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

Isa be1 Sd-Nlogueira, Teresa V. Nogueira, Sónia Soares and Hermínia de Lencastre

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

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 arabino-sidases, 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...
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 subject 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 lenA, 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^T), 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 araI, 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 I. 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, 30 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, plates were plated on minimal C medium containing 0.1% (w/v) L-arabinose. The Ara^- 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 OD₆₀₀. 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 polylinking 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'- CCCTCTGCTATTACGCC'C 3', complementary to the coding sequence of lacZ, was used to sequence the transcriptional lacZ fusions.

**Plasmid constructions.** pSNL7 was constructed by subconling a 959 bp Smal–PstI DNA fragment (nt 938–1897, Fig. 1) from pSNL1 (Sá-Nogueira & Lencastre, 1989) between the Smal and PstI sites of the integrational vector pJM783 (Perego, 1983). To construct pSS2, we digested pSNL1 (Sá-Nogueira & Lencastre, 1989) with HindIII and XhoI and cloned a purified fragment of 965 bp (nt 3815–4780, Fig. 1) between the HindIII and SalI sites of the integrating vector pJH101 (Ferrari et al., 1983). pTN10 was obtained by subcloning a 789 bp HindIII–HindII DNA fragment (nt 6545–7334, Fig. 1) from pSS3 between the HindIII and EcoRV sites of the integrational vector pJH101 (Ferrari et al., 1983). pTN14was constructed by subcloning the 678 bp Smal–BglII DNA fragment (nt 8242–8920, Fig. 1) from pTN13 between the BamHI and SstI (fill-in) sites of pJM783 (Perego, 1993). pSNL10 was obtained by subcloning a 1 kb EcoRI–HindII fragment (nt 2681–4416, Fig. 1) from pSNL1 (Sá-Nogueira & Lencastre, 1989) between the EcoRI and Smal sites of pMK4 (Sullivan et al., 1984).
pSNL11 and pSNL12 were obtained as follows. A 4.5 kb BamHI–HindIII (fill-in) fragment extracted from pMC11 (Debarbouille et al., 1990), containing lacZ and erm from pTV32 (Perkins & Youngman, 1986), was subcloned in both orientations at the unique EcoRV restriction site (nt 3214, Fig. 1) of pSNL10. pSNL13 and pSNL14 were obtained by subcloning a 470 bp DraI–EcoRV DNA fragment (nt 82–552, Fig. 1) from pSNL9 at the unique SmaI site of the integrational vector pJ783 (Ferego, 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 EcoRV–HindIII fragment (nt 3214–4416, Fig. 1) from pSS3 into the Smal site of pAH248 [a pGem-7zf(+) (Promega) derivative that contains a Km' gene cloned between its XhoI and EcoRI 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 EcoRV fragment from pSN5 (nt 10632–12332, Fig. 1) was inserted into the HindIII site of pAH250 [a pBluescript SK+ (Stratagene) derivative that contains a Sp' gene (spc) cloned into the EcoRV site (A. O. Henriques, B. W. Beall & C. P. Moran Jr, personal communication)]. To construct pSN22, we digested pSN20 with PstI and NsiI and cloned a purified fragment of about 2790 bp, containing the Km' gene, in the Smal site of pSN21. pSNL9, pSS3, pTN13 and pSN5 were obtained by cutting chromosomal DNA from *B. subtilis* strains pIQ800, pIQ205, pIQ204 and pIQ203 (Table 1) with HindIII, EcoRI, NcoI and Smal, 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 (OD₆₀₀ = 0.1–0.15), 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, t₀, 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 OD₆₀₀ and at 30 min intervals, 100 µl of cell culture samples was removed and stored at −70 °C until the cultures reached an OD₆₀₀ of 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 (OD₆₀₀ ~ 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-³²P]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' GAGACTGTAAACTGCCCC 3'), complementary to nt 216–234 (Fig. 1), and primer B (5' CCAGCGCTAATAACTGCCCC 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...
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
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 Smal 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

