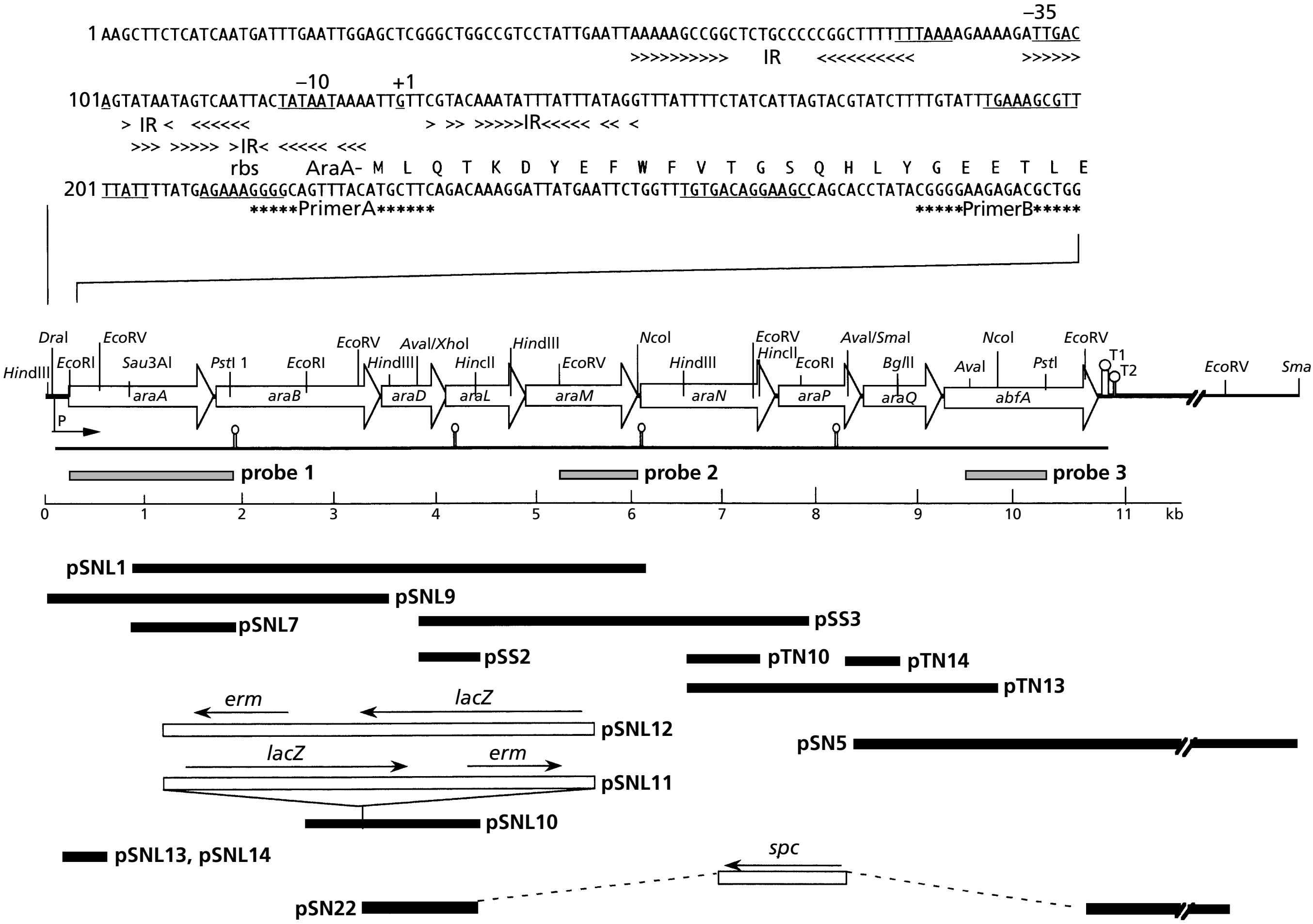


Fig. 1. Physical and genetic map of the *ara* region of the chromosome. The location and direction of transcription of the nine ORFs (*araA*, *B*, *D*, *L*, *M*, *N*, *P*, *Q* and *abfA*), predicted from the analysis of the nucleotide sequence, are indicated by arrows. The promoter (P) of the *ara* operon, defined by primer extension, is located upstream from *araA* and the two regions of dyad symmetry (T1 and T2) that could represent the terminators of the *ara* transcriptional unit are located downstream from *abfA*. Relevant restriction sites are given in the partial restriction map. The region to the right of the *Eco*RI site (position 11755) is not drawn to scale. Immediately below the physical map the *ara* operon transcript is schematically shown and putative secondary structures of the mRNA are indicated by stem–loop structures. The grey boxes, below the physical map, represent the three fragments used as probes for Northern analysis of the *ara* transcripts and the black boxes represent the extent of the inserts in the indicated plasmids. The sites of different insertion-deletion mutations resulting from replacement of wild-type sequences, by double cross-over events (confirmed by Southern blot analysis, data not shown), with *in vitro*-engineered fragments of the *ara* region, present in plasmids pSNL11, pSNL12 and pSN22, are also shown. Plasmids pSNL7, pSS2, pTN10, pTN14, pSNL13 and pSNL14 were integrated into the host chromosome by means of a single cross-over (Campbell-type) recombinational event that occurred in the region of homology (confirmed by Southern blot analysis, data not shown). The *ara* operon promoter nucleotide sequence of the non-transcribed strand is shown in the 5′–3′ direction above the physical map. The predicted N-terminal region of the polypeptide encoded by *araA* is given in single letter code. The transcription start site (+1), defined by primer extension analysis, the −35 and −10 regions of the promoter and the putative ribosome binding site (rbs) are underlined. Convergent arrows represent different regions of dyad symmetry (IR) and the complementary sequence of the two primers A and B, used in primer extension analysis are represented below the sequence. The two putative catabolic-repression-associated sequences (positions 191–204 and 260–273) are underlined.

Sambrook *et al.* (1989). Analysis of the extended products was carried out on 7·5 % polyacrylamide urea gels.

Computer analysis. Amino acid sequences were deduced from the nucleotide sequence using DNASIS V2.0 (Hitachi Software Engineering, 1991). The GenBank and EMBL databases were accessed using the GCG package of sequence analysis software (Genetics Computer Group, Madison, Wisconsin, USA).

RESULTS

Insertional inactivation of *araB* and cloning of an intact copy of *araA*

The location of the *araA* locus at one end of the cloned fragment in pSNL1 (Fig. 1), together with the absence of *araA* complementation with pSNL1, suggested that only

Table 2. Percentage amino acid identity between the predicted sequences of the Ara proteins and similar proteins

