JOURNAL OF BACTERIOLOGY, May 1991, p. 2915–2919
0021-9193/91/092915-05$02.00/0

Vol. 173, No. 9

Gene Encoding Two Alkali-Soluble Components of the Spore Coat from *Bacillus subtilis*

SIMON CUTTING, LIANGBIAO ZHENG, AND RICHARD LOSICK*

Department of Cellular and Developmental Biology, The Biological Laboratories, Harvard University, Cambridge, Massachusetts 02138

Received 26 November 1990/Accepted 28 February 1991

We report the cloning and characterization of a gene called *cotF* from *Bacillus subtilis* that encodes alkali-soluble polypeptides of 5 and 8 kDa that are components of the spore coat. The 5- and 8-kDa polypeptides are generated by proteolytic cleavage of the primary product of the *cotF* gene, which is 160 codons in length and is capable of encoding a polypeptide of 19 kDa. Amino acid sequence analysis indicates that the 5-kDa species is derived from the NH_2-terminal portion of the primary gene product and that the 8-kDa species is derived from the COOH-terminal portion. A mutant bearing an in vitro-constructed *cotF* null mutation produced normal-looking spores that contained an apparently complete set of coat proteins except for the absence of the 5- and 8-kDa polypeptides. The map position of *cotF* is 349°. Transcription of *cotF* commenced coincidently (during h 6 of sporulation) with genes known to be under the control of sporulation transcription factor σ^K.

Endospores of the gram-positive bacterium *Bacillus subtilis* are encased in a tough protein shell known as the coat that protects the spore from lysozyme and facilitates the capacity of the spore to germinate (1, 6). The coat is composed of a dozen or more different proteins, which are organized into an electron-dense outer layer and a lamella-like inner layer. So far, genes for six coat proteins have been identified: *cotA*, *-B*, *-C*, *-D* (4), *-E* (12), and *-T* (2), which encode polypeptides of 65, 59, 10, 9, 24, and 8 kDa, respectively. Transcription of *cot* genes is regulated by a complex hierarchical cascade consisting of four coordinately controlled sets (regulons) of coat protein and transcriptional regulatory genes, which are induced successively at intermediate to late stages of sporulation (13). Interestingly, with the exception of CotE, which plays a key role in coat morphogenesis (12), the absence of individual coat proteins as a result of *cot* gene null mutations causes little or no impairment of the resistance and germination properties of the spore (4). Here we report the identification of an additional coat protein gene, which we call *cotF*. The predicted primary product of *cotF* is a 19-kDa polypeptide, which is processed to generate alkali-soluble coat proteins of 5 and 8 kDa. Transcription of *cotF* is induced coincidently with genes under the control of sporulation transcription factor σ^K. On the basis of this and other considerations, *cotF* is tentatively assigned to the regulon of σ^K-controlled genes, whose members include coat protein genes *cotA* and *cotD* and the regulatory gene *gerE* (3, 7, 13).

MATERIALS AND METHODS

Methods for the preparation and electrophoretic separation of spore coat proteins, for the use of synthetic oligonucleotides as hybridization probes, for DNA sequencing, and for primer extension analysis were as previously described (4, 8, 10, 13).

Nucleotide sequence accession number. The GenBank accession number for the *cotF* sequence is M58592.

* Corresponding author.

RESULTS AND DISCUSSION

Gene encoding an alkali-soluble spore coat polypeptide of 5 kDa. Two procedures for the extraction of coat proteins from intact spores are treatment with alkali and treatment with sodium dodecyl sulfate (SDS) and dithiothreitol, procedures which preferentially solubilize different components of the spore coat (4, 6). We were interested in identifying the structural gene for an alkali-soluble coat polypeptide of approximately 5 kDa, which, because of its solubility properties and size, was anticipated to be distinct from coat proteins whose structural genes had previously been cloned. To clone the gene encoding the 5-kDa polypeptide, coat proteins were extracted from intact, purified spores with NaOH and purified by SDS-polyacrylamide gel electrophoresis. The purified 5-kDa protein was eluted from the gel and subjected to sequential Edman degradation in the Harvard Microchemistry Facility. The partial NH_2-terminal amino acid sequence is shown by the thick underlining in the third row of Fig. 1. The amino acid sequence was used to design a synthetic oligonucleotide of 44 nucleotides in length. The 44-mer was, in turn, used as a hybridization probe to identify the structural gene for the 5-kDa polypeptide in a library of size-fractionated, *Hin*dIII-*Eco*RI fragments of *B. subtilis* DNA cloned in *Escherichia coli* as previously described (4). The presence of the 5-kDa protein-coding sequence in a *Hin*dIII-*Eco*RI fragment of 2.3 kb (Fig. 2) was confirmed by nucleotide sequencing, which revealed an open reading frame whose deduced amino acid sequence from codons 5 through 20 conformed exactly to the partial NH_2-terminal amino acid sequence determined for the 5-kDa protein (Fig. 1). The open reading frame was capable of encoding a polypeptide of 160 residues, whose predicted molecular weight (18,725) was substantially larger than that of the purified coat protein. We designate the putative coat protein-coding sequence *cotF*. The primary product (CotF) of the *cotF* gene is expected to be markedly hydrophilic and to have an isoelectric point of 7.9. Using standard computer homology searches, we were unable to identify significant similarities between the predicted CotF amino acid sequence

```
                                                    Pr3
                                    GTGCATGTTCTCAACCCGTATCCTATATAAGGTGAATTTCCCTTTTTAAA
                                                      P      Sau3A ClaI
AACTTCAAAACATCATGAACAACATTGAGCGTTGGGCATATGCTGATATGGAATCTGATCGATAAAGGAGAATTTTCAC

                          Pr1                                                 Pr2
ATG GAT GAA CGC AGA ACA TTG GCT TGG CAT GAA ACA TTA GAA ATG CAC GAG CTG GTT GCT
Met Asp Glu Arg Arg Thr Leu Ala Trp His Glu Thr Leu Glu Met His Glu Leu Val Ala

                                                                        Pr4
TTT CAA TCA AAC GGA CTC ATT AAA CTG AAG AAA ATG ATA AGA GAA GTA AAA GAC CCT CAG
Phe Gln Ser Asn Gly Leu Ile Lys Leu Lys Lys Met Ile Arg Glu Val Lys Asp Pro Gln

CTC AGA CAG CTT TAT AAC GTG TCT ATT CAG GGT GTT GAG CAA AAT TTG AGA GAG CTT CTT
Leu Arg Gln Leu Tyr Asn Val Ser Ile Gln Gly Val Glu Gln Asn Leu Arg Glu Leu Leu

CCG TTC TTT CCA CAG GCT CCG CAC AGA GAG GAT GAG GAA GAA GAA CGC GCA GAT AAC CCA
Pro Phe Phe Pro Gln Ala Pro His Arg Glu Asp Glu Glu Glu Glu Arg Ala Asp Asn Pro

                  BstEII
TTT TAC AGC GGT GAC CTG CTC GGT TTT GCC AAA ACA TCT GTC CGC AGC TAT GCC ATC GCA
Phe Tyr Ser Gly Asp Leu Leu Gly Phe Ala Lys Thr Ser Val Arg Ser Tyr Ala Ile Ala

ATT ACA GAA ACA GCA ACA CCT CAA TTA AGA AAC GTA CTG GTC AAA CAG CTG AAT GCT GCC
Ile Thr Glu Thr Ala Thr Pro Gln Leu Arg Asn Val Leu Val Lys Gln Leu Asn Ala Ala

                                  ClaI
ATC CAG CTG CAC GCC CAA GTT TAT CGA TAC ATG TAT