Nucleotide Sequence of Sporulation Locus spoIIA in Bacillus subtilis

By P. FORT*† AND P. J. PIGGOT‡

Microbiology Unit, Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, UK

(Received 14 December 1983; revised 7 March 1984)

We have determined a sequence of 2073 bp from two recombinant plasmids carrying the whole spoIIA locus from Bacillus subtilis, the expression of which is required for spore formation. The sequence contains three long open reading frames (ORFs), each of them being preceded by a ribosome binding site. These three putative proteins (mol. wts 13100, 16300 and 22200) are likely to be expressed and are probably encoded on the same mRNA. The stop codon of ORF1 overlaps with the start codon of ORF2 suggesting that there might be translational coupling between the two ORFs. Although some known promoter sequences were found, the only one upstream from the first open reading frame is about 260 bp from it.

INTRODUCTION

The known number of genetic loci in Bacillus subtilis specifically concerned with spore formation and germination is about 50 (Piggot & Coote, 1976; Young & Mandelstam, 1979; Piggot et al., 1981). The spoIIA locus is particularly interesting as mutations in it lead to various phenotypes, including asporogeny or oligosporogeny, and a lack of production of several enzymes (Ionesco & Schaeffer, 1968; Coote, 1972a,b; Piggot, 1973; Akrigg & Mandelstam 1978; Yudkin & Turley, 1980; Errington & Mandelstam, 1983). A part of the locus was cloned in a plasmid and was shown to complement mutations at one end of the locus (Liu et al., 1982). The rest of the locus has now been cloned in a plasmid (Piggot et al., 1984) and the whole of the locus in a phage (Savva & Mandelstam, 1984). In this paper, we publish the nucleotide sequence of the whole of the spoIIA locus.

METHODS

Bacterial strains, phages and plasmids. Bacillus subtilis strains carrying the spoIIA mutations I, 37, 42 and 69, have been described previously (Piggot, 1973; Yudkin & Turley, 1980; Errington & Mandelstam, 1983). Escherichia coli strains containing the plasmids pHM2 and pPP33 are described elsewhere (Liu et al., 1982; Piggot et al., 1984). Escherichia coli JM103 [Δ(lac-pro) thi strA supE endA sbcB15 hsdR4 F’ traD36 proA.B lacP, ZAM15] was kindly supplied by I. Jones. M13 mp8, mp9, mp10 and mp11 replicative form (RF) DNA was either purified in the laboratory (see below) or purchased from BRL. Plasmic pUC13, which contains the polylinker sites for the restriction endonucleases HindIII, PstI, SalI, XbaI, BamHI, SmaI, SstI and EcoRI (Messing & Vieira, 1982), was a gift from J. Errington.

DNA preparation. Plasmid DNA and M13 RF DNA were isolated by alkali denaturation and phenol extraction (McMaster et al., 1980), with the following modifications (G. K. McMaster, personal communication). Cultures (500 ml) were grown to an OD600 value of 0.4-0.5, and then centrifuged at 5000 r.p.m. for 10 min. Subsequent manipulations were at 0°C. The pellet was resuspended in 10 ml 0.1% (w/v) tri-isopropyl-naphthalenesulphonic acid.

* Permanent address: Laboratoire de Biologie Moléculaire, Université de Sciences et Techniques du Languedoc, Place E. Bataillou, 34060 Montpellier Cedex, France.
‡ Permanent address: Division of Microbiology, National Institute for Medical Research, Mill Hill, London NW7 1AA, UK.