B. subtilis AraA protein	Homologue (species/accession no.)*	Function	Identity (%)	Amino acid overlap
AraA	AraA (*E. coli*/M15263)	L-Arabinose isomerase	52·9	495
	AraA (*Sal. typhimurium*/M11047)	L-Arabinose isomerase	52·9	495
AraB	AraB (*E. coli*/M15263)	L-Ribulokinase	25·7	552
	AraB (*Sal. typhimurium*/M11045)	L-Ribulokinase	30·6	350
AraD	AraD (*E. coli*/M15263)	L-Ribulose-5-P 4-epimerase	57·1	231
	AraD (*Sal. typhimurium*/M11046)	L-Ribulose-5-P 4-epimerase	58·0	205
AraL	NagD (*E. coli*/X14135)	Unknown	25·5	251
AraN	LacE (*Agrobacterium radiobacter*/X66596)	Lactose-binding protein	26·2	302
	MalX (*Streptococcus pneumoniae*/L08611)	Maltose-binding protein	24·1	345
	AmyE (*Thermoanaerobacterium thermosulfurigens*/M57692)	Starch-binding protein	21·7	369
AraP	LacF (*Agrobacterium radiobacter*/X66596)	Membrane protein	29·6	284
	UgpA (*E. coli*/X13141)	Membrane protein	26·2	286
	AmyD (*Thermoanaerobacterium thermosulfurigens*/M57692)	Membrane protein	25·4	284
	MalC (*Streptococcus pneumoniae*/L08611)	Membrane protein	25·2	298
AraQ	LacG (*Agrobacterium radiobacter*/X66596)	Membrane protein	32·7	254
	UgpE (*E. coli*/X13141)	Membrane protein	22·9	279
	AmyC (*Thermoanaerobacterium thermosulfurigens*/M57692)	Membrane protein	28·2	262
	MalD (*Streptococcus pneumoniae*/L08611)	Membrane protein	25·6	262
AbfA	AbfA (*Streptomyces lividans*/U04630)	α-L-Arabinofuranosidase	52·6	500

part of *araA* was present in this plasmid (Sá-Nogueira & Lencastre, 1989). To clone the entire *araA* gene, plasmid pSNL7 (Fig. 1) was integrated, as single copy, into the *B. subtilis* 168T$^+$ chromosome at the *araA* and *araB* region of homology. This procedure causes disruption of the transcriptional unit and the structure of the resulting strain IQB100 that was unable to grow on minimal medium containing L-arabinose as sole carbon source, confirming the polar effect of the insertion on the genes located downstream from *araA*. Furthermore, strain IQB100 showed resistance to ribitol in the presence of L-arabinose on minimal medium plates supplemented with 1% casein hydrolysate. In *B. subtilis* (Paveia & Archer, 1992a), like in *E. coli* (Katz, 1970), these results indicate a defective *araB*. Chromosomal DNA from IQB100 was used to rescue the entire *araA* gene and its upstream region (see Methods). The structure of the recircularized plasmid, pSNL9, was analysed and it contains a 950 bp fragment of DNA upstream from the previously cloned DNA in plasmid pSNL7 (Fig. 1).

Cloning of the chromosomal region extending downstream from *araD*

To clone the region located downstream from *araD*, an integrational plasmid, pSS2, carrying sequences of *araD* and *araL* (Fig. 1), was transformed into the wild-type strain 168T$^+$. After integration as single copy, the resulting strain IQB202 presented an Ara$^+$ phenotype although the growth on minimal medium plates with L-arabinose as sole carbon source was slower than that observed with the wild-type strain 168T$^+$ (see Discussion below). The digestion of total chromosomal DNA from IQB202 followed by circularization of the fragments yielded plasmid pSS3 that includes a 3·0 kb insert located downstream to the fragment cloned in pSS2 (Fig. 1). To obtain a fragment that would contain the downstream region from *araN*, we performed a second chromosome walking step, using integrational plasmid pTN10 (Fig. 1). This procedure created plasmid pTN13 that carried an additional 3·2 kb of DNA adjacent to the previously cloned fragment in plasmid pTN10 (Fig. 1). Strain IQB204, which resulted from the integration of plasmid pTN10 (Fig. 1) into the chromosome of the wild-type strain 168T$^+$ showed a Ara$^+$ phenotype similar to that seen with IQB202. A third chromosome walking step rightwards from pTN13, using integrational plasmid pTN14 (Fig. 1), isolated a 4·7 kb *Sma*I fragment (plasmid pSN5). Plasmid pTN14, when integrated into the chromosome of strain IQB205 as single copy, caused an Ara$^+$ phenotype. The structure of the inserts in pSS3, pTN13 and pSN5 was compared to that of the corresponding areas 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 and deduced products of *ara* genes

Appropriate restriction fragments, selected on the basis of the physical maps of pSNL1, pSNL9, pSS3, pTN13 and pSN5, were subcloned into M13mp18 and M13mp19 and used as templates to determine the nucleotide sequence of the 11 kb DNA region shown in

Fig. 1. Sequence analysis revealed the presence of nine ORFs; the first three, by their position in the sequenced fragments of pSNL1 and pSNL9 and according to our previous results (Sá-Nogueira & Lencastre, 1989), were identified as *araA*, *araB* and *araD* (Fig. 1). *araA*, *araB* and *araD* could encode 496, 560 and 229 aa products of 56·2, 60·9 and 25·7 kDa, respectively. The six ORFs found downstream from *araD*, here named *araL*, *M*, *N*, *P*, *Q* and *abfA* (Fig. 1), of 269, 394, 433, 313, 281 and 499 codons, are capable of encoding putative products of 29, 43·1, 48·7, 35, 31·8 and 57 kDa, respectively. All ORFs are preceded by strong ribosome binding sites with the exception of *araL* which possesses a weak ribosome binding site. The intercistronic regions are very short and overlaps were observed between the *araD* and *araL* coding sequences, and between *araL* and *araM*, suggesting translational coupling. Two potential hairpin–loop structures, situated next to the UAA stop codon of *abfA* (T_1 and T_2, Fig. 1, with ΔG values of $-27{\cdot}4$ and $-18{\cdot}7$ kcal mol^{-1}, respectively, according to Tinoco *et al.*, 1973), probably correspond to transcription terminators. The absence of transcriptional signals among the nine coding regions suggested that they form a large operon transcribed from a promoter (described below) positioned 104 nt upstream from the *araA* start codon (Fig. 1).

Comparison of the primary structures of the products predicted to be encoded by the *ara* genes with GenBank sequences revealed significant similarities with other bacterial proteins of known function and the results are summarized in Table 2. The putative product of *araM*, a hydrophilic protein, did not show any significant similarity. The *araA*, *araB* and *araD* gene products exhibited a high level of identity to the L-arabinose isomerase, L-ribolukinase and L-ribulose-5-phosphate 4-epimerase, respectively, of *E. coli* and *Salmonella typhimurium*. The product of *araL*, a hydrophilic protein, displayed similarity to the *nagD* gene product of unknown function, which belongs to the *nag* regulon of *E. coli* involved in the metabolism of *N*-acetyl glucosamine (Plumbridge, 1989). The N-terminal region of the predicted sequence also shared 28·1 % and 29·2 % identity (over 121 and 106 aa, respectively, data not shown) with two 4-nitrophenylphosphatases, Pho2 and Pho13, from *Schizosaccharomyces pombe* (Yang *et al.*, 1991) and *Saccharomyces cerevisiae* (Kaneko *et al.*, 1989), respectively.

The predicted primary structure of *araN* showed similarity to known sugar-binding proteins that belong to the family of binding-protein-dependent transport systems (Table 2). Although the identity was not very high, there was significant sequence conservation within the N-terminal region of these proteins which display a signature sequence, according to Tam & Saier (1993). On the basis of this signature sequence (Fig. 2a) AraN can be included in the cluster 1 binding proteins (according to Tam & Saier, 1993), together with the above-mentioned proteins involved in the transport of malto-oligosaccharides and multiple sugars. The hydropathy profile of AraN indicated that it is mainly a hydrophilic protein; however its N-terminal region displayed characteristics of signal peptides of secretory precursor proteins: a positively charged N terminus, a hydrophobic core and a sequence, IAGCSA (starting at aa 19), which corresponds to the consensus sequence for the precursors of lipoproteins (reviewed in Hayashi & Wu, 1990).

