Structure and Function in a *Bacillus subtilis* Sporulation-specific Sigma Factor: Molecular Nature of Mutations in *spoIIAC*

By M. D. YUDKIN

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

(Received 7 August 1986; revised 5 October 1986)

The *spoIIAC* gene was cloned from chromosomal DNA of seven *spoIIAC* mutants of *Bacillus subtilis*, and the complete sequence of the gene was determined for each mutant. Three of the mutants proved to have chain-terminating mutations (one of which, previously shown to be suppressible by *sup-3*, was identified as *amber*); these led in every case to complete failure either to manufacture spores or to synthesize two enzymes normally associated with stage II of sporulation. The four remaining mutations were missense, and these corresponded to a phenotype in which a few spores are formed and about half the wild-type quantities of the two enzymes are made. Of the four missense mutations, two were near the promoter-distal end of the gene, in a region believed to correspond to the DNA-binding domain of the sigma factor that *spoIIAC* encodes. The remaining two mutations were in the region of the gene that is thought to correspond to the domain of the protein that interacts with core RNA polymerase.

INTRODUCTION

*spoIIA* is one of the earliest sporulation-specific loci to be expressed after cells of *Bacillus subtilis* are transferred to a medium that induces sporulation (Savva & Mandelstam, 1985). Determination of the sequence of the locus (Fort & Piggot, 1984) showed that it consists of three genes, *spoIIAA*, *spoIIB* and *spoIIAC*. The last of these encodes a protein whose predicted sequence has striking homologies with that of the product of another sporulation gene, *spoIIG*, and also with sigma factors of both *B. subtilis* and *Escherichia coli* (Errington et al., 1985; Stragier et al., 1985; Stragier, 1986). Given the role of sigma factors in determining the specificity of transcription by RNA polymerase, it seems plausible to suppose that the *spoIIAC* gene product directs the transcription, at a certain point during sporulation, of loci whose products are needed for the developing spore. [Such a role for sporulation-specific sigma factors in promoting the transcription of particular genes has for several years been emphasized by R. Losick; see, for example, Losick (1981).]

Mutations in *spoIIA* can give rise to a variety of phenotypes (Errington & Mandelstam, 1983). Some render strains asporogenous (altogether incapable of making spores), others oligosporogenous (making fewer spores than the wild-type). Some prevent synthesis of two enzymes – alkaline phosphatase and DNAase – that normally appear about 2 h after cells are transferred to sporulation medium; others allow about half the wild-type quantities of these enzymes to be made.

I have started a programme of identifying the mutations that give rise to these various phenotypes, as a contribution to studying the structure and function of the *spoIIA* polypeptides. An earlier paper reported the identification of three mutations in the *spoIIAA* gene (Yudkin et al., 1985). I now describe the nature of seven mutations in *spoIIAC*.

METHODS

Bacterial strains. These are listed in Table 1.

Preparation of DNA. Chromosomal DNA from *B. subtilis* was prepared by the method of Errington (1984),
modified by including an additional extraction with phenol before the treatment with a phenol/chloroform mixture. Plasmid was prepared from *E. coli* by the method of Birnboim & Doly (1979).

Restriction endonuclease digestion. Enzymes were obtained from Amersham, Boehringer or BRL and used as recommended by the suppliers.

Ligation. The ligation buffer contained (final concentrations) 66 mM-Tris/HCl pH 7.5, 5 mM-MgCl₂, 5 mM-DTT and 1 mM-ATP. The reaction volume was 10 μl, and the mixture was incubated at 4 °C for 5–6 h for ligating cohesive ends (in the presence of 0.1–0.2 unit T4 DNA ligase) or for 20–24 h for ligating blunt ends (in the presence of 1 unit ligase).

Transformation of *E. coli*. Cells were made competent by treatment with CaCl₂ (Maniatis et al., 1982) and used within 2 h.