Abbreviations: ORF, open reading frame; RF, replicative form; SD sequence, Shine and Dalgarno sequence.
acid (Eastman Kodak). A solution of 0.5 M NaOH, 50 mM EDTA was then added dropwise to adjust the pH to 12.5–13. The mixture was stirred for 15 min then adjusted to pH 7-5 with a solution containing 2 volumes of NaCl (5 M), 1 volume of 2 M-Tris, pH 7.5, and 1 volume of HCl (1 M). Proteins and denatured chromosomal DNA were removed by extracting twice with phenol saturated with 0.5 M NaCl. The aqueous phase was then treated with 0.05 volume of ammonium acetate (2.5 M) and pH adjusted to 12.5–13. The mixture was stirred for 15 min then adjusted to pH 7-5 containing 0.1 M NaCl, 10 mM-Tris/HCl and 1 mM-EDTA and incubated for 30 min at 37 °C with 100 μg pancreatic RNAase A (Sigma), previously heat-shocked at 90 °C for 10 min. At this stage, the DNA preparation was almost free of chromosomal DNA but a further step of purification was performed by centrifugation in a density gradient (Haseltine et al., 1980) for pPP33 or in a velocity gradient for pHM2.

Restriction enzymes. All restriction enzymes were obtained from commercial suppliers and used under conditions provided by the supplier.

Ligation. Restricted M13 RF DNA (100 ng, 0.02 pmol) and a one- to fivefold molar excess of fragments to be cloned were mixed in 50 μl ligation buffer (50 mM-Tris/HCl, pH 7.8, containing 10 mM-MgCl₂, 50 μg BSA (nuclease-free) ml⁻¹, 1 mM-spermidine, 1 mM-rATP and 20 mM-DTT) with 1–2 units of ligase (Boehringer). The reaction was left at 15 °C for 6 to 12 h. A 2–5 μl volume of the ligation mixture was used to transform competent cells and the remainder was stored at −20 °C.

Transformation. Transformation of JM103 was performed using a modification of the low pH method (M13 Cloning Handbook, New England Biolabs): cultures were grown to an OD₆₀₀ value of 0–4–0.5, and then centrifuged at 7000 g for 5 min at 4 °C. The pellet was resuspended in 1/5 of the original volume in a solution (pH 5.6) containing M₅Cl₂ (50 mM), NaCl (5 mM) and sodium acetate (10 mM). Cells were left on ice for 20 min, then centrifuged as before. The pellet was resuspended in 1/6 of the original volume of a solution (pH 5.6) containing CaCl₂ (70 mM), M₅Cl₂ (100 mM), glycerol (5%, v/v) and sodium acetate (10 mM). Samples (100 μl) were mixed with the DNA in 1.5 ml Eppendorf tubes and allowed to stand on ice for 40 min. Samples were then heat-shocked at 42 °C for 3 min, mixed with 5 ml of TY top agar (Bactotryptone 20 g l⁻¹, yeast extract 10 g l⁻¹, NaCl 10 g l⁻¹, agar 16 g l⁻¹), and 200 μl fresh exponential culture of JM103 cells, 40 μl IPTG (100 mM) and 40 μl X-Gal (5-bromo-4-chloro-3-indolyl-β-D-galactoside; 2%, w/v) were added. The mixture was plated on to a previously warmed 2 × TY plate and incubated overnight at 37 °C.

In the case of transformation with plasmid pUC13, 1 ml 2 × TY medium was added to the cells after they had been heat-shocked. The culture was then incubated at 37 °C for 1 h to allow the expression of drug resistance. Samples were then plated on 2 × TY agar medium with ampicillin (100 μg ml⁻¹), IPTG and X-Gal (40 μg ml⁻¹ each).

Preparation of single-stranded M13 DNA. Colourless plaques were picked and added to 2 ml of an overnight culture of JM103 which had been diluted 100-fold with fresh medium. After 5 to 6 h incubation at 37 °C with vigorous shaking, 1.5 ml of the culture was transferred to a microfuge tube and phage DNA was purified as described by Sanger et al. (1980).

Screening of recombinants. Three methods were used: colony hybridization (Jeffreys & Flavell, 1977); Southern transfers of phage DNA from agarose gels (Southern, 1975); and T tracking (Sanger et al., 1980). In the case of hybridization, the probe used was checked for its ability to rescue spoIIA mutations in B. subtilis strains by transformation. The probe was labelled by nick translation (Rigby et al., 1977) using [α-³²P]dCTP (Amersham).