(a)

```
Signature sequence  L X X L G K X F E X D X X G I K V X V  (68-81)
                    I     I A D   Y T   E     N V   I   L
                    V     V I Q     N   A     D Y       P
                    A     W V                   P

              AraN  Y V E M V K E W N K K Y P D R K I K L N T V V Y P Y  (75)
```

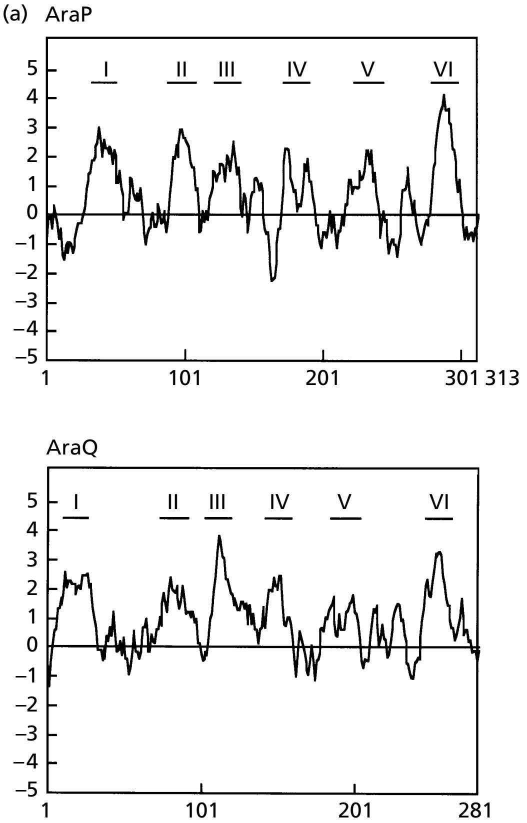
(b)

```
                   10        20        30        40
Abf–B. su  MKKARMIVDKEYKIGEVDKRIYGSFIEHMGRAVYEGIYEPDHPEADEDG
             ::: ::..:..::.:.:::::::::::.:::::.:::::.::.:::.:
Abf–B. st ATKKATMIIEKDFKIAEIDKRIYGSFIEHLGRAVYGGIYEPGHPQADENG
                  10        20        30        40        50
```

Fig. 2. (a) Alignment of a segment of the predicted sequence of the AraN protein with the signature sequence of cluster 1 binding proteins, from binding-protein-dependent transport systems, according to Tam & Saier (1993). Numbers in parentheses indicate the positions of the last amino acid residues. The highly conserved lysine residue (K) is in bold and the amino acid residues that match the signature sequence are underlined. (b) Alignment of the N-terminal sequence (deduced from the nucleotide sequence) of AbfA from *B. subtilis* (B. su) with the N-terminal sequence of α-L-arabinofuranosidase from *B. stearothermophilus* (B. st). Double dots represent identical amino acids and single dots represent conservative changes.

The predicted products of *araP* and *araQ* exhibited hydropathy profiles (according to Kyte & Dootlittle, 1982) characteristic of integral membrane proteins: six major regions of high hydrophobicity (hydropathic index > 1·0), each composed of at least 20 aa which could be capable of spanning the membrane (Fig. 3a). AraP and AraQ shared an identity of 19·6 % and showed significant similarity with integral cytoplasmic membrane proteins involved in prokaryotic binding-protein-dependent transport systems (Table 2). In common with most of these integral membrane proteins, AraP and AraQ have a conserved hydrophilic segment (Fig. 3b) at approximately 100 residues from the C terminus with the consensus EAA---G---------I-LP (Dassa & Hofnung, 1985). Furthermore, on the basis of this signature sequence, they can be included in the disaccharide subcluster proposed by Saurin *et al.* (1994) together with the above-mentioned proteins involved in the transport of malto-oligosaccharides, multiple sugars and α-glycerol phosphate.

The deduced product of *abfA*, a hydrophilic protein, displays a N-terminal region (Fig. 2b) which resembles a signal peptide of exoproteins (reviewed on Gierasch, 1989 and Nagarajan, 1993): a positively charged N terminus, a hydrophobic core and a potential cleavage site (AV, position 32–33, Fig. 2b). The primary structure of the putative product of *abfA* is closely related to the



(b)

```
AraP      198 GLQNVPKELYEAADIDGANTMKKFLHITLPFLKPVTVYVL (109)
              .::..:..: :::.::::.....: .:.::.:::: . .:
Consensus     ALQSIPDSLIEAAKIDGAGPFQRFWNIVLPLLKPVLAVLL
                 ..: .:...:..:: . :  :..:. ::.::......
AraQ      165 YALGLPRDLLDSARMDGCTEFGIFFRIMAPLMKPAFGAMI (100)
```

Fig. 3. (a) Hydropathic index for the deduced amino acid sequences of AraP and AraQ according to the algorithm of Kyte & Doolittle (1982). The hydropathy profiles are plotted from the N terminus to the C terminus by averaging hydropathy values over a window of 10 residues. Hydrophobic segments which could correspond to membrane-spanning regions are labelled I–VI. (b) Alignment of a hydrophilic segment, at approximately 100 residues from the C terminus of the predicted sequences of AraP and AraQ, with the consensus sequence for the group of integral cytoplasmic membrane proteins from binding-protein-dependent transport systems (Saurin *et al.*, 1994), which includes permeases involved in the transport of disaccharides and glycerol phosphate. The general consensus for integral membrane proteins from binding-protein-dependent permeases, EAA---G---------I-LP, where (-) represents any amino acid (Dassa & Hofnung, 1985), is underlined. The distance of the invariant glycine residue from the C terminus is represented in parentheses. Double dots represent identical amino acids and single dots represent conservative changes.

α-L-arabinofuranosidase of *Streptomyces lividans* (Table 2) and the N-terminal region (Fig. 2b) is 74% identical and 96% similar to the sequenced N terminus of purified α-L-arabinofuranosidase from *Bacillus stearothermophilus* (Gilead & Shoham, 1995). These observations strongly suggest that *abfA* encodes an α-L-arabinofuranosidase.

RNA transcript analysis of the L-arabinose gene region

Total RNA from cells grown in the presence and absence of L-arabinose was isolated, blotted and hybridized to three different DNA probes (probes 1, 2 and 3, Fig. 1) each specific to one gene of the *ara* region (*araA*, *araM* and *abfA*, respectively). Northern blot analysis (Fig. 4) revealed that *ara* genes are organized in a large polycistronic operon, and that transcripts could be detected only if the cells were grown in the presence of L-arabinose. In addition to a transcript of 11 kb comprising all genes and detected with the three probes, several other signals of different intensities were obtained depending on the probe used (Fig. 4). Using the *araA*-specific probe, we detected five different transcripts of about 8·2, 6·4, 5·8, 4 and 1·9 kb, considering a margin of error of 10–15% for the size determination of transcripts. Two additional transcripts of about 8·2 and 6·4 kb were visualized with the *araM*-specific probe and three hybridization signals were obtained with the *abfA*-specific probe: 8·3, 4·8 and 1·1 kb. Interestingly, stable secondary structures were identified at the corresponding sites within the *araB*, *araL*, *araN* and *araQ* sequences (Fig. 1). The exact nature of these different minor transcripts is unknown but they might be generated by premature transcription termination and/or processing of the multicistronic messenger or RNA degradation. Another possible explanation is the presence of transcription initiation sites located downstream from the promoter defined by primer extension analysis (see below).

