The Nucleotide Sequence and the Transcription During Sporulation of the gerE Gene of Bacillus subtilis

By SIMON CUTTING and JOEL MANDELSTAM*

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

(Received 11 March 1986; revised 22 May 1986)

We have determined the nucleotide sequence of a 1496 bp stretch of Bacillus subtilis chromosome that complements the gerE36 mutation. The sequence contains three open reading frames. One of these is part of the sdhC gene, the other two, whose functions are unknown, are capable of encoding proteins with $M_r$ values of approximately 17000 and 8500. The use of integrational plasmids to delimit the gerE transcriptional unit shows that the gerE locus consists of one gene, which encodes a polypeptide of 74 amino acid residues and which is switched on after $t_3$ during sporulation of wild-type B. subtilis 168.

INTRODUCTION

Strains of Bacillus subtilis carrying the gerE36 mutation produce spores that germinate poorly, have defective coats (as seen by electron microscopy) and are sensitive to lysozyme (Moir, 1981). Transduction experiments have shown that the gerE36 mutation is closely linked to the sdhC gene, and that the order of loci in this region of the B. subtilis chromosome is $sdhA(citF78)-sdhB(citF2)-sdhC(citF83)-gerE36$ (Moir, 1981; Hederstedt et al., 1982). Further work has shown that this mutation has other effects: the synthesis of an intracellular protease (protease e) at about stage $V$ is prevented, and there are reduced amounts of a coat polypeptide ($M_r$ 36000) which is synthesized at about $t_6$ (Jenkinson et al., 1981; Jenkinson & Lord, 1983; James & Mandelstam, 1985). The spore coats of gerE mutants lack a number of polypeptides that are normally present, but have four additional polypeptides (Jenkinson & Lord, 1983). Evidence has been presented (Jenkinson & Lord, 1983) that the gerE locus is in some way required for the synthesis of late sporulation proteins, including protease e and the '36K' spore-coat polypeptide. The gerE gene has recently been cloned by two groups of workers (James & Mandelstam, 1985; Hasnain et al., 1985), who have both found that gerE+ is dominant over gerE36.

In this paper we present the complete nucleotide sequence of the gerE locus and show that the gerE gene encodes a small protein. Transcription of the gene is detectable by Northern hybridization 4 h after the induction of sporulation.

METHODS

Bacterial strains, plasmids and phages. These are listed in Table 1.

General methods. Restriction endonucleases and other enzymes were obtained from Amersham or New England Biolabs, and were used as recommended by the suppliers. Replicative forms of M13mp10 and M13mp11 DNA were obtained from Amersham. General cloning techniques, preparation of plasmid DNA and B. subtilis chromosomal DNA, and ligation and recovery of DNA from agarose gels were done as described by Fort & Piggot (1984), Errington (1984) and Savva & Mandelstam (1984). Preparation of phage φ105 gerE and its DNA, and other general phage techniques were done as described previously (Jenkinson & Mandelstam, 1983; Errington, 1984).

Abbreviation: ORF, open reading frame.

0001-3328 © 1986 SGM
### Table 1. Bacterial strains, bacteriophages and plasmids

<table>
<thead>
<tr>
<th>Bacterial strain</th>
<th>Genotype</th>
<th>Source/reference/construction</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Escherichia coli</em> MC1061</td>
<td>araD139 Δ(ara leu)7697 Δlac X 74 galU galK hsr hsm Δ strA</td>
<td>Casadaban &amp; Cohen (1980)</td>
</tr>
<tr>
<td>JM103</td>
<td>Δ(lac-pro) thi strA supE endA sbcB15 F’ traD36 proAB lacPZAM15</td>
<td>Messing et al. (1981)</td>
</tr>
<tr>
<td><em>Bacillus subtilis</em> 168</td>
<td>trpC2</td>
<td>Laboratory stock</td>
</tr>
<tr>
<td>522</td>
<td>trpC2 gerE36</td>
<td>R. Sammons*</td>
</tr>
<tr>
<td>4709</td>
<td>leuC2 citF78</td>
<td>Laboratory stock</td>
</tr>
<tr>
<td>4700</td>
<td>trpC2 citF2</td>
<td></td>
</tr>
<tr>
<td>4822</td>
<td>leuC2 citF83</td>
<td></td>
</tr>
<tr>
<td>43.2</td>
<td>leu-8 spo4A43</td>
<td></td>
</tr>
</tbody>
</table>