Subcloning from clones with large inserts. When two clones were shown to carry the same insert in opposite orientations, 1–2 μg amounts of the two single-stranded DNAs from each clone were mixed and allowed to hybridize for 2 h at 65 °C in the appropriate restriction buffer, after which only the insert was double-stranded. The mixture was then restricted and a portion was ligated to suitably cut RF M13 DNA and used to transform competent cells.

Shortening of inserts. The 2.2 kbp HindIII–SalI fragment of pHM2 was subcloned in pUC13. The new plasmid was linearized by digestion with HindIII, and shortened using a combination of exonuclease III and nuclease S1 (Guo et al., 1983). The shortened insert was removed from the vector by digestion with SalI, and recloned in pUC13, digested with SalI and Smal. DNA from clones of suitable size was subcloned in M13 mp11 for sequencing.

DNA sequencing. DNA sequencing was performed using the dideoxy chain termination method (Sanger et al., 1977). An improved definition in gels was obtained by the use of buffer gradient gels and nucleotides labelled with ³⁵S instead of ³²P (Biggin et al., 1983).

RESULTS AND DISCUSSION

Localization of the spoIIA locus

Two plasmids, pHM2 (Liu et al., 1982) and pPP33 (Piggot et al., 1984) contained between them the entire spoIIA locus, and were used to isolate the M13 clones employed for sequencing.
DNA sequence of B. subtilis spoIIA locus

Fig. 1. Partial restriction map of the spoIIA region showing directions of sequencing and positions of ORFs. The top of the Figure represents the partial restriction maps of plasmids pHM2 and pPP33. The arrows indicate the start points and the directions of sequencing; broken lines represent sequences obtained by the exonuclease III-nuclease S1 method (see Methods). Within the six possible reading frames the positions of ORFs encoding products larger than 5 kDa are indicated as white boxes.

Mutagenesis of pHM2 with Tn1000 (Lencastre et al., 1983) had shown that the locus began, at the most, 200 to 300 bp to the left of the EcoRI site (Fig. 1). The sequence reported here starts from this point and ends at the BclI site located 110 bp to the right of the PstI site (Fig. 1); no spoIIA mutations lie beyond this PstI site (D. Savva, personal communication; P. J. Piggot, unpublished results).

Determination of the sequence

As some information about the restriction map of the locus (Liu et al., 1982; Piggot et al., 1984) was available we used the method of 'forced cloning' in phages M13 mp8 and mp9 (Messing & Vieira, 1982). When the insert was too large to be read out, a subcloning step was performed as described in Methods. As there is a background of white clones resulting from chromosomal contamination and from the ability of the single-stranded DNA to transform competent cells, a further T-screening (Sanger et al., 1980) was necessary. The region on each side of the EcoRI site (Fig. 1) contains few other useful restriction sites so the sequence was determined by progressive shortening of the HindIII-BglII fragment (see Methods). The sequence presented in Fig. 2 was determined on both strands using the M13 clones shown in Fig. 1, except for two small regions, at positions 1230–1365 and 1607–1722 (Fig. 2), which were sequenced in one direction only; both were read from more than one gel using two different clones, and both gave unequivocal sequences.