The promoter region and transcriptional start site of the *ara* operon

To determine the transcriptional start site of the *ara* operon, total RNA was extracted during the exponential growth of wild-type cultures in the presence and in the absence of L-arabinose. Reverse transcripts were obtained using an end-labelled 17-mer (primer B, Fig. 1), designed to hybridize to part of the *ara*A mRNA. A single extension product was detected with RNA isolated from cells grown in the presence of L-arabinose, the size of which suggests that transcription of the *ara* operon starts at a G residue situated 97 nt upstream from the *araA* start codon (Fig. 5). No extension product was seen when RNA was isolated from cells grown in the absence of L-arabinose. The same transcription start point was obtained using a second primer (primer A, Fig. 1) designed to hybridize to part of the mRNA 50 bases upstream from the first primer (Fig. 5). The synthesis of the *ara* operon mRNA is induced by L-arabinose and driven by a strong promoter as evaluated by the intensity of the reverse transcript signal obtained. Situated 7 and 30 bp upstream from the *ara* operon transcription start site are sequences identical to the consensus −35 and −10 regions (TTGACA-17 bp-TATAAT), respectively, of promoters recognized by *B. subtilis* σ^A-containing RNA polymerase (Moran *et al.*, 1982). Sequence analysis of the promoter region revealed the existence of three inverted repeats, putative

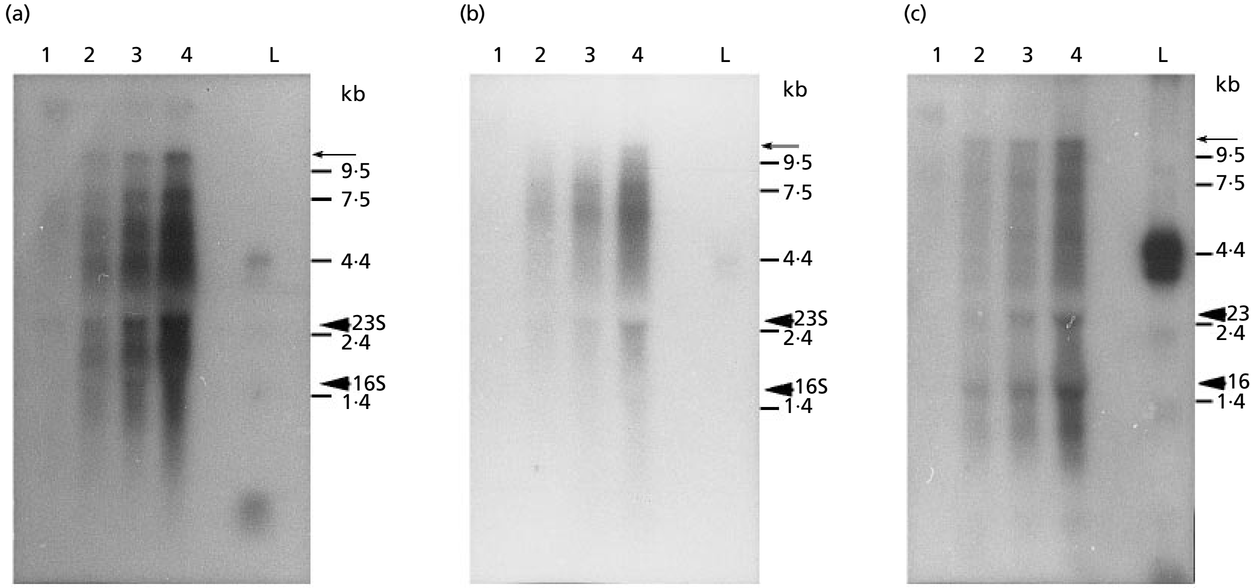


Fig. 4. Northern analysis of the *ara* operon-specific transcripts. Lanes: 1, 10 μg total RNA extracted from the uninduced wild-type strain *B. subtilis* 168T$^+$; 2, 3 and 4, 2·5 μg, 5 μg and 10 μg, respectively, of total RNA extracted from the induced wild-type strain *B. subtilis* 168T$^+$ grown on L-arabinose (see Methods); L, 4 μg RNA ladder (0·24–9·5 kb; Gibco/BRL). The samples were run in 1 % (a, b) and 1·2 % (c) agarose formaldehyde denaturing gel. The ^{32}P-labelled probes used were synthesized from (a) a 1·6 kb *Eco*RI–*Pst*I fragment (position 249–1897, probe 1), (b) a 0·8 kb *Nco*I–*Eco*RV fragment (position 5270–6079, probe 2) and (c) a 0·7 kb *Pst*I–*Ava*I fragment (position 9538–10275, probe 3). The RNA ladder was probed with ^{32}P-labelled λ DNA and also visualized by staining with ethidium bromide. The transcript of about 11 kb comprising all genes and detected with the three probes is indicated by an arrow.

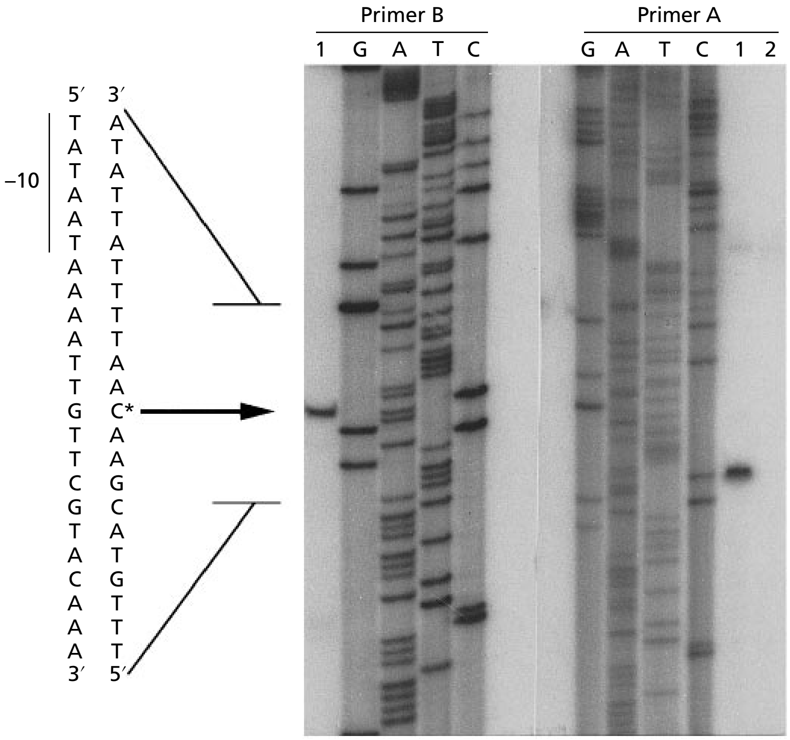


Fig. 5. Primer extension analysis of the *ara* operon promoter. Two radiolabelled oligonucleotide primers, A and B, complementary to two different regions downstream from the *araABD* promoter [primer A, 5′ GAAGCATGTAAACTGCCCC 3′, complementary to a region of *araA* mRNA located between nucleotides 216 and 234 (Fig. 1) and B, 5′ CCAGCGTCTCTTCCCCG 3′, complementary to a region of the *araA* mRNA located between nucleotides 283 and 300 (Fig. 1)] were hybridized with *B. subtilis* BR151 RNA isolated from exponentially growing cells in the presence (lane 1) or absence (lane 2) of L-arabinose. After extension, the products were analysed by gel electrophoresis, together with a set of dideoxynucleotide chain-termination sequencing reactions using the same primers and a single-stranded M13 DNA template which includes the entire *araA* gene and an additional 228 bp of its 5′ flanking sequence.

operator-like sequences, in the −35 and −10 regions (Fig. 1). A potential hairpin–loop structure with a ΔG value of −19·2 kcal mol^{-1} (Tinoco *et al.*, 1973), centred 27 bp upstream from the −35 region (Fig. 1), probably corresponds to a transcription terminator of a gene located upstream from the cloned DNA fragment.