**Bacteriophage**

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Source/reference/construction</th>
</tr>
</thead>
<tbody>
<tr>
<td>φ105gerE†</td>
<td>Cloning vector with unique BarnHI site</td>
</tr>
<tr>
<td>φ105J23</td>
<td>Chain terminating sequencing vectors</td>
</tr>
<tr>
<td>φ105J45, φ105J46, φ105J47</td>
<td>Cloning vector</td>
</tr>
</tbody>
</table>

**Stock plasmid**

<table>
<thead>
<tr>
<th>Plasmid</th>
<th>Relevant characteristics</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>pSGMU2</td>
<td>bla cat, 3.7 kbp</td>
<td>James &amp; Mandelstam (1985)</td>
</tr>
<tr>
<td>pUC19</td>
<td>bla, 2.7 kbp</td>
<td>J. Errington &amp; D. Jones (unpublished)</td>
</tr>
</tbody>
</table>

**Construction**

<table>
<thead>
<tr>
<th>Plasmid</th>
<th>Size (kbp)</th>
<th>Construction‡</th>
</tr>
</thead>
<tbody>
<tr>
<td>pSGMU101</td>
<td>4.2</td>
<td>pUC19, PsI, XbaI + (φ105gerE) 1.50 kbp PsI (1), XbaI (1497)</td>
</tr>
<tr>
<td>pSGMU102</td>
<td>4.1</td>
<td>pSGMU2, EcoRI, HindIII + (M13mp10) 0.39 kbp PsuI (897), HpaII (1289)</td>
</tr>
<tr>
<td>pSGMU103</td>
<td>3.9</td>
<td>pSGMU2, EcoRI, HindIII + (M13mp10) 0.21 kbp HpaII (1289), XbaI (1497)</td>
</tr>
<tr>
<td>pSGMU104</td>
<td>4.1</td>
<td>pSGMU2, EcoRI, HindIII + (M13mp11) 0.40 kbp HindIII (502), PsuI (897)</td>
</tr>
<tr>
<td>pSGMU105</td>
<td>4.2</td>
<td>pSGMU2, EcoRI, HindIII + (M13mp11) 0.49 kbp HindIII (502), EcoRV (988)</td>
</tr>
<tr>
<td>pSGMU106</td>
<td>4.2</td>
<td>pSGMU2, EcoRI, HindIII + (M13mp11) 0.51 kbp EcoRV (988), XbaI (1497)</td>
</tr>
<tr>
<td>pSGMU107</td>
<td>4.7</td>
<td>pSGMU2, EcoRI, HindIII + (M13mp10) 1.00 kbp HindIII (502), XbaI (1497)</td>
</tr>
<tr>
<td>pSGMU108</td>
<td>4.3</td>
<td>pSGMU2, EcoRI, HindIII + (M13mp11) 0.60 kbp PsuI (897), XbaI (1497)</td>
</tr>
<tr>
<td>pSGMU109</td>
<td>4.95</td>
<td>pSGMU2, EcoRI, HindIII + (M13mp10) 1.25 kbp NarI (250), XbaI (1497)</td>
</tr>
<tr>
<td>pSGMU110</td>
<td>2.95</td>
<td>pUC19, EcoRI, HindIII + (M13mp11) 0.25 kbp PsuI (1), NarI (250)</td>
</tr>
</tbody>
</table>

* R. Sammons, Department of Genetics, University of Birmingham, UK.
† The 1.5 kbp cloned DNA fragment in phage φ105gerE was in fact obtained from an MboI genomic library of *B. subtilis* 168 DNA cloned in the phage vector φ105J9, rather than from a BclI genomic library as stated previously (James & Mandelstam, 1985).
‡ The vector and the enzymes used to digest it are given first. After the '+' symbol, the phage in which the DNA insert was subcloned is given in parentheses. This is followed by the size of the inserted DNA fragment and the restriction endonucleases used to generate it. The numbers in parentheses refer to the positions of the cleavage sites in the nucleotide sequence in Fig. 2. The PsuI and XbaI sites were located in the polylinker of plasmid pSGMU101 adjacent to the MboI sites at the ends of the cloned *B. subtilis* DNA.