Translational initiation sites

The sequence was analysed for open reading frames (ORFs) using the computer programme RUTRANMT (Staden, 1977). The largest ORFs are shown in Fig. 1. They all start with an ATG codon which is preceded by a ‘Shine and Dalgarno’ (SD) sequence (Shine & Dalgarno, 1974); short ORFs with the capacity to code for proteins of less than 5000 molecular weight are ignored. It has been reported that the interaction between the SD sequences and the 3' end of 16S
Fig. 2. Nucleotide sequence of the \textit{spolI}A locus. The sequence was determined from subclones of plasmids pHM2 and pPP33 in M13. Restriction sites used for subcloning are underlined and the putative Shine and Dalgarno sequences are doubly underlined. Regions of homology with known promoter sequences are indicated as follows: \textbullet, promoters recognized by \textit{Ea}3' (Johnson et al., 1983); \textdagger, promoters recognized by \textit{Ed}5 (Moran et al., 1982b). Amino acids are indicated by standard single-letter abbreviations.

rRNA is extensive in Gram-positive organisms (McLaughlin et al., 1981; Moran et al., 1982b), with calculated free energies ($\Delta G$) of base pairing between 16S RNA and its putative binding sites ranging from $-14$ to $-23$ kcal mol$^{-1}$. Our data, calculated according to the same rules (Tinoco et al., 1973) gave values of $-20.4$, $-16.6$, and $-12.8$ kcal mol$^{-1}$ for ORFs 1, 2 and 3, respectively (Table 1; 1 cal = 4.2 J).

The initiation codon, ATG, of ORF2 overlaps with the terminal TAG of ORF1 in the
Table 1. Nucleotide sequences of putative ribosome binding sites

<table>
<thead>
<tr>
<th>Protein encoded in:*</th>
<th>Nucleotide sequence†</th>
<th>Position</th>
<th>(\Delta G) (kcal mol(^{-1}))‡</th>
</tr>
</thead>
<tbody>
<tr>
<td>ORF1 AAGGAGGaaTGA</td>
<td>372</td>
<td>-20.4</td>
<td></td>
</tr>
<tr>
<td>ORF1 GGAGG</td>
<td>535</td>
<td>-14.4</td>
<td></td>
</tr>
<tr>
<td>ORF2 GGGGTGgcATC</td>
<td>724</td>
<td>-16.6</td>
<td></td>
</tr>
<tr>
<td>ORF3 AAGGAG</td>
<td>1175</td>
<td>-12.8</td>
<td></td>
</tr>
<tr>
<td>ORF3 GGAGG</td>
<td>1196</td>
<td>-14.4</td>
<td></td>
</tr>
</tbody>
</table>

† The nucleotides that are complementary to the 16S rRNA are shown in upper case. The position refers to the 16S rRNA 3’ terminus (McLaughlin et al., 1981).

‡ \(\Delta G\) is the free energy of base-pairing between the Shine and Dalgarno sequence (Shine & Dalgarno, 1974) and the 16S rRNA 3’ terminus (McLaughlin et al., 1981).

sequence ATGA. Similar overlaps of termination and initiation codons have been found in sequences of the E. coli trp operon (Oppenheim & Yanofsky, 1980) and of bacteriophage \(\lambda\) (Sanger et al., 1982). In each case of overlap in the trp operon, the pairs of gene products were made in equimolar amounts, and were sub-units of an enzyme complex; further, in the case of trpE and trpD, their translation was shown to be coupled (Oppenheim & Yanofsky, 1982). The overlap between ORF1 and ORF2 could indicate similar interactions, although there is no direct evidence as to whether this is the case.

**Gene products**

From the DNA sequence, three proteins, corresponding to ORF1, 2 and 3, are likely to be expressed. The location of the spoIIA69 complementing region defined by Tn1000 mutagenesis (Lencastre et al., 1983) corresponds very well to ORF1; a number of mutations lie within this complementing region (Chak et al., 1982; Errington & Mandelstam, 1983) and we anticipate that these are in ORF1. As yet no spo mutations are known to lie within ORF2. Various reasons can be put forward for this: it may be just chance; it may be that ORF2 is not expressed; it may be that ORF2 is expressed but is not involved in sporulation (at least not in the way that ORF1 and ORF3 are involved). The location of ORF3 (cloned in pPP33 but not pHM2) probably corresponds to the map position of a series of spoIIA mutations that are corrected by pPP33 but not pHM2. As one of these, spoIIAl, is suppressible by a nonsense suppressor (Yudkin & Turley, 1981), this would confirm that ORF3 is indeed expressed as a protein.