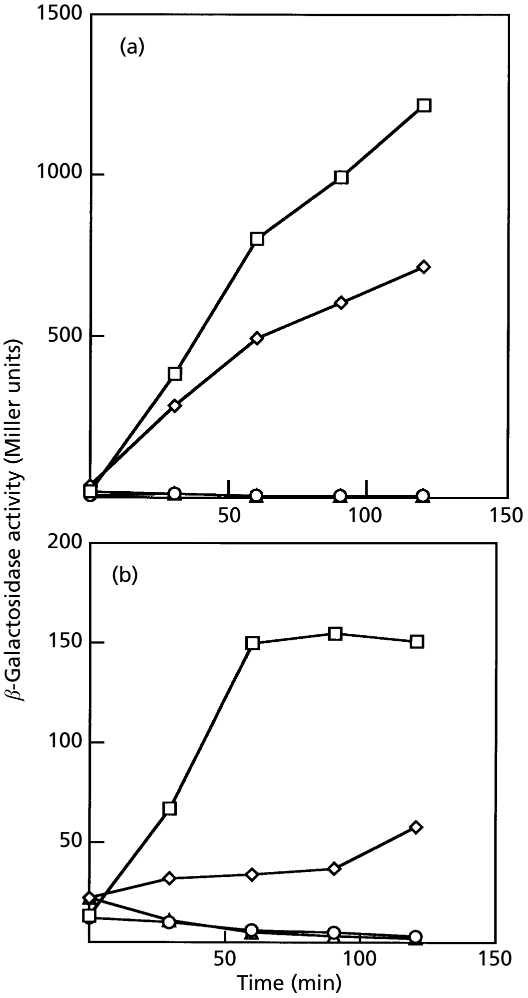


Fig. 6. Expression of the *ara* operon measured by determination of the levels of β-galactosidase activity (Miller units) present in exponentially growing cells. Strains of *B. subtilis* harbouring transcriptional *lacZ* fusions were grown on minimal C medium supplemented with 1 % casein hydrolysate and either (a) 0·4 % L-arabinose or (b) 0·4 % L-arabinose plus 0·4 % glucose (see Methods). Time is expressed in minutes after induction. ◇, IQB101 (*araB′–lacZ erm*; Ara⁻ Em^r LacZ$^+$); □, IQB103 (*araA′–lacZ cat*; Ara$^+$ Cm^r LacZ$^+$); △, IQB102 (*araB′–erm lacZ*; Ara⁻ Em^r LacZ$^-$; negative control); ○, IQB104 (*araA′–cat lacZ*; Ara$^+$ Cm^r LacZ$^-$; negative control). For each strain the results represent the mean, in Miller units, of two independent experiments.

Expression of the *ara* operon is induced by L-arabinose and repressed by glucose

To study the regulation of expression of the operon we constructed transcriptional *lacZ* fusions at this locus. The replicative plasmids pSNL11 and pSNL12, carrying *lacZ* and *erm* (Fig. 1), were linearized and used separately to transform the wild-type 168T$^+$ strain. This resulted in the integration of *lacZ* and *erm* into the chromosome at the *araB* locus. The resulting strains, IQB101 (*araB′–lacZ erm*) and IQB102 (*araB′–erm lacZ*), were unable to grow on L-arabinose as sole carbon source, which confirmed the insertional inactivation of *araB*. The integrational plasmids pSNL13 and pSNL14, carrying the same DNA fragment in opposite orientations (Fig. 1), were integrated as single copy into the chromosome of the wild-type strain 168T$^+$. The resulting strains, IQB103 (*araA′–lacZ cat*) and IQB104 (*araA′–cat lacZ*), respectively, displayed an Ara$^+$ phenotype because the integration was not disruptive. The LacZ phenotype of the four strains was tested on minimal C medium plates supplemented with 1 % casein hydrolysate and X-Gal. Upon addition of L-arabinose to the medium, strains IQB101 and IQB103 presented a dark blue phenotype, whereas those of IQB102 and IQB104 remained white, confirming that the expression of the operon is driven from a promoter located upstream from *araA* and induced by L-arabinose. Furthermore, addition of other pentoses such as D-xylose and D-ribose failed to induce a LacZ$^+$ phenotype in strain IQB103. The regulation of *ara* operon expression was examined in cultures during mid-exponential phase in minimal C medium supplemented with 1 % casein hydrolysate as described in Methods. The levels and patterns of *lacZ* expression in IQB101 (*araB′–lacZ erm*; Ara⁻ LacZ$^+$), IQB103 (*araA′–lacZ cat*; Ara$^+$ LacZ$^+$), IQB102 (*araB′–erm lacZ*; Ara⁻ LacZ⁻; negative control) and IQB104 (*araA′–cat lacZ*; Ara$^+$ LacZ⁻; negative control) determined in the presence of L-arabinose and L-arabinose plus glucose are shown in Fig. 6. When the four strains were grown in the absence of inducer, the level of accumulated β-galactosidase activity, at time $t = 120$ min, was 4·4, 4·8, 2·8 and 1·8 Miller units, respectively. In the presence of L-arabinose the pattern of expression observed in strains IQB101 (*araB′–lacZ*; Ara⁻) and IQB103 (*araA′–lacZ*; Ara$^+$) was very similar (Fig. 6) but the levels of accumulated β-galactosidase activity in the *araB* null mutant were less than 60 % relative to the wild-type strain (discussed below). Addition of glucose reduced the level of expression to less than 12 % in both Ara$^+$ and Ara⁻ backgrounds (Fig. 6). These data demonstrate that L-arabinose is an inducer which stimulates the expression of the *ara* operon at the transcriptional level and transcription is subjected to catabolite repression by glucose. Furthermore, the prediction that the expression of the *ara* operon is driven from a strong promoter, made on the basis of the intensity of the reverse transcript signal observed in primer extension analysis, was confirmed when β-galactosidase activity was measured in strain IQB103 (*araA′–cat lacZ*; Ara$^+$).

araL, M, N, P, Q and *abfA* are not required for L-arabinose utilization

Strains IQB202 and IQB204 in which the integration of plasmids pSS2 and pTN10, respectively, interrupted the transcription unit at *araL* and *araN* (Fig. 1), exhibited an Ara$^+$ phenotype, however, their growth on minimal medium plates with L-arabinose as sole carbon source was slightly slower than the one observed with the wild-type strain 168T$^+$. This phenotype was not observed with strain IQB205 in which pTN13 disrupted the operon at the end of *araQ*. To confirm that *araL*, *M*, *N*, *P*, *Q* and *abfA* are not required for L-arabinose utilization, we constructed a deletion in the region downstream from *araD* by replacing *in vitro* the wild-type sequences of *araL*, *M*, *N*, *P*, *Q* and *abfA* with a Sp^r cassette and then using it to replace the corresponding chromosomal sequences (see Methods). Plasmid pSN22 (Fig. 1) was linearized and used to transform the wild-type strain $168T^+$ Sp^r. The resulting strain IQB206, was Km^s which indicated that the Sp^r phenotype was the result of a double cross-over event that occurred on both sides of the cassette inserted between the *araL* and *abfA* sequences (Fig. 1). This mutant strain was able to grow on minimal medium plates with L-arabinose but displayed a phenotype even more drastic than the one exhibited with strains IQB202 and IQB204. To quantify this observation we determined the specific growth rate of the deletion-insertion mutant and the wild-type strain in liquid minimal C medium with L-arabinose as sole carbon source, as described in Methods. The doubling time of strain IQB206 was 1·8-fold higher than the wild-type strain $168T^+$, $193{\cdot}4 \pm 7{\cdot}2$ and $107{\cdot}7 \pm 3{\cdot}6$ min (means of three independent experiments ± SEM), respectively. These results confirmed that the genes located downstream from *araD* in the operon are not essential for L-arabinose utilization, however their absence in the deletion mutant affects the specific growth rate in minimal medium with L-arabinose as the sole carbon source when compared to the wild-type strain.