---

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

<table>
<thead>
<tr>
<th>B. subtilis AraA protein</th>
<th>Homologue (species/accession no.)*</th>
<th>Function</th>
<th>Identity (%)</th>
<th>Amino acid overlap</th>
</tr>
</thead>
<tbody>
<tr>
<td>AraA</td>
<td>AraA (E. coli/M15263)</td>
<td>t-Arabinose isomerase</td>
<td>52.9</td>
<td>495</td>
</tr>
<tr>
<td>AraA</td>
<td>AraA (Sal. typhimurium/M11047)</td>
<td>t-Arabinose isomerase</td>
<td>52.9</td>
<td>495</td>
</tr>
<tr>
<td>AraB</td>
<td>AraB (E. coli/M15263)</td>
<td>t-Ribulokinase</td>
<td>25.7</td>
<td>552</td>
</tr>
<tr>
<td>AraB</td>
<td>AraB (Sal. typhimurium/M11045)</td>
<td>t-Ribulokinase</td>
<td>30.6</td>
<td>350</td>
</tr>
<tr>
<td>AraD</td>
<td>AraD (E. coli/M15263)</td>
<td>t-Ribulose-5-P 4-epimerase</td>
<td>57.1</td>
<td>231</td>
</tr>
<tr>
<td>AraD</td>
<td>AraD (Sal. typhimurium/M11046)</td>
<td>t-Ribulose-5-P 4-epimerase</td>
<td>58.0</td>
<td>205</td>
</tr>
<tr>
<td>AraL</td>
<td>NagD (E. coli/X14135)</td>
<td>Unknown</td>
<td>23.5</td>
<td>251</td>
</tr>
<tr>
<td>AraN</td>
<td>LacE (Agrobacterium radiobacter/X66596)</td>
<td>Lactose-binding protein</td>
<td>26.2</td>
<td>302</td>
</tr>
<tr>
<td></td>
<td>MalX (Streptococcus pneumoniae/L08611)</td>
<td>Malto-binding protein</td>
<td>24.1</td>
<td>345</td>
</tr>
<tr>
<td></td>
<td>AmyE (Thermoanaerobacterium thermosulfurigenes/M57692)</td>
<td>Starch-binding protein</td>
<td>21.7</td>
<td>369</td>
</tr>
<tr>
<td>AraP</td>
<td>LacF (Agrobacterium radiobacter/X66596)</td>
<td>Membrane protein</td>
<td>29.6</td>
<td>284</td>
</tr>
<tr>
<td></td>
<td>UgpA (E. coli/X13141)</td>
<td>Membrane protein</td>
<td>26.2</td>
<td>286</td>
</tr>
<tr>
<td></td>
<td>AmyD (Thermoanaerobacterium thermosulfurigenes/M57692)</td>
<td>Membrane protein</td>
<td>25.4</td>
<td>284</td>
</tr>
<tr>
<td></td>
<td>MalC (Streptococcus pneumoniae/L08611)</td>
<td>Membrane protein</td>
<td>25.2</td>
<td>298</td>
</tr>
<tr>
<td></td>
<td>LacG (Agrobacterium radiobacter/X66596)</td>
<td>Membrane protein</td>
<td>32.7</td>
<td>254</td>
</tr>
<tr>
<td></td>
<td>UgpE (E. coli/X13141)</td>
<td>Membrane protein</td>
<td>22.9</td>
<td>279</td>
</tr>
<tr>
<td></td>
<td>AmyC (Thermoanaerobacterium thermosulfurigenes/M57692)</td>
<td>Membrane protein</td>
<td>28.2</td>
<td>262</td>
</tr>
<tr>
<td></td>
<td>MalD (Streptococcus pneumoniae/L08611)</td>
<td>Membrane protein</td>
<td>25.6</td>
<td>262</td>
</tr>
<tr>
<td>AbfA</td>
<td>AbfA (Streptomyces lividans/U04630)</td>
<td>a-1-Arabinofuranosidase</td>
<td>52.6</td>
<td>500</td>
</tr>
</tbody>
</table>
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 (Sa-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, 48.4, 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 (T1 and T2, Fig. 1, with ΔG values of -27.4 and -18.7 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-ribulokinase 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 (Kaneo 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 hydrophathy 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).

The predicted products of araP and araQ exhibited hydrophathy profiles (according to Kyte & Doolittle, 1982) characteristic of integral membrane proteins: six major regions of high hydrophobicity (hydrophatic 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 YO 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-EP (Dassa & Hofnung, 1985). Furthermore, on the basis of this signature sequence, they can be included in the disaccharide sub-cluster 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

---

**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. steaethermophilus (B. st). Double dots represent identical amino acids and single dots represent conservative changes.
**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 82, 64, 58, 4 and 19 kb, considering a margin of error of 10–15% for the size determination of transcripts. Two additional transcripts of about 82 and 64 kb were visualized with the araM-specific probe and three hybridization signals were obtained with the abfA-specific probe: 83, 48 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 araA 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). 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).
Fig. 4. Northern analysis of the ara operon-specific transcripts. Lanes: 1, 10 µg total RNA extracted from the uninduced wild-type strain \textit{B. subtilis} 168T\textsuperscript{+}; 2, 3 and 4, 2.5 µg, 5 µg and 10 µg, respectively, of total RNA extracted from the induced wild-type strain \textit{B. subtilis} 168T\textsuperscript{+} 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 \textsuperscript{32}P-labelled probes used were synthesized from (a) a 1.6 kb EcoRI-PstI fragment (position 249-1897, probe 1), (b) a 0.8 kb Ncol-EcoRV fragment (position 5270-6079, probe 2) and (c) a 0.7 kb PstI-Aval fragment (position 9538-10275, probe 3). The RNA ladder was probed with \textsuperscript{32}P-labelled \textit{A} 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 \textit{araA} mRNA located between nucleotides 216 and 234 (Fig. 1) and B, 5' CCAGCGTCTCTTCCCCG 3', complementary to a region of the \textit{araA} mRNA located between nucleotides 283 and 300 (Fig. 1)) were hybridized with \textit{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 \textit{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 $\Delta G$ value of $-19.2$ kcal mol$^{-1}$ (Tinoco \textit{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.
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 strain 168T+. 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−).
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' 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'. The resulting strain IQB206, was Km' which indicated that the Sp' 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+; 193.4 ± 7.2 and 107.7 ± 3.6 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 hydrophathy 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 (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 sub-cluster (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 t-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, t-arabinose and/or t-arabinose oligomers, for the B. subtilis AraN binding protein is suggested. The last gene of the ara operon, abfA, probably encodes a x-t-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 t-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 σA-dependent promoters (Moran et al., 1982). These sequences were shown to be important for the interaction of σA 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 t-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 t-arabinose was proposed for the downstream genes araN, araP and araQ, this effect could be due to less accumulated intracellular t-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 t-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 (TGWNANCNGTNWCA; 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 Investiga@o Cientifica e Tecnologica (JNICT). This work was supported by grants 87.218/Bio and 1287/92/Bio, from Junta Nacional de Investiga@o Cientifica e Tecnologica (JNICT).

REFERENCES


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