Nucleotide Sequence of the \textit{Bacillus subtilis} Developmental Gene \textit{SpoVE}

By U. D. BUGAICHUK* AND P. J. PIGGOT†

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

(Received 25 October 1985; 19 February 1986)

We have determined the nucleotide sequence of a 1159 bp DNA fragment containing the \textit{spoVE} locus of \textit{Bacillus subtilis}. The locus contained a single open reading frame of 293 codons. On the basis of the predicted amino acid sequence, the product of the \textit{spoVE} gene is believed to be a protein with an $M_r$ of 31 539. The amino-terminal portion of the \textit{spoVE} gene was used to construct a translational fusion with the \textit{lacZ'} gene. The hybrid \textit{spoVE–lacZ'} gene was shown to be expressed in \textit{Escherichia coli} and, therefore, it seems reasonable to conclude that the proposed open reading frame for the \textit{spoVE} gene does indeed function \textit{in vivo}.

\textbf{INTRODUCTION}

Bacterial endospore formation, triggered by nutrient limitation, is a primitive type of cellular differentiation. The study of this process in the Gram-positive bacterium \textit{Bacillus subtilis} is particularly amenable to analysis as this organism is probably the most extensively studied of the differentiating prokaryotes.

The completion of the forespore engulfment, resulting in a two-celled structure, marks a significant stage in the process of developmental change in \textit{B. subtilis}. Co-existence of the two genomes in such close proximity inevitably raises a number of questions about their participation in the subsequent biochemical and morphological events. Two major morphological changes that take place after compartmentalization are cortex formation, designated as stage IV, and deposition of the coat layers, designated as stage V. The existence of sporulation mutants which may be cortex+ coat− in one instance and cortex− coat+ in another, coupled with the results of epistasis experiments with double sporulation mutants, indicates the possibility that more than one dependent sequence of gene expression might be actively involved in the process of spore maturation at later stages (Coote & Mandelstam, 1973; Piggot & Coote, 1976).

There have been extensive studies of cloned sporulation-associated genes (Losick & Youngman, 1984; Piggot, 1985). Most of these are concerned with genes expressed early in sporulation. The \textit{spo} loci involved in the later stages of spore formation are, perhaps, of particular interest because of the complexity of the regulation of the developmental programme with two cell types involved, the mother cell and the forespore. Although most of the late loci are probably expressed only in the mother cell (Lencastre & Piggot, 1979; Dancer & Mandelstam, 1981), only one locus specific to the mother cell, \textit{spoVE}, has been cloned (Yamada et al., 1983; Piggot et al., 1986).

We report here the nucleotide sequence of the \textit{spoVE} locus and the proposed amino acid sequence for its product. We have established that this locus consists of a single gene 879 bp long. The promoter-proximal portion of this gene was used to construct a translational fusion to the \textit{lacZ'} gene. The hybrid \textit{spoVE–lacZ'} gene was expressed in \textit{Escherichia coli}, providing strong evidence that the proposed open reading frame for the \textit{spoVE} locus can be translated \textit{in vivo} and, presumably, is functionally active.

† Present address: Department of Microbiology and Immunology, Temple University School of Medicine, 3400 North Broad Street, Philadelphia, Pennsylvania 19140, USA.
Table 1. *Escherichia coli* strains

<table>
<thead>
<tr>
<th>E. coli strain</th>
<th>Genotype</th>
<th>Source/reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>JM103</td>
<td>supE sbcB15 strA thi endA Δ(lac-proAB) [F' traD36 proAB lacF' ZAM15]</td>
<td>Messing et al. (1981)</td>
</tr>
<tr>
<td>JM107</td>
<td>endA1 gyrA96 hsdR17 supE44 relA1 thi, λ- Δ(lac-proAB) [F' traD36 proAB lacF' ZAM15]</td>
<td>Yanisch-Perron et al. (1985)</td>
</tr>
</tbody>
</table>

**METHODS**

_Bacterial strains, phages and plasmids._ The *E. coli* strains used are listed in Table 1.

Plasmid pPP7 is a derivative of the bifunctional vector pHV33, containing a 4.7 kbp chromosomal fragment of _B. subtilis_ (Piggot et al., 1986).

Plasmids pUB3 and pUB4 were constructed by subcloning the 1.35 kbp _XhoI_ fragment from pPP7 into the _SalI_ site of pUC9 (Vieira & Messing, 1982). The orientation of the insertion was verified by multiple digests with the appropriate restriction enzymes. A physical map of plasmid pUB4 is shown in Fig. 3. Plasmid pUB3 contains the same _XhoI_ fragment in the reverse orientation. The construction of plasmid pUB5 is discussed in Results.