DISCUSSION

In this study we have described a new catabolic operon involved in the utilization of L-arabinose in *B. subtilis*, which we designated *ara*. The arabinose metabolic genes *araA*, *araB* and *araD*, encoding L-arabinose isomerase, L-ribulokinase and L-ribulose-5-phosphate 4-epimerase, respectively, were cloned previously and by complementation experiments the products of *araB* and *araD* were shown to be functionally homologous to their *E. coli* counterparts (Sá-Nogueira & Lencastre, 1989). These genes, whose inactivation leads to an Ara^- phenotype, were found to be the first three ORFs of a nine cistron transcriptional unit whose total length is 11 kb. To our knowledge this operon is the largest catabolic operon described in *B. subtilis*. As expected from the occurrence of genetic complementation, the deduced products of *araA*, *araB* and *araD* from *B. subtilis* display a very high level of identity to the corresponding enzymes from *E. coli* and *Sal. typhimurium*, which indicates that this metabolic pathway was fundamentally conserved during evolution. In *B. subtilis* the metabolic gene order, *araABD*, coincides with the order of the enzymic steps carried out by the proteins they encode. This order is different from the one found in the operons of the *Enterobacteriaceae* members *E. coli* and *Sal. typhimurium*, *araBAD*, so it seems that the three genes did not act as a unitary block in the evolution of the eubacterial *ara* genes.

The six ORFs found downstream from *araD*, here named *araL*, *M*, *N*, *P*, *Q* and *abfA*, are not required for L-arabinose utilization. This was shown in a mutant strain, IQB206, bearing a deletion in the region downstream from *araD* comprising all genes. The function of *araL* and *araM* is unknown. The putative product of *araM* did not show any significant similarity with other bacterial proteins of known function and the weak similarities displayed by *araL* did not suggest any particular function. Interestingly, the N-terminal sequence of *araL* shares an identity of 18·7% over 193 residues with the C-terminal sequences of *araM* (data not shown). The primary sequences of the products of *araN*, *araP* and *araQ* strongly suggest that they have a similar function to that of a superfamily of membrane-bound nutrient transport systems (Higgins *et al.*, 1990). Sequence similarities to known import proteins and the organization of the genes in the operon revealed the presence of three components of these transport systems. Firstly, the N terminus of AraN has a predicted signal peptide and sequences typical of Gram-positive lipoproteins (IAGCSA, starting at aa 19). We therefore suggest that AraN might be anchored in the cytoplasmic membrane via an amino-lipid group (Gilson *et al.*, 1988; Perego *et al.*, 1991). Secondly, *araP* and *araQ* gene products, as other characterized integral cytoplasmic membrane proteins, have hydropathy profiles which are virtually superimposable and some of their residues are apparently conserved (Fig. 3). Finally, *araN*, *araP* and *araQ* belong to the same operon and the ligand-specific binding protein, AraN, is encoded by the promoter-proximal gene, a situation common to these systems. In *B. subtilis* the phosphotransferase system is not involved in the transport of L-arabinose into the cell (Gay *et al.*, 1973). Therefore, it is tempting to propose that AraN, AraP and AraQ are components of a high affinity transport system for L-arabinose. However, no evident ATP-binding protein connected with energy coupling of the transport system was found in the operon.

The transport of L-arabinose across the *E. coli* cytoplasmic membrane requires the expression of either the high-affinity transport operon, *araFGH*, a binding-protein-dependent system (Horazdovsky & Hogg, 1989; Kolodrubetz & Schleif, 1981) or the low-affinity transport operon, *araE*, a proton symporter (Novotny & Englesberg, 1966). The existence of two parallel uptake systems thwarts usual genetic attempts to isolate mutants defective in either of the transport systems. The Ara^+ phenotype displayed by the *B. subtilis* deletion-insertion mutant strain IQB206 (Δ*ara–abfA*::*spc*) together with the 1.8-fold increase in doubling time observed on liquid minimal medium with L-arabinose as the sole carbon source, relative to the wild-type strain, is typical of a transport mutant when the micro-organism has alternative transport systems for the same substrate. An additional explanation for this phenotype observed in the deletion-insertion mutant is that insertion of *spc* might result in a less stable mRNA encoding *araABD*, leading to decreased amounts of their products. Interestingly, the primary structure of AraP and AraQ showed weak similarity with AraH, the integral cytoplasmic membrane protein from *E. coli*, and the same result was observed between AraN and AraF, the *E. coli* arabinose binding protein (data not shown). Furthermore, on the basis of their signature sequences, AraN, AraP and AraQ can be included in the disaccharide subcluster (Figs 2 and 3) together with proteins involved in the high-affinity transport of malto-oligosaccharides and multiple sugars. *B. subtilis* secretes three enzymes involved in the degradation of L-arabinose polymers, an endo-arabanase and two arabinosidases, and the purified endo-arabanase has been shown to be capable of releasing arabinosyl oligomers from plant cell walls (Kaji & Saheki, 1975; Weinstein & Albersheim, 1979). To account for these observations a wider substrate range, L-arabinose and/or L-arabinose oligomers, for the *B. subtilis* AraN binding protein is suggested. The last gene of the *ara* operon, *abfA*, probably encodes a α-L-arabinofuranosidase, based on the strong similarity observed between the primary structure of its putative product and other bacterial arabinosidases. Whether this enzyme is extracellular or intracellular is unknown.

Expression of the *ara* operon is induced by L-arabinose and driven by a promoter located upstream of *araA*. This has been demonstrated in this study by Northern blotting and primer extension analysis. Examination of the *ara* operon promoter reveals −35 and −10 sequences, relative to its transcriptional start site (shown in Fig. 1), separated by an optimal spacing of 17 bp, identical to the consensus sequences derived from the analysis of many σ^A-dependent promoters (Moran *et al.*, 1982). These sequences were shown to be important for the interaction of σ^A with their cognate promoters (reviewed in Moran, 1993). The presence of a strong promoter raises the possibility that transcription of *ara* is negatively regulated like in other well characterized *B. subtilis* catabolic operons, such as *xyl* (Gärtner *et al.*, 1992) and *gnt* (Fujita & Fujita, 1987); in fact the product of *araC* recently cloned, is a negative regulator of the *ara* operon (I. Sá-Nogueira & L. J. Mota, unpublished). To characterize the regulation of *ara* expression in greater detail we constructed transcriptional fusions of the *ara* promoter to the *E. coli lacZ* gene in Ara^+ and Ara^- strains. The induction by L-arabinose in the Ara^+ background was approximately 100-fold and the pattern of expression observed in Ara^- and Ara^+ strains was very similar. Interestingly however, the levels of accumulated β-galactosidase activity in the Ara^- background were less than 60 % of the fully induced level in the wild-type strain. Since in this strain the *ara* transcription unit is interrupted at the level of *araB* (Fig. 1), and a role in the transport of L-arabinose was proposed for the downstream genes *araN*, *araP* and *araQ*, this effect could be due to less accumulated intracellular L-arabinose which prevents full expression of the *ara* promoter. Another possible explanation is that the products of *araL* and *araM* could stimulate transcription from the *ara* promoter. Addition of glucose reduced the level of expression to less than 12 % in both Ara^+ and Ara^- backgrounds, indicating that repression of the *ara* operon by glucose acts at the transcriptional level.