The amino acid sequences of the three proteins are listed in Fig. 2. The amino acid distribution shows no obvious deviation from the average based on 1048 proteins (Doolittle, 1981), except for the high proportion of hydrophobic residues in ORF1 and ORF2. However, this hydrophobicity is evenly distributed along the protein, and therefore does not indicate any special properties.

There is also an ORF in frame 6 capable of coding for a small protein (Fig. 1). However, this does not fit with the location of the spoIIA69 complementing region delimited by Tn1000, nor is it compatible with a polycistronic transcriptional unit. Its involvement in the sporulation process seems unlikely.

**Transcriptional signals**

The three main ORFs are on the same strand and, as there is no gap between them, they are probably transcribed as a single mRNA. We then looked for regions showing homologies with known promoters (Fig. 2 and Table 2). Their positions are rather odd, with respect to the localization of the ORFs. The first promoter is located about 260 bp upstream of ORF1: this length of 5’ non-translated sequence is very unusual. The efficient use of any of the other promoters would abolish the transcription, and therefore the expression of ORF1, which is not consistent with the Tn1000 insertion results. However, the list of promoters used in B. subtilis is not complete, and it is possible that the transcription of the spoIIA locus may require a special sigma factor.
Table 2. Putative promoter sequences

Nucleotides occurring in at least two sequences are indicated in upper case.

<table>
<thead>
<tr>
<th>-35</th>
<th>Spacer*</th>
<th>-10</th>
<th>Locus†</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAATC</td>
<td>17</td>
<td>TAgTGCTccgA</td>
<td>spoIIA</td>
</tr>
<tr>
<td>AAATC</td>
<td>15</td>
<td>TAtGCTTTTA</td>
<td>spoVG (P1)</td>
</tr>
<tr>
<td>AAATC</td>
<td>15</td>
<td>TAATGTTGTA</td>
<td>spoVC (P1)</td>
</tr>
<tr>
<td>AGGAaT</td>
<td>13</td>
<td>GAAATTGAc</td>
<td>spoIIA</td>
</tr>
<tr>
<td>AGGATT</td>
<td>14</td>
<td>GAAATTGAT</td>
<td>spoVG (P2)</td>
</tr>
<tr>
<td>AGGTTT</td>
<td>14</td>
<td>GGATTGTT</td>
<td>spoVC (P2)</td>
</tr>
<tr>
<td>TTGAIA</td>
<td>13</td>
<td>TAAAT</td>
<td>spoIIA</td>
</tr>
<tr>
<td>TTGACAC</td>
<td>17 or 18</td>
<td>TATAAT</td>
<td>Vegetative</td>
</tr>
</tbody>
</table>

* Number of nucleotides between the last base of -35′ sequence and the first base of -10′ sequence.
† The loci are described in the following references: spoIIA (this paper); spoVG (Moran et al., 1981; Johnson et al., 1983); spoVC (Moran et al., 1982a); vegetative (Moran et al., 1982b).

In the same way, none of the usual signals of transcriptional termination was found after the stop codon of ORF3. We know that the spoIIA and spoVA loci are adjacent (Piggot et al., 1983; Errington & Mandelstam, 1984; Savva & Mandelstam, 1984). The nearest spoVA mutation is at most 1.6 kb from the end of ORF3. These loci are known to be different transcriptional units (Piggot et al., 1984) and we are now sequencing the junction between spoIIA and spoVA to get further information.

We are grateful to J. Mandelstam, D. Savva and J. Errington for many helpful discussions, and to I. Jones and K. Jones for advice on the computer treatment of the data. This work was supported by the Science and Engineering Research Council and by Elf Aquitaine (UK).

REFERENCES


DNA sequence of B. subtilis spoIIA locus