Recovery of DNA from agarose gels. Low-melting-point agarose was used and the DNA was extracted as described by Savva & Mandelstam (1984).

Cloning of the wild-type *spoIIA* region. This has been described by Yudkin (1986).

Cloning of the *spoIIA* region from *spoIIAC* mutants. Chromosomal DNA was digested to completion with *BglII* and *PstI* and subjected to electrophoresis on low-melting-point agarose alongside a 0.96 kb marker obtained by digestion of the wild-type *spoIIA* clone (Yudkin, 1986). DNA corresponding in size to the marker was extracted and ligated into plasmid pUC19 (Yanisch-Perron et al., 1985), and the mixture was transformed into *E. coli* strain JM103. Selection on nutrient agar containing ampicillin, isopropyl-β-D-thiogalactoside and 5-bromo-4-chloro-3-indolyl-β-D-galactoside yielded clones of which the large majority were white and therefore apparently contained inserts. These were probed with 1–6 kb *spoIIA* DNA which had been purified from the wild-type clone (Yudkin, 1986) and labelled with [32P]dCTP with a nick-translation kit from Amersham. Preparation of filters, hybridization and autoradiography were as described by Savva & Mandelstam (1984). About 1–3% of the clones hybridized with the probe. These were picked and purified, and small-scale plasmid preparations were made from them and checked by restriction analysis for the presence of the *spoIIAC* region. Two independent clones were kept from each *spo* mutant and were carried through the remainder of the procedure. The plasmids were transferred to *E. coli* strain GM48 (which is unable to methylate DNA at the *BclI* site); plasmid was prepared from 100 ml cultures, purified by centrifugation through CsCl (Maniatis et al., 1982), precipitated twice and finally stored at 4 °C.

DNA sequencing. DNA fragments were obtained by restriction of the purified plasmids, subcloned into M13mp18 or M13mp19 (Yanisch-Perron et al., 1985), and sequenced with a kit from Amersham. The whole *spoIIAC* gene was sequenced (for each of the two clones from each mutant) by reading from the *ClaI* site towards the end of the gene, and from the *PstI* site and either the *BclI* or the *HaeIII* site towards the beginning of the gene (see Fig. 1). In each mutant a single nucleotide change was found within the gene; the identification was confirmed by reading the sequence from a different restriction site, as described in detail in Results.

RESULTS

Wild-type *spoIIA* sequence

In a Note Added in Proof, Errington et al. (1985) made two corrections to the sequence of the *spoIIA* gene published by Fort & Piggot (1984). By reading the wild-type *spoIIA* sequence from the *PstI* site I have confirmed the presence of the two additional nucleotides mentioned by Errington et al. (1985): these are, on the non-transcribed strand, an additional T after residue

---

# Table 1. Bacterial strains

<table>
<thead>
<tr>
<th>Strain</th>
<th>Genotype</th>
<th>Reference or source</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>E. coli</em></td>
<td></td>
<td></td>
</tr>
<tr>
<td>JM103</td>
<td><em>supE</em> <em>thi</em> <em>endA</em> Δ(<em>lac–pro</em>) <em>sbcB15</em> <em>hsdR4/F</em> <em>traD36</em> <em>proAB</em></td>
<td>Messing et al. (1981)</td>
</tr>
<tr>
<td>GM48</td>
<td><em>supE</em> <em>thi</em> <em>thr</em> <em>leu</em> <em>lacY</em> <em>galK</em> <em>galT</em> <em>ara</em> <em>tonA</em> <em>txs</em> <em>dam</em> <em>dcm</em></td>
<td>Backman (1980)</td>
</tr>
<tr>
<td><em>B. subtilis</em></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2000.01</td>
<td><em>lys−1</em> <em>pyrD1</em> <em>rif</em> <em>spoIIA1</em></td>
<td>Yudkin &amp; Turley (1981)</td>
</tr>
<tr>
<td>2000.63</td>
<td><em>lys−1</em> <em>pyrD1</em> <em>rif</em> <em>spoIIA63</em></td>
<td>Yudkin &amp; Turley (1981)</td>
</tr>
<tr>
<td>560</td>
<td><em>metC3</em> <em>spoIIA560</em></td>
<td>Errington &amp; Mandelstam (1983)</td>
</tr>
<tr>
<td>561</td>
<td><em>metC3</em> <em>spoIIA561</em></td>
<td>Errington &amp; Mandelstam (1983)</td>
</tr>
<tr>
<td>563</td>
<td><em>metC3</em> <em>spoIIA563</em></td>
<td>Errington &amp; Mandelstam (1983)</td>
</tr>
<tr>
<td>564</td>
<td><em>metC3</em> <em>spoIIA564</em></td>
<td>Errington &amp; Mandelstam (1983)</td>
</tr>
<tr>
<td>578</td>
<td><em>metC3</em> <em>spoIIA578</em></td>
<td>J. Errington (unpublished)</td>
</tr>
</tbody>
</table>
Molecular identification of spoIIA C mutations

Fig. 1. Restriction map of the spoIIAC gene. The numbering is that of Fort & Piggot (1984).

1722 and an additional A after residue 1880. (The non-transcribed strand is that strand of the DNA that has the same sequence, T being substituted for U, as that of the mRNA.) The corrected sequence of the ClaI–PstI fragment of spoIIA and the sequence of the predicted spoIIAC polypeptide are given in Fig. 2. The remainder of this paper uses the corrected numbering.

**Mutant sequences**

By cloning spoIIAC from each mutant and determining its sequence, I found in every case one and only one change from the wild-type sequence. The same change was found in both of the two independent clones prepared from each mutant, and was confirmed by sequencing the opposite strand of DNA in the region in which the mutation occurred. All the mutations turned out to be G-C to A-T transitions.

**spo-I.** In the sequence of the transcribed strand read from the PstI site, the change G to A was apparent at position 1920. The corresponding C to T change in the non-transcribed strand was found by reading from the BclI site towards the PstI site.

**spo-63.** The mutation C to T in the non-transcribed strand was found at nucleotide 1266 in the sequences read both from the ClaI site and from the PvuII site. In the transcribed strand the sequence read from the Rsal site showed G to A at the same position.

**spo-560.** The transcribed strand read from the PstI site showed a C to T change at position 1897. The non-transcribed strand had G to A at this position, easily seen by sequencing from the BclI site.

**spo-561.** Reading from the PstI site showed the mutation C to T in the transcribed strand at position 1884, which was confirmed as G to A on the sequence read from the BclI site.

**spo-563.** This DNA was sequenced both from the ClaI site and from the PvuII site, and contained a C to T substitution on the non-transcribed strand at nucleotide 1405. Confirmation was obtained by finding the complementary G to A change on the sequence read from the HaeIII site.

**spo-564.** The transcribed strand sequenced from the HaeIII site showed the mutation C to T at position 1519. Confirmation of the complementary change was obtained by running a long sequencing gel from the ClaI site.

**spo-578.** Sequencing the transcribed strand from the BclI site towards the ClaI site showed the mutation G to A at nucleotide 1482. The non-transcribed strand read from the ClaI site had C to T at the same position.

**DISCUSSION**

The introduction of two additional nucleotides into the sequence published by Fort & Piggot (1984) changes the reading frame and consequently extends the predicted polypeptide to 255 amino-acid residues. As pointed out by Stragier (1986), this extension greatly increases the homology between the spoIIAC product and the sigma proteins of both *E. coli* and *B. subtilis*, inasmuch as it shows that the spoIIAC protein is strikingly homologous with the other sigma factors near its C-terminus as well as at an internal region. Stragier (1986) has, moreover, called attention to the fact that the C-terminal region of the spoIIAC product has the characteristic helix-turn-helix motif that is a feature of DNA-binding proteins (reviewed by Pabo & Sauer,
Fig. 2. Nucleotide sequence of the spoIIA region between the Clal and the PstI sites shown in Fig. 1, with the predicted amino-acid sequence (in one-letter code) of the polypeptide encoded by the spoIAC gene.
Molecular identification of spoIIAC mutations

Table 2. Identity of spoIIAC mutations

<table>
<thead>
<tr>
<th>Isolation no. of mutation</th>
<th>Spore incidence*</th>
<th>Alkaline phosphatase and DNAase†</th>
<th>Wild-type amino-acid residue</th>
<th>Mutant residue</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>&lt;10^-8</td>
<td>&lt;5</td>
<td>Gln-245</td>
<td>amber</td>
</tr>
<tr>
<td>63</td>
<td>&lt;10^-8</td>
<td>&lt;5</td>
<td>Gln-27</td>
<td>ochre</td>
</tr>
<tr>
<td>560</td>
<td>10^-3</td>
<td>~50</td>
<td>Arg-237</td>
<td>Lys</td>
</tr>
<tr>
<td>561</td>
<td>10^-6</td>
<td>~50</td>
<td>Val-233</td>
<td>Met</td>
</tr>
<tr>
<td>563</td>
<td>10^-6</td>
<td>~50</td>
<td>Ser-73</td>
<td>Phe</td>
</tr>
<tr>
<td>564</td>
<td>10^-6</td>
<td>~50</td>
<td>Arg-111</td>
<td>Gln</td>
</tr>
<tr>
<td>578</td>
<td>&lt;10^-8</td>
<td>&lt;5</td>
<td>Arg-99</td>
<td>TGA</td>
</tr>
</tbody>
</table>

* This was determined by Errington & Mandelstam (1983), except for the incidence for strain 578, which was determined during the present work. The wild-type is taken as 1.
† These enzymes were measured during sporulation by Errington & Mandelstam (1983), except for the activity in strain 578, which was determined during the present work. The wild type is taken as 100.

1984), implying that the C-terminal region represents the domain of this sigma factor that selects promoters to be transcribed (Gitt et al., 1985) and the internal region the domain that interacts with the core RNA polymerase.

How far do the results described in this paper support these proposals? Table 2 lists the mutations that I have identified at the nucleotide level, the changes in amino-acid residue to which they give rise, and the phenotype of the corresponding mutants. The three mutations spo-l, spo-63 and spo-578 are all chain-terminating. The corresponding phenotypes are the most pleiotropic of all the mutations in spoIIAC: spores are not made in these mutants, and no synthesis of the sporulation-associated enzymes alkaline phosphatase and DNAase is detectable. Clearly the absence of protein of wild-type length leads to a completely negative phenotype. Even spo-l-carrying mutants, which lack only 11 residues from the C-terminus, are unable to carry out any of the sporulation functions that normally require the presence of the spoIIAC product: it is probably relevant that of the 11 missing amino-acid residues, the two immediately distal to the point of mutation are highly conserved among sigma proteins (see Fig. 3a). [It follows incidentally from the identification of spo-l that sup-3, which suppresses spo-l (Yudkin & Turley, 1981), is an amber suppressor.]

spo-560 and spo-561 are both located in the putative DNA-binding domain of the protein at residues 237 and 233, just distal to the glycine residue that forms the centre of the helix-turn-helix motif (see Fig. 3a). In the wild-type spoIIAC protein, each of these residues differs from that present at the corresponding position in the major sigma protein found in vegetative cells (line 3 in Fig. 3a). It is tempting to suggest that the mutations affect residues that interact with the promoter(s) that the spoIIAC product helps to transcribe; this temptation is strengthened by the facts that mutations in corresponding positions in the trp repressor of E. coli affect the repressor's binding to DNA (Kelley & Yanofsky, 1985), and that the corresponding residues in both the cl repressor and the cro product of phage λ interact with specific base pairs in their target operators (Ptashne, 1986). Moreover, studies of two other DNA-binding proteins from E. coli suggest that the position corresponding to residue 233 of the spoIIAC product is functionally important: the only known missense mutation in the sigma factor encoded by htpR occurs here (Yura et al., 1984), and changes in the corresponding residue of the catabolite gene activator alter its sequence specificity for the lac promoter (Ebright et al., 1984). The changes imposed by spo-560 and spo-561 are remarkably conservative: spo-560 changes one positively-charged residue to another and spo-561 one hydrophobic residue to another. The fact that such apparently trivial changes have so dramatic an effect on the incidence of sporulation (see Table 2) tends to support the view that residues 233 and 237 help to determine the specificity of interaction between the spoIIAC product and its target promoter(s).

At first sight the results with spo-563 and spo-564 speak against this interpretation. These two mutations are located in a quite different region of the gene from spo-561, and yet the sporulation
Fig. 3. The amino-acid sequence of parts of the product of the htpR gene of E. coli, the rpoD gene of E. coli, the rpoD gene of B. subtilis, the spoIIG gene of B. subtilis and the spoIIAC gene of B. subtilis. The asterisks mark positions where identical or similar residues are present in four or five proteins. The numbering refers to the amino-acid sequence of the spoIIAC product. The alignments are due to Stragier (1986). (a) the C-terminal sequence; (b) an internal region corresponding to the putative binding domain for the core polymerase.

phenotype to which they give rise is apparently identical to that of the spo-561 mutant. However, if the spoIIAC product is indeed a sigma factor necessary for the transcription of subsequent spo gene(s), then it is reasonable to suppose that missense mutations in it might disrupt sporulation whether they occur in the domain that interacts with DNA or in the domain that interacts with the core RNA polymerase: in both types of mutant the sigma factor would be unable to fulfil its task of ensuring specific spo transcription. Thus the facts that spo-563 occurs within the highly conserved region that is presumed to represent the site of binding of sigma proteins to the core polymerase (Stragier et al., 1985; Stragier, 1986), and that spo-564 is on the edge of the same conserved region (see Fig. 3b), suggest an obvious interpretation for these two mutations. Indeed, the disruptive effect of spo-563 is particularly easy to understand, since it substitutes the bulky hydrophobic residue Phe for the small polar residue Ser at a position where the only other residue known to be permitted is Ala (Fig. 3b). The precise effect of spo-564 is less obvious.

The view that spo-560 and spo-561 can properly be distinguished from spo-563 and spo-564, despite their similarity of sporulation phenotype, finds support from the fact that in one respect the phenotype to which they give rise is quite different. I have previously reported that the wild-type spoIIAC product is toxic to E. coli (Yudkin, 1986). D. Harrison in this laboratory has now shown that both spo-563 and spo-564 abolish this toxicity, whereas spo-561 affects it only slightly and spo-560 not at all (unpublished results). These findings lend credence to the suggestion that spo-563 and spo-564 prevent the product of spoIIAC from binding to the core RNA polymerase of E. coli.
Molecular identification of spoIAC mutations

Even the most conservative of the amino-acid changes described here, spo-560, reduces the incidence of sporulation by a factor of $10^3$, whereas even the most radical of them, spo-563, reduces the synthesis of alkaline phosphatase and DNAase by only a factor of 2. This unresponsiveness of the synthesis of alkaline phosphatase and DNAase to missense mutations in spoIAC stands in contrast to the fact that all nonsense mutations in the gene render these enzymes undetectable. It may well be that the spoIAC product is involved in the synthesis of the enzymes not directly but at several removes, and that additional factors operate to regulate their synthesis.

I am grateful to Dr P. Butler, Dr J. Errington and Professor J. Mandelstam for much useful advice, and to Mr Alex Coaker for conscientious technical assistance.

REFERENCES