The regulatory system mediating catabolite repression in *B. subtilis* seems to be accomplished by a negative regulatory mechanism (reviewed in Hueck & Hillen, 1995; Saier *et al.*, 1996). This evidence is based on the location and the sequences of *cis*-acting sites (CREs) responsible for catabolite repression of several *B. subtilis* genes and operons. Moreover, catabolite repression of most genes regulated via these *cis*-acting sites is also affected by the *trans*-acting factors CcpA, a DNA-binding protein, and HPr, an intermediate in the phosphotransferase sugar transport system. It has been proposed that HPr-Ser-P might interact with CcpA and that this interaction might allow CcpA to bind to the CRE (Deutscher *et al.*, 1994). Strong evidence for this proposal, but also contradictory results, have been obtained recently (Saier *et al.*, 1996; and references therein). CREs of catabolic genes and operons are located either in the promoter regions, where the binding of a regulatory protein probably interferes with transcription initiation, or in the downstream regions (reviewed in Hueck & Hillen, 1995). In the case of the *hut* operon two active CREs were found, one at the promoter and the other within *hutP*, and a looping mechanism involving co-operatively bound CREs has been proposed to interfere with transcription initiation (Wray *et al.*, 1994). Furthermore, the transition-state regulator AbrB is capable of specifically binding to *hut* CRE *in vitro* and an *abrB* null mutation leads to more efficient catabolite repression of some genes in *B. subtilis*, including L-arabinose isomerase. Thus, AbrB has been suggested to compete for binding to CRE with CcpA (Fisher *et al.*, 1994). The promoter region of the *ara* operon contains a sequence very similar to the CRE consensus sequence (TGWNANCGNTNWCA; W = A, T; Weickert & Chambliss, 1990) located between the transcription start site and the translation start site of *araA* (position 191–204, Fig. 1). A second sequence, which shows weak similarity with the CRE consensus sequence was found within *araA* (position 260–273, Fig. 1). Since inducer exclusion does not play a major role in carbon regulation of expression of the *ara* metabolic genes (Sá-Nogueira *et al.*, 1988), as observed in the *hut* operon (Chasin & Magasanik, 1968), it will thus be interesting to investigate the role of CcpA, HPr and AbrB in the catabolite repression of the *ara* operon and whether these sequences are *cis*-acting sites responsible for catabolite repression of the *ara* genes.

ACKNOWLEDGEMENTS

We would like to thank Drs M. Debarbouillé, Adriano O. Henriques, Helena Paveia for helpful discussions and A. O. Henriques and C. P. Moran Jr for the gift of plasmids. T. V. Nogueira was the recipient of a fellowship from Junta Nacional de Investigação Científica e Tecnológica (JNICT). This work was supported by grants 87.218/Bio and 1287/92/Bio, from Junta Nacional de Investigação Científica e Tecnológica (JNICT).

REFERENCES

Anagnostopoulos, C. & Spizizen, J. (1961). Requirements for transformation in *Bacillus subtilis*. *J Bacteriol* **81**, 741–746.

Chasin, L. A. & Magasanik, B. (1968). Induction and repression of the histidine-degrading enzymes of *Bacillus subtilis*. *J Biol Chem* **243**, 5165–5178.

Dale, R. M. K., McClure, B. A. & Houchins, J. P. (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.

Dassa, E. & Hofnung, M. (1985). Sequence of *malG* gene in *E. coli* K12: homologies between integral membrane components from binding protein-dependent transport systems. *EMBO J* **4**, 2287–2293.

Debarbouillé, M., Arnaud, M., Foust, A., Klier, A. & Rapoport, G. (1990). The *sacT* gene regulating the *sacPA* operon in *Bacillus subtilis* shares strong homology with transcriptional anti-terminators. *J Bacteriol* **172**, 3966–3973.

Deutscher, J., Reizer, J., Fischer, C., Galinier, A., Saier, M. H., Jr & Steinmetz, M. (1994). Loss of protein kinase-catalyzed phosphorylation of Hpr, a phospho-carrier protein of the phosphotransferase system, by mutation of the *ptsH* gene confers catabolite repression resistence to several catabolic genes of *B. subtilis*. *J Bacteriol* **176**, 3336–3344.

Englesberg, E., Squires, C. & Meronk, F. (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.

Ferrari, E., Nguyen, A., Lang, D. & Hoch, J. (1983). Construction and properties of an integrable plasmid for *Bacillus subtilis*. *J Bacteriol* **154**, 1513–1515.

Fisher, S. H., Strauch, M. A., Atkinson, M. R. & Wray, L. V., Jr (1994). Modulation of *Bacillus subtilis* catabolite repression by transition state regulatory protein AbrB. *J Bacteriol* **176**, 1903–1912.

Fujita, Y. & Fujita, T. (1987). The gluconate operon *gnt* of *Bacillus subtilis* encodes its own transcriptional negative regulator. *Proc Natl Acad Sci USA* **84**, 4524–4528.

Gärtner, D., Degenkolb, J., Rippberger, J., Allmansberger, R. & Hillen, W. (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.

Gay, P., Cordier, P., Marquet, M. & Delobbe, A. (1973). Carbohydrate metabolism and transport in *Bacillus subtilis*. A study of *ctr* mutations. *Mol Gen Genet* **121**, 355–368.

Gierasch, L. M. (1989). Signal sequences. *Biochemistry* **28**, 923–930.

Gilead, S. & Shoham, Y. (1995). Purification and characterization of α-L-arabinofuranosidase from *Bacillus stearothermophilus* T-6. *Appl Environ Microbiol* **61**, 170–174.

Gilson, E., Alloing, G., Schmidt, T., Claverys, J.-P., Dudler, R. & Hofnung, M. (1988). Evidence for high-affinity binding-protein dependent systems in Gram-positive bacteria and *Mycoplasma*. *EMBO J* **7**, 3971–3974.

Hayashi, S. & Wu, H. C. (1990). Lipoproteins in bacteria. *J Bioenerg Biomembr* **22**, 451–471.

Higgins, C. F., Hyde, S. C., Mimmack, M. M., Gileadi, U., Gill, D. R. & Gallagher, M. P. (1990). Periplasmic binding-protein dependent systems. *J Bioenerg Biomemb* **22**, 571–592.

Horazdovsky, B. & Hogg, R. (1989). Genetic reconstitution of the high-affinity L-arabinose operon in *Escherichia coli*. *J Bacteriol* **171**, 3053–3059.

Hueck, C. J. & Hillen, W. (1995). Catabolite repression in *Bacillus subtilis*: a global regulatory mechanism for the Gram-positive bacteria? *Mol Microbiol* **15**, 395–401.