_Preparation of plasmid DNA._ Plasmid DNA was prepared as described by Birnboim & Doly (1979).

_Transformation._ *E. coli* strains were made competent and transformed by the method of Hanahan (1983). Recombinant M13 or pUC9 vectors were selected as described by Fort & Piggot (1984), except that ampicillin was used at a concentration of 50 μg ml⁻¹ for plasmid selection.

_Restriction enzyme analysis and isolation of specific DNA fragments._ All restriction enzymes were purchased from BRL, New England Biolabs or Pharmacia. DNA ligase was purchased from Boehringer. Enzymes were used as recommended by the suppliers. Mapping of restriction sites was done by multiple enzyme cutting and analysis of the digestion products by agarose (0.7–1.4%, w/v) gel electrophoresis using Tris/borate buffer (Maniatis et al., 1982). Specific DNA fragments for sub-cloning or sequencing were isolated on horizontal (0.5–1% w/v) low melting point agarose slabs (BRL) as described by the manufacturers.

_DNA sequencing._ Specific restriction fragments were sub-cloned into the appropriate M13 (mp10, mp11 or mp19) vector (Messing, 1983) and the nucleotide sequence was determined using the dideoxy chain termination method of Sanger _et al._ (1977). The sequencing kit was purchased from Amersham and was used in accordance with the supplier's recommendations.

**RESULTS**

The _spoVE_ locus has been shown to be contained within a 2.9 kbp _EcoRI_ fragment isolated by the 'prophage transformation' method in the temperature phage φ105 (Yamada _et al._, 1983). The independent isolation of this locus on a plasmid, pPP7 (Piggot _et al._, 1986), led us to the conclusion that the _spoVE_ locus is probably situated on a 1.35 kbp _XhoI_ fragment and that its minimum size is about 560 bp. In order to identify the product(s) of the _spoVE_ locus and to gain some information concerning the regulation of its expression we have determined the nucleotide sequence of a 1159 bp DNA fragment covering an area where all known _spoVE_ mutations were located. A detailed restriction map of the _spoVE_ locus and the sequencing strategy are presented in Fig. 1. Plasmids pPP7, pUB3 and pUB4 served as sources of specific DNA fragments for sub-cloning into appropriately cleaved M13 vectors. The areas around all the restriction sites used for sequencing were re-sequenced from a nearby restriction site in order to avoid potential loss of a sequence due to closely spaced restriction sites.

The nucleotide sequence of the 1159 bp chromosomal fragment is shown in Fig. 2. Analysis of the sequence has revealed that there is a single long open reading frame of 293 codons from which a protein (M, 31 539) can be translated. The open reading frame is preceded by a sequence which has a substantial degree of complementarity to the 3' end of the _B. subtilis_ 16S rRNA and is, therefore, a potential ribosome binding site (Moran _et al._, 1982).

With a view to obtaining experimental evidence that we have correctly identified the open reading frame for the _spoVE_ gene, we attempted to demonstrate that the N-terminus of this frame can function as a translational start for synthesis of the α-peptide of β-galactosidase (Messing _et al._, 1977). The hybrid plasmid pUB4 was chosen for subsequent intramolecular alterations aimed at creating a translational in-frame fusion of the _spoVE_ and _lacZα_ genes. A
DNA sequence of *B. subtilis* spoVE locus

Fig. 1. Physical map of the *B. subtilis* spoVE locus. The top line indicates the positions of restriction sites used in sequencing experiments; horizontal arrows represent the sequencing strategy. The position of the large open reading frame (ORF) is indicated by the open box above the restriction map. Av, AciI; Nc, NcoI; Nr, NruI; X, XhoI; H, HindIII. The four HindIII fragments are lettered A–C in correspondence to their size.

Fig. 2. Nucleotide and predicted amino acid sequence of the *B. subtilis* spoVE locus (only the non-transcribed strand is shown). The sequence was determined on both strands. The strong ribosome-binding site is doubly underlined. The predicted amino acid sequence is shown using standard three-letter abbreviations. Broken lines indicate the putative promoter sequences.
Fig. 3. Construction of spoVE-lacZ' translational gene fusion. The upper portion of the figure shows the structure of plasmid pUB4. The fragment of the sequence, shown on the right, represents the junction site and demonstrates the frame shift leading to translational failure. The proposed structure of plasmid pUB5 after excision of DNA between AvaI and SmaI (AvaI, XmaI) sites of pUB4 is shown in the lower portion of the figure. The sequence across the new junction site is presented on the right. The hatched boxes indicate the lacZ' gene. The stippled box represents the area downstream from spoVE translational start (TLS2), while the open box indicates the sequence upstream from its ribosome binding site. The direction of translation is from left to right in all cases. Av, AvaI; X, XhoI; S, SalI; Sm, SmaI; Nr, NruI; TLS1, translational start of the lacZ' gene.

medium supplemented with X-gal. We postulated that correlation of the frame shift and reduction in size of the spoVE gene portion could lead to at least partial restoration of the α-complementation function of a hybrid peptide (Fig. 3). The experimental approach to test this hypothesis was largely influenced by our discovery that the AvaI site within the spoVE gene and the AvaI (SmaI, XmaI) site in the multiple covalence site (MCS) area of the vector are in the same translational phase, so that after removal of the DNA fragment between these sites, rejoining of the remaining ends would create an in-frame junction between the spoVE and lacZ' genes.

Plasmid pUB4 was cut by AvaI and the 3.5 kbp fragment was isolated on low melting point agarose. The purified fragment was circularized by T4 DNA ligase and introduced into E. coli strain JM103. Altogether, 56-ampicillin resistant transformants appeared on a selective medium containing X-gal and 53 of them showed pale blue colouration. Ten pale blue clones that we analysed contained a plasmid of 3.5 kbp. This new plasmid was designated pUB5. The identity of the remaining B. subtilis chromosomal fragment in the plasmid was confirmed by the presence of a unique NruI site which is 80 bp upstream from the translational start point of the spoVE gene. We verified that this plasmid had the expected structure by subcloning the 0.8 kbp PstI–EcoRI fragment of plasmid pUB5 into M13 mp11 and determining the sequence of the junction site between the spoVE and lacZ' genes.
**DNA sequence of B. subtilis spoVE locus**

<table>
<thead>
<tr>
<th>Position†</th>
<th>Holoenzyme</th>
<th>‘−35’</th>
<th>‘−10’</th>
</tr>
</thead>
<tbody>
<tr>
<td>38−50</td>
<td>Ea28</td>
<td>CTAAA</td>
<td>CC GATAT†</td>
</tr>
<tr>
<td></td>
<td></td>
<td>⋆⋆⋆⋆⋆</td>
<td>⋆⋆ ⋆⋆⋆</td>
</tr>
<tr>
<td>49−79</td>
<td>Ea43</td>
<td>TTGACA</td>
<td>TATAAT</td>
</tr>
<tr>
<td></td>
<td></td>
<td>⋆⋆⋆ ⋆</td>
<td>⋆ ⋆ ⋆</td>
</tr>
<tr>
<td>125−155</td>
<td>Ea37</td>
<td>AGNNTT</td>
<td>GGNATTNTT</td>
</tr>
<tr>
<td></td>
<td></td>
<td>⋆ ⋆ ⋆</td>
<td>⋆ ⋆⋆</td>
</tr>
<tr>
<td>181−207</td>
<td>Ea32</td>
<td>AAATC</td>
<td>TANTGNNTNTA</td>
</tr>
<tr>
<td></td>
<td></td>
<td>⋆ ⋆ ⋆</td>
<td>⋆ ⋆ ⋆</td>
</tr>
</tbody>
</table>

Fig. 4. Regions showing homology to known B. subtilis promoters. Consensus sequences of the B. subtilis promoters controlled by minor forms of RNA polymerase were taken from Cowing et al. (1985). Nucleotides that match are designated with an asterisk (*). †, The position is given in correspondence to the sequence presented in Fig. 2. ‡, The one residue gap has been introduced in order to make homology more explicit.

**DISCUSSION**

Two of the spoV loci have been cloned so far: spoVG (Segall & Losick, 1977) and spoVA (Savva & Mandelstam, 1984). A substantial amount of data concerning transcriptional aspects of the spoVG locus is now available (Banner et al., 1983). Fort & Errington (1985) have determined the nucleotide sequence of the spoVA locus, which is the largest polycistronic sporulation operon yet characterized.

In this paper we present the nucleotide sequence of another stage V locus, spoVE, and the predicted amino acid sequence for its product (Fig. 2). The identification of this DNA fragment as the spoVE locus is unambiguous since all spoVE mutations tested map within this sequence (Piggot et al., 1986). The entire sequence of 1159 bp (Fig. 2) is based on sequencing of individual clones, as indicated in Fig. 1. Computer analysis has revealed one long open reading frame consisting of 293 codons. It starts with an ATG codon at position 240-242. Two nucleotides upstream there is a strong ribosome binding site, GGAGG, for which the free energy of base pairing (AG) with the 3' end of the B. subtilis 16S rRNA is calculated to be −14.4 kcal mol⁻¹ (−60.3 kJ mol⁻¹) using the rules of Tinoco et al. (1973). This AG value falls into a range published for Bacillus (Moran et al., 1982; Fort & Piggot, 1984). An interesting feature of this part of the sequence is the unusual closeness of the ribosome binding site to the translational start point.