**DNA sequencing.** This was done by the dideoxy chain termination method (Sanger et al., 1977). Restriction fragments from plasmid pSGMU101 were isolated from low-melting-point agarose gels and subcloned into M13 phage sequencing vectors (Messing, 1983). Sequencing reactions were done according to Sanger et al. (1980), except that gradient buffer gels and 35S-labelled nucleotides were used (Biggin et al., 1983).
Genetic transformation of B. subtilis and selection of lysozyme-resistant spores. Chromosomal DNA was prepared as described by Errington (1984). Competent cells were produced by the method of Anagnostopoulos & Spizizen (1961), as modified by Jenkinson (1983). When transforming strains carried the citF mutation, CitF+ transformants were selected on minimal agar containing 0.5% (w/v) sodium lactate (Rutberg & Hoch, 1970); when they carried the gerE36 mutation, the transformation mixture was spread onto Irie’s sporulation agar (Irie et al., 1982) for the production of spores. After 4 d incubation at 37 °C, the spores were harvested in lysozyme buffer [0.054 M-potassium phosphate pH 5.8, 0.15 M-NaCl] and lysozyme (200 µg ml⁻¹). The suspension was diluted in lysozyme buffer to give an OD₆₀₀ of approximately 3-0, incubated at 42 °C for 60 min and then at 80 °C for 20 min to destroy vegetative cells and lysozyme-sensitive spores. Lysozyme-resistant c.f.u. were estimated by spreading serial dilutions on nutrient agar plates and incubating them at 37 °C for 12-16 h. Chloramphenicol-resistant transformants were selected on Oxoid nutrient agar containing chloramphenicol (5 µg ml⁻¹).

Growth and sporulation. B. subtilis cultures of wild-type or sporulation mutants were induced to sporulate by the method of Sterlini & Mandelstam (1969).

Isolation of mRNA. Cultures (100 ml) were removed at various times after resuspension in sporulation medium and RNA was isolated from the cells as described by Savva & Mandelstam (1985).

Construction of recombinant plasmids. (a) Plasmid pSGMU101. DNA (10 µg) from phage φ105gerE was digested to completion with PstI and XhoI. The 1.5 kbp DNA fragment was isolated from a low-melting-point agarose gel, ligated between the PstI and XhoI sites of pUC19, and used to transform Escherichia coli MC1061 (Casadaban & Cohen, 1980).

(b) Integrational plasmids. Replicative forms of the M13mp10 and M13mp11 subclones used for nucleotide sequencing were digested with HindIII and EcoRI (see Table 1). The DNA fragment to be cloned was isolated from a low-melting-point agarose gel, ligated to plasmid pSGMU2 (Fort & Errington, 1985) and used to transform E. coli strain JM103.

(c) Plasmid pSGMU110. This was constructed as described under (b) but plasmid pUC19 was used instead of pSGMU2.

Transformation of strains of E. coli. E. coli strains JM103 and MC1061 were transformed as described by Fort & Piggot (1984) and by Casadaban & Cohen (1980) respectively.

Screening for recombinants. (a) E. coli JM103. Transformed cells containing plasmids pUC19 or pSGMU2, which carry the ampicillin-resistance gene bla, were selected after transformation on nutrient agar plates containing ampicillin (75 µg ml⁻¹). M13 phage transformants were isolated by plating the mixture as an overlay on nutrient agar plates as described by Fort & Piggot (1984). Insertion of DNA into any of the restriction sites in a polynucleotide region in the lacZ coding region of M13 or pUC vectors results in loss of β-galactosidase (lacZ) activity, which can be detected using the chromogenic substrate 5-bromo-4-chloro-3-indolyl β-d-galactopyranoside (X-gal) (Messing, 1983). Recombinant phage or plasmids appear as colourless plaques or as white colonies respectively.

(b) E. coli MC1061. This strain does not possess the F' episome needed for complementing activity of the lacZ gene (Messing, 1983): recombinant plasmids were detected in this strain by colony hybridization (see below).

Screening cells for the GerE phenotype. Colonies were picked on to sporulation agar plates (Difco nutrient agar, 2.3%, w/v; thymine, 0.001%, w/v; manganese sulphate, 0.001%, w/v) (Irie et al., 1982) and incubated at 37 °C for 4 d. The colonies were transferred to a Whatman no. 1 filter paper which was placed, colony side uppermost, on a ‘heat shock’ agar plate (Bacto-agar, 30% w/v, Sigma). These plates were incubated for 1 h at 37 °C after which GerE+ colonies turned pink, while GerE- colonies remained colourless.

DNA–RNA and DNA–DNA hybridizations. (a) Northern hybridizations. Samples of RNA (8 µg) were electrophoresed through 2% agarose gels containing formaldehyde, as described by Maniatis et al. (1982). The RNA was transferred to nitrocellulose filters as described by Savva & Mandelstam (1984). Probe DNA (200 ng, see Results) was labelled by nick translation (Rigby et al., 1977) with [α-³²P]dCTP (3000 Ci mmol⁻¹, 110 TBq mmol⁻¹; Amersham). RNA–DNA hybridizations were done as described by Savva & Mandelstam (1985).

(b) Colony hybridizations. These were done by the method of Grunstein & Hogness (1975) as modified by Maniatis et al. (1982).

(c) Southern hybridizations. These were done as described by Southern (1975).

Construction of recombinant phage vectors. The phage vectors φ105J45, φ105J46 and φ105J47 were constructed as follows. The appropriate integrational plasmids (see Results) were linearized by digestion with BamHI and ligated to BamHI-digested phage φ105123 (J. Errington & D. Jones, unpublished results) in a 50 µl reaction volume (16 h at 15 °C). This ligation mixture was used to transfect protoplasts of B. subtilis CU267 (non-lysogenic) as described by Errington (1984). Recombinant phage particles were recovered (Errington, 1984) and used to transduce strain 522 to chloramphenicol-resistance.
RESULTS

Construction of a plasmid carrying the gerE region

A fragment of *B. subtilis* 168 DNA that complements in trans the gerE36 mutation has recently been cloned in a derivative of the temperate bacteriophage φ105 (James & Mandelstam, 1985). Preliminary restriction mapping of the phage DNA showed that the whole of the cloned insert (1.5 kbp) could be excised by double digestion with the restriction enzymes PstI and XbaI. Repeated attempts to clone this fragment in plasmid pUC19 using strain JM103 as recipient failed, and it was concluded that the 1.5 kbp *B. subtilis* DNA insert probably contained an EcoK restriction endonuclease site.

Fort & Errington (1985) have reported a similar problem in cloning a gene encoding chloramphenicol resistance in *E. coli* strain JM103. Accordingly, a restrictionless strain of *E. coli*, MC1061, was transformed with the same ligation mixture. Transformed cells containing a plasmid were selected for ampicillin resistance. Strain MC1061 of *E. coli* is not suitable for the detection of recombinant pUC plasmids in the usual manner, i.e. by insertional inactivation of lacZ α-complementing activity (see Messing, 1983). Therefore, the ampicillin-resistant colonies were transferred to a nitrocellulose filter and screened by colony hybridization with a radioactively labelled probe of φ105gerE DNA: of 88 colonies tested, 68 were found to hybridize to the probe. Plasmid DNA was prepared from one of these and characterized by restriction endonuclease mapping. This plasmid, pSGMU101, when digested with PstI and XbaI, yielded a 1.5 kbp fragment. Restriction sites for EcoRI, PvuII, HindIII, AvaII and NarI were found; their positions are shown in Fig. 1.

The recombinant plasmid was examined to see if it still contained the gerE gene. Earlier work (Ohné et al., 1973; Moir, 1981) has shown that the gerE locus is closely linked to the citF operon (recently designated sdh for succinate dehydrogenase), and that, of the three sdh genes (sdhA, sdhB and sdhC) sdhC is the most closely linked to gerE. Transformation experiments in this laboratory (results not shown) indicate that gerE and sdhC are about 80% linked, and that gerE and sdhB are about 65% linked.

*B. subtilis* strain 522.1 carrying both the gerE36 mutation and the sdhC (citF83) mutation was transformed to CitF+ with DNA (1 pg) from phage φ105gerE or plasmid pSGMU101. In both crosses, about 75% of the CitF+ transformants obtained were also Ger+ (394/500 for pSGMU101 and 360/500 for φ105gerE). This showed that the 1.5 kbp cloned fragment in phage φ105gerE and plasmid pSGMU101 contained at least part of the sdhC and gerE genes. Neither of the other two sdh mutants [sdhA (citF78) and sdhB (citF2)] were transformed to prototrophy by phage or plasmid DNA.

Southern blot hybridizations were used to ensure that no chromosomal deletion or rearrangement had occurred in the cloning process. *B. subtilis* 168 DNA was restricted with enzymes BglII, EcoRV, PvuII and HindIII, or combinations of these, and the digestion products were examined using a radioactively labelled probe of plasmid pSGMU101 DNA. The radioactive bands produced were compared with those predicted from the restriction map of the cloned insert in pSGMU101. The results (not shown) indicated that the cloned fragment had not originated from a BclI digest of 168 DNA as previously reported (James & Mandelstam, 1985). Nucleotide sequence analysis (see below) showed that, in fact, it arose from a partial MboI digestion which was done at the same time. No chromosomal rearrangement or deletion was detected. Single and double digestions of chromosomal DNA with HindIII and PvuII gave the positions of two further restriction sites, shown in Fig. 1, adjacent to the left and right hand ends of the cloned insert. The position of the HindIII site adjacent to the left hand end of the cloned insert is in agreement with that found by Hasnain et al. (1985).

Determination of the nucleotide sequence

The source of DNA for sequencing was the 1.5 kbp cloned DNA fragment contained in plasmid pSGMU101. Restriction fragments from this plasmid (see Fig. 1) were isolated and subcloned into either phage M13mp10 or M13mp11. The complete sequence of 1496 bp was determined on both strands, and each sequence determination was obtained from at least two independent M13 clones. The complete sequence together with the predicted amino acid
sequence is shown in Fig. 2. Also shown are the restriction sites used for generating M13 sequencing clones and an EcoK restriction site that was presumably responsible for the initial failure to clone this fragment of DNA in E. coli strain JM103.

Localization of gerE and sdhC mutations

Plasmids containing different fragments of B. subtilis DNA (Table 1 and Fig. 3) were used to transform strain 522 and selection for gerE+ was made by treatment with lysozyme (see Methods). The results from a typical experiment using these plasmids are shown in Table 2. These transformations showed that the gerE36 mutation lies between the EcoRV and HpaII sites of the cloned insert (Fig. 1). This conclusion is consistent with that obtained by analysis with integrational plasmids (see below).

Four recombinant plasmids, pSGMU101, pSGMU107, pSGMU109 and pSGMU110, were used to transform strain 4822 citF83 leuC2 to CitF+ on minimal-lactate plates. The results of these transformations (Table 2) showed that the sdhC (citF83) mutation lay not more than 250 bp from the left of the cloned DNA as shown in Fig. 1. This is in agreement with a previous study (Hasnain et al., 1985) and suggests that open reading frame 1 (ORF1) is part of the sdhC gene at the distal end of the sdh operon.

Use of integrational plasmids to define the gerE transcriptional unit

The nucleotide sequence indicated three possible ORFs, as shown in Fig. 1. It was necessary to determine which of these was the gerE gene and also whether the cloned insert that had been sequenced contained the entire gerE locus. This was done with the plasmids constructed by subcloning fragments from the M13 sequencing clones into the integrational plasmid pSGMU2 (Fort & Errington, 1985).

Plasmid pSGMU2 containing a fragment of the 1.5 kbp cloned insert can integrate into the chromosome of B. subtilis 168 by a Campbell-like mechanism and thus produce duplication of the homologous DNA. If, for example, the plasmid contained homologous DNA the ends of which were completely within the functional (transcriptional) unit of the gerE operon, then duplication of this homologous DNA in the recipient's chromosome would disrupt the operon and give a Ger- phenotype. Conversely, if the homologous DNA contained either end or both ends, integration and duplication of homologous DNA would give an undisrupted gerE gene.
The plasmids containing different sectors of the 1.5 kbp clone were therefore used to define the transcriptional unit of the \textit{gerE} locus.

\textit{B. subtilis} strain 168 (\textit{gerE}+) was transformed with DNA (1 \(\mu\)g) prepared from each of the constructed integrational plasmids, pSGMU102, pSGMU103, pSGMU104, pSGMU105, pSGMU106, pSGMU107, pSGMU108 and pSGMU109. Transformants containing integrated plasmids were selected on the basis of chloramphenicol resistance (encoded by the plasmid vector), and screened for their Ger phenotype by the tetrazolium method (see Methods). In all crosses the phenotype of the chloramphenicol-resistant transformants was Ger+. [Integration of pSGMU plasmids by a Campbell-like mechanism has been shown by Fort \\& Errington (1985) and so we did not attempt to confirm this by Southern hybridizations.] To explain these results we must presume that the \textit{gerE} locus is a monocistronic operon containing one, small, gene. Thus, plasmid pSGMU102, which contained a 390 bp cloned DNA fragment spanning ORF3, gave a Ger+ phenotype. This fragment must therefore contain one or both ends of the \textit{gerE}
**Transcriptional unit.** Plasmid pSGMU103, which contained a small cloned DNA fragment of 210 bp stretching from the HpaII site at the end of ORF3 to the end of the cloned insert, also gave a Ger+ phenotype, so we concluded that the gerE functional unit did not extend beyond the end (3' end) of the DNA fragment cloned in phage φ105gerE and in plasmid pSGMU101. If the cloned fragment had contained only part of a gerE polycistronic operon extending downstream from ORF3, then plasmid pSGMU103 would have given a GerE− phenotype after integration into the *B. subtilis* chromosome.

To determine the approximate position of the gerE promoter, three recombinant phage vectors were constructed using phage φ105J23 as described in Methods. Phages φ105J45, φ105J46 and φ105J47 contained the integrational plasmids pSGMU106, pSGMU108 and pSGMU107 respectively (see Fig. 3). Lysates of these phages were used to transduce strain 522 carrying the gerE36 mutation to chloramphenicol resistance. Transductants were picked onto selective media and scored for their GerE phenotype by the tetrazolium method (see Methods).
Table 3. Analysis of the GerE phenotype

Phage lysates were used to transduce strain 522 carrying the gerE36 mutation to chloramphenicol resistance, and the Ger phenotype was determined by the tetrazolium method (see Methods). Phage lysates were also used to transduce strain 522, with direct selection for lysozyme resistance. Typically, strain 522 gave approximately $10^2$ lysozyme-resistant survivors ml$^{-1}$.

<table>
<thead>
<tr>
<th>Phage</th>
<th>Cloned insert*</th>
<th>Ger phenotype</th>
<th>No. of lysozyme-resistant survivors ml$^{-1}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>φ105J45</td>
<td>EcoRV–MboI</td>
<td>-</td>
<td>$9.3 \times 10^2$</td>
</tr>
<tr>
<td>φ105J46</td>
<td>PvuII–MboI</td>
<td>+</td>
<td>$3.8 \times 10^9$</td>
</tr>
<tr>
<td>φ105J47</td>
<td>HindIII–MboI</td>
<td>+</td>
<td>$5.8 \times 10^9$</td>
</tr>
</tbody>
</table>

* The B. subtilis cloned DNA insert contained in these phages is shown by the restriction sites at the left hand and right hand ends of the insert (see Fig. 1) respectively.

Lysates from each of these phages were also incubated with strain 522, and the lysozyme resistance of the resulting lysogens was determined by plating them on non-selective media as described in Methods. The results of these experiments are shown in Table 3. Phage φ105J47, incorporating a 1.0 kbp segment of B. subtilis cloned DNA spanning both ORF2 and ORF3, and also φ105J46, which contained a 0.6 kbp fragment downstream from the PvuII site (see Fig. 1), complemented the GerE phenotype. However, φ105J45, which contained a 0.51 kbp fragment downstream from the EcoRV site and which therefore spanned ORF3, did not complement the GerE phenotype.

DNA from these phages was digested with BamHI to check that they contained the correct cloned DNA insert. In the construction of these phage vectors (see Methods) the integrational plasmids pSGMU106, pSGMU108 and pSGMU107 were inserted into the unique BamHI site of φ105J23 to construct the vectors φ105J45, φ105J46 and φ105J47 respectively. Therefore, digestion of the phage DNA of these vectors would liberate the insert to give a restriction fragment of 4-21 kbp, 4-3 kbp and 4-7 kbp for the vectors φ105J45, φ105J46 and φ105J47 respectively. Since complementation of the GerE phenotype could result only if the phage vector contained an entire gerE gene and its promoter, the promoter for the gerE gene must lie downstream from the PvuII site (Fig. 1) and after ORF2, but before the EcoRV site. Restriction analysis of DNA from these phages showed that the orientations of the cloned inserts were the same (results not shown). The possibility that there is transcription from a phage promoter can be ruled out (J. Errington, unpublished results).