Igo, M. M. & Losick, R. (1986). Regulation of a promoter that is utilized by minor forms of RNA polymerase holoenzyme in *Bacillus subtilis*. *J Mol Biol* **191**, 615–624.

Kaji, A. & Saheki, T. (1975). Endo-arabanase from *Bacillus subtilis* F-11. *Biochim Biophys Acta* **410**, 354–360.

Kaneko, Y., Toh-e, A., Banno, I. & Oshima, Y. (1989). Molecular characterization of a specific *p*-nitrophenylphosphatase gene, *PHO13*, and its mapping by chromosome fragmentation in *Saccharomyces cerevisiae*. *Mol Gen Genet* **220**, 133–139.

Katz, L. (1970). Selection of *araB* and *araC* mutants of *Escherichia coli* B/r by resistance to ribitol. *J Bacteriol* **102**, 593–595.

Kolodrubetz, D. & Schleif, R. (1981). L-Arabinose transport systems in *Escherichia coli* K12. *J Bacteriol* **148**, 472–479.

Kyte, J. & Doolittle, R. F. (1982). A simple method for displaying the hydropathic character of a protein. *J Mol Biol* **157**, 105–132.

Lepesant, J. A. & Dedonder, R. (1967a). Metabolisme du L-arabinose chez *Bacillus subtilis* Marburg Ind$^-$ 168. *C R Acad Sci Ser D*, 2683–2686.

Lepesant, J. A. & Dedonder, R. (1967b). Isolement de mutants du système du L-arabinose chez *Bacillus subtilis* Marburg Ind$^-$ 168. *C R Acad Sci Ser D*, 2832–2835.

Martin, I., Debarbouillé, M., Ferrari, E., Klier, A. & Rapoport, G. (1987). Characterization of the levanase gene of *Bacillus subtilis* which shows homology to yeast invertase. *Mol Gen Genet* **208**, 177–184.

Miller, J. H. (1972). *Experiments in Molecular Genetics*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.

Moran, C. P., Jr (1993). RNA polymerase and transcription factors. In *Bacillus subtilis and Other Gram-positive Bacteria: Biochemistry, Physiology and Molecular Genetics*, pp. 653–667. Edited by A. L. Sonensheim, J. A. Hoch & R. Losick. Washington, DC: American Society for Microbiology.

Moran, C. P., Jr, Lang, N., LeGrice, S. F. J., Lee, G., Stephens, M., Sonensheim, A. L., Pero, J. & Losick, R. (1982). Nucleotide sequences that signal the initiation of transcription in *Bacillus subtilis*. *Mol Gen Genet* **186**, 339–346.

Nagarajan, V. (1993). Protein secretion. In *Bacillus subtilis and Other Gram-positive Bacteria: Biochemistry, Physiology and Molecular Genetics*, pp. 713–726. Edited by A. L. Sonensheim, J. A. Hoch & R. Losick. Washington, DC: American Society for Microbiology.

Novotny, C. & Englesberg, E. (1966). The L-arabinose permease system in *Escherichia coli* B/r. *Biochim Biophys Acta* **117**, 217–230.

Pascal, M., Kunst, F., Lepesant, J. A. & Dedonder, R. (1971). Characterization of two sucrase activities in *Bacillus subtilis* Marburg. *Biochem* **53**, 1059–1066.

Paveia, H. & Archer, L. (1992a). Genes for L-arabinose utilization in *Bacillus subtilis*. *Brotéria Genética Lisboa* **XIII** (**LXXX**), 149–159.

Paveia, H. & Archer, L. (1992b). Mapping of *ara* genes in *Bacillus subtilis*. *Brotéria Genética Lisboa* **XIII** (**LXXX**), 161–167.

Perego, M. (1993). Integrational vectors for genetic manipulation in *Bacillus subtilis*. In *Bacillus subtilis and Other Gram-positive Bacteria: Biochemistry, Physiology and Molecular Genetics*, pp. 615–624. Edited by A. L. Sonensheim, J. A. Hoch & R. Losick. Washington, DC: American Society for Microbiology.

Perego, M., Higgins, C. F., Pearce, S. R., Gallagher, M. P. & Hoch, J. A. (1991). The oligopeptide transport system of *Bacillus subtilis* plays a role in the initiation of sporulation. *Mol Microbiol* **5**, 173–185.

Perkins, J. B. & Youngman, P. J. (1986). Construction and properties of Tn*917–lac*, a transposon derivative that mediates transcriptional gene fusions in *Bacillus subtilis*. *Proc Natl Acad Sci USA* **83**, 140–144.

Plumbridge, J. A. (1989). Sequence of the *nagBACD* operon in *Escherichia coli* K12 and pattern of transcription within the *nag* regulon. *Mol Microbiol* **3**, 505–515.

Saier, M. H., Jr, Chauvaux, S., Cook, G. M., Deutscher, J., Paulsen, I. T., Reizer, J. & Ye, J.-J. (1996). Catabolite repression and inducer control in Gram-positive bacteria. *Microbiology* **142**, 217–230.

Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989). *Molecular Cloning: a Laboratory Manual* 2nd edn. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.

Sanger, F., Nicklen, S. & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibition. *Proc Natl Acad Sci USA* **74**, 140–144.

Sá-Nogueira, I. & Lencastre, H. (1989). Cloning and characterization of *araA*, *araB* and *araD*, the structural genes for L-arabinose utilization in *Bacillus subtilis*. *J Bacteriol* **171**, 4088–4091.

Sá-Nogueira, I., Paveia, H. & Lencastre, H. (1988). Isolation of constitutive mutants for L-arabinose utilization in *Bacillus subtilis*. *J Bacteriol* **170**, 2855–2857.

Saurin, W., Köster, W. & Dassa, E. (1994). Bacterial binding protein-dependent permeases: characterization of distinctive signatures for functionally related integral cytoplasmic membrane proteins. *Mol Microbiol* **12**, 993–1004.

Sullivan, M. A., Yasbin, R. E. & Young, F. E. (1984). New shuttle vectors for *Bacillus subtilis* and *Escherichia coli* which allow rapid detection of inserted fragments. *Gene* **29**, 21–26.

Tam, R. & Saier, M. H., Jr (1993). Structural, functional, and evolutionary relationships among extracellular solute-binding receptors of bacteria. *Microbiol Rev* **57**, 320–346.

Tinoco, I., Borer, P. N., Dengler, B., Levine, M. D., Uhlenbeck, O. C., Crothers, D. M. & Gralla, J. (1973). Improved estimation of secondary structure in ribonucleic acids. *Nature New Biol* **246**, 40–41.

Weickert, M. J. & Chambliss, G. H. (1990). Site-directed mutagenesis of a catabolic repression operator sequence in *Bacillus subtilis*. *Proc Natl Acad Sci USA* **87**, 6238–6242.

Weinstein, L. & Albersheim, P. (1979). Structure of plant cell walls. IX. Purification and partial purification of a wall-degrading endoarabanase and an arabinosidase from *Bacillus subtilis*. *Plant Physiol* **63**, 425–432.

Wray, L. V., Jr, Pettengill, F. K. & Fisher, S. H. (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.

Yang, J., Dhamija, S. S. & Schweingruber, M. E. (1991). Characterization of a specific *p*-nitrophenylphosphatase gene and protein of *Schizosaccharomyces pombe*. *Eur J Biochem* **198**, 493–497.

Yanisch-Perron, C., Vieira, J. & Messing, J. (1985). Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene* **33**, 103–119.

Received 10 July 1996; revised 7 October 1996; accepted 9 October 1996.