Potential transcription initiation signals can be discerned from the determined sequence. Preceding the spoVE gene, within 207 bp upstream from the ribosome binding site, are four regions that have a substantial degree of homology to known B. subtilis promoters. The positions of these regions are indicated in Fig. 4. The sequences at positions 38−50, 125−155 and 181−207 display quite a close resemblance to promoters recognized by Ea28, Ea37 and Ea32 respectively. All of them appear to belong to the class of promoters utilized by minor forms of RNA polymerase holoenzyme. The sequence at position 49−79 exhibits substantial homology with the consensus sequence of promoters recognized by the major form of holoenzyme Ea43. Although the exact position(s) of the transcriptional start point(s) will require further investigation, the arrangement of σ37 and σ32 promoter sequences upstream from the translational start point of the spoVE gene is reminiscent of the regulatory region of the spoVG locus, which includes the same two types of promoters (Banner et al., 1983).

A great deal of information concerning various aspects of the expression of sporulation genes can be gained by fusing the gene in question to a second gene, whose product can be assayed
Fig. 5. Graphical display of the structural properties predicted for the *spoVE* gene product. (a) The distribution of charged residues along the amino acid sequence. (b) The hydrophatic character of the predicted protein (Kyte & Doolittle, 1982; R. Staden, personal communication). Both profiles were obtained with a segment of seven residues.

We have employed this technique in order to establish whether or not the open reading frame we propose for the *spoVE* locus is translated *in vivo*. Analysis of the *spoVE-lacZ* structure contained on plasmid pUB4 in terms of translational possibilities provides a convincing explanation for the inability of this plasmid to synthesize a functionally active α-peptide. On the basis of our sequencing data we conclude that disruption of α-peptide synthesis has a translational and not a transcriptional basis. Translation from the original *lacZ* start point (TLS1) comes to a halt shortly after passing the junction site (XhoI/SalI), resulting in a peptide containing 11 amino acids. Translation from the *spoVE* start point (TLS2) will not produce functional α-peptide because of the frame shift (Fig. 3a). In both cases synthesis of a hybrid protein containing α-peptide as a C-terminus appears not to be possible. However, when a portion of the DNA between the *AvaI* site in the *spoVE* gene and the *AvaI* (*SmaI*, *XmaI*) site in the MCS area is removed, the resulting plasmid, pUB5, is apparently able to produce at least partially active hybrid α-peptide, indicated by the appearance of pale blue colouration when plasmid-containing colonies are plated on agar supplemented with X-gal.

We conclude that this excision has created a new in-frame junction between the *spoVE* and *lacZ* genes (Fig. 3b). Therefore, while the translation from TLS1 remains impeded, translation from TLS2 can now lead to the synthesis of a functionally active hybrid α-peptide whose amino-terminus is substituted for the 118 amino acid stretch of the *spoVE* gene product. It seems reasonable, therefore, to conclude that the proposed open reading frame for the *spoVE* gene does indeed function *in vitro*.

The protein product predicted from the nucleotide sequence appears to be quite unusual and shows several features of interest. The ratio of acidic (glutamate, aspartate) or basic (lysine, arginine, histidine) residues is 0.55, indicating a rather basic protein with a calculated isoelectric point of 9.1. This protein also has a high content of hydrophobic residues, which is a
characteristic feature of membrane-bound proteins (Fig. 5). Moreover, it contains several stretches of 18–24 hydrophobic residues flanked by charged residues and this is strongly suggestive of membrane-spanning amino acid sequences (Eisenberg et al., 1984).

The availability of the entire nucleotide sequence of the spoVE locus should facilitate the rapid elucidation of those features that are important in the regulation of expression of this gene.

We are grateful to Professor J. Mandelstam and Dr M. D. Yudkin for critical reading of the manuscript and many helpful discussions. We thank Drs J. Errington, J. Gagnon, J. Rees and M. J. Geisow for help with computer analysis and advice. U. D. B. acknowledges the technical help of Mrs M. Deadman with sequencing experiments. This work was supported by grants from the Science and Engineering Research Council.

REFERENCES