Use of Northern hybridization to determine the size and approximate time of transcription of gerE

From the predicted size of gerE, one would expect the appearance during sporulation of a transcript of about 0.2 to 0.4 kb in size and possessing the expected hybridization properties. To test this, mRNA was isolated from sporulating cells (100 ml) of B. subtilis 168 at hourly intervals and passed through an agarose gel (see Methods) containing formaldehyde. After electrophoresis, the RNA was transferred to a nitrocellulose filter and tested for hybridization to one of three radioactively labelled probes. These probes were made from plasmids pSGMU102, pSGMU108 and pSGMU107. With each of these probes, radioactive bands became detectable by $t_4$ and were present till $t_6$. The approximate size of the RNA transcript was determined by using markers of pSGMU107 and pSGMU102 DNA which had been cleaved with HindIII and EcoRI before electrophoresis. Thus, pSGMU107 produces a band of 3-7 kbp and another of 0-9 kbp and pSGMU102 bands of 3-7 and 0-39 kbp which hybridize to the probe DNA. The size of the transcript was estimated to be about 0.3 kb. Since hybridization was obtained with the probe DNA made from plasmid pSGMU102, containing only the ORF3 region, it appears that ORF3 must be the gerE gene. The results of a typical experiment with a labelled probe of pSGMU102 are shown in Fig. 4. As a control, RNA isolated at $t_0$, $t_2$, $t_4$ and $t_6$ from a mutant (strain 43.2) blocked at stage 0 of sporulation was electrophoresed and hybridized to the same labelled probes; as expected, mRNA hybridization was not detected. Since plasmids pSGMU107 and
Nucleotide sequence of the gerE gene

pSGMU108 both contain cloned DNA extending into the ORF2 gene, it follows that if ORF2 had encoded an additional protein coded for by the gerE locus we would have expected to see a transcript of at least 0.6 kb in size. It therefore seems that the gerE locus contains a single gene represented by ORF3. ORF2 is not expressed during sporulation and its function, if any, during vegetative growth remains unknown.

Analysis of gene products and regulatory signals

The nucleotide sequence of 1496 bp (Fig. 2) was analysed using the ANALYSEQ program of Staden (1984). Three ORFs were predicted from the sequence, all translated from left to right as shown in Fig. 1. The first of these extended beyond the left end of the cloned region. Using the codon preference method of Staden & McLachlan (1982) and a standard file of codon preferences for a gene known to encode a sporulation protein (the spoVAD gene; Fort & Errington, 1985), each of the three ORFs was found to have a high probability of coding for a protein.

The nucleotide sequence was searched for putative ribosome binding sites (Shine & Dalgarno, 1974). These are listed in Table 4, together with the free energies (ΔG) of base-pairing between the putative ribosome binding sites and the 3' terminus of B. subtilis 16S rRNA (McLaughlin et al., 1981) calculated by the method of Tinoco et al. (1973). The putative ribosome binding site preceding ORF3 is well placed to act as a translational initiation site and would be expected to form a stable complex with the 16S rRNA. There are two possible ribosome binding sites preceding ORF2 and either could form a stable complex with 16S rRNA. (Another ribosome binding site at position 284 could form a translational initiation site since it precedes an ATG start codon.)

Both ORF1 and ORF3 were followed by regions of dyad symmetry (Fig. 2). Their resemblance to E. coli and B. subtilis transcription terminators (Rosenberg & Court, 1979;
Table 4. Putative ribosome binding sites

<table>
<thead>
<tr>
<th>Position*</th>
<th>Sequence†</th>
<th>Spacer</th>
<th>ΔG (kcal mol⁻¹)‡</th>
<th>Start signal</th>
</tr>
</thead>
<tbody>
<tr>
<td>284</td>
<td>AAGGcGG</td>
<td>9</td>
<td>-16-4</td>
<td>TTG</td>
</tr>
<tr>
<td>474</td>
<td>AAGGAGaTGA</td>
<td>7</td>
<td>-13-6</td>
<td>ATG</td>
</tr>
<tr>
<td>475</td>
<td>AAGGAGAT</td>
<td>8</td>
<td>-14-2</td>
<td>ATG</td>
</tr>
<tr>
<td>1029</td>
<td>AGcAAGGAGG</td>
<td>10</td>
<td>-20-2</td>
<td>TTG</td>
</tr>
</tbody>
</table>

16S rRNA: UCUUUCUCCACUAG

* No. of the first nucleotide in the given sequence, with reference to Fig. 2.
† The nucleotides that are complementary to the 16S rRNA are shown in upper case.
‡ ΔG 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). Values were determined as described by Tinoco et al. (1973). 1 kcal mol⁻¹ = 4.18 kJ mol⁻¹.

Table 5. Properties of putative translation products

<table>
<thead>
<tr>
<th>Position of triplet*</th>
<th>Start</th>
<th>Stop</th>
<th>No. of amino acids</th>
<th>Mr</th>
<th>Hydrophobicity†</th>
</tr>
</thead>
<tbody>
<tr>
<td>ORF2</td>
<td>490</td>
<td>930</td>
<td>147</td>
<td>17174</td>
<td>+7.23</td>
</tr>
<tr>
<td>ORF3</td>
<td>1049</td>
<td>1270</td>
<td>74</td>
<td>8544</td>
<td>-8.74</td>
</tr>
</tbody>
</table>

* First nucleotide of start codon and last nucleotide before stop codon.
† Determined using the ANALYSEQ program of Studen (1984).

Brennan & Geiduschek, 1983) and their position near the 3’ termini of these ORFs suggest that they are likely to act as transcriptional stop signals for the ORF1 and ORF3 genes. The calculated thermal stability of the predicted RNA ‘hairpin’ structures that these terminators might form were calculated for ORF1 and ORF3 as -11.8 kcal mol⁻¹ (-49.4 kJ mol⁻¹) and -19.8 kcal mol⁻¹ (-82.8 kJ mol⁻¹) respectively (Tinoco et al., 1973).

Some of the properties of the proteins predicted by the sequences of ORF2 and ORF3 are shown in Table 5. ORF2, whose function is unknown, could code for a protein of 147 amino acids. The amino acid sequence was compared with those of other proteins but no significant homologies were found. ORF3, which seems to be the gerE gene, encodes a protein of 74 amino acids. The amino acid composition of the ORF3 protein shows a high proportion of charged amino acids, with no cysteine.

The upstream regions of ORF2 and ORF3 were searched for possible homologies to known B. subtilis promoter sequences. Upstream of ORF2 there are several sequences showing partial homology with recognized B. subtilis sequences. Between positions 931 and 1021 upstream of ORF3 partial homologies with sequences recognized by B. subtilis σ²⁸ (Gilman et al., 1981) were seen. This would be within the region possibly encoding the regulatory region of the gerE gene.

DISCUSSION

The gerE locus has been characterized extensively (Moir, 1981; Jenkinson & Lord, 1983; James & Mandelstam, 1985); a mutation in it causes a defect in the germination response of the spore in the presence of L-alanine. The mutant spore is also sensitive to lysozyme, and its spore coat is altered both in composition and structure (Jenkinson & Lord, 1983). Lastly, the spore lacks a specific sporulation-associated protease, protease e (M, 30000; James & Mandelstam, 1985), which is synthesized at stage V in sporation.

The nucleotide sequence of a fragment of B. subtilis chromosomal DNA, presented in this paper, which complements the gerE36 mutation shows that the gerE locus encodes a single protein. The gene starts with a TTG triplet and codes for a 74 amino acid polypeptide.
gerE36 mutation is complemented by a fragment of DNA covering this ORF. Integrational plasmids were used to delimit the gerE mRNA; this mRNA was analysed by Northern hybridization and shown to be about 300 bases in size, which agreed with the size of the transcript predicted from the sequence. As far as we know, the gerE protein is the smallest B. subtilis polypeptide found to date.

With regard to its formation, Jenkinson & Lord (1983) suggested that the gerE protein was probably regulatory, on the basis of its pleiotropic effects, namely, the absence of several coat polypeptides including the 36K polypeptide synthesized at stage VI and required for lysozyme resistance (Jenkinson, 1981) and also protease e.

Another possibility is that the gerE product could be a coat polypeptide; the M, of the gerE product (8500) is similar to that of one of the coat polypeptides absent from the gerE mutant. However, as Jenkinson & Lord (1983) postulated, it is then difficult to account for the other phenotypic effects of gerE. A more likely explanation is that gerE is a regulatory gene that is switched on late in the course of sporulation and that is responsible for the appearance of several protein components of the spore, including protease e. It is consistent with this interpretation that preliminary computer studies indicate that the product of the gerE gene has close structural resemblances to DNA binding proteins such as the lexA repressor and the lambda repressor.

We are indebted to Dr J. Errington for many helpful discussions. This work was supported by the Science and Engineering Research Council and by Unilever (UK). S. C. was the holder of an SERC-CASE studentship.

REFERENCES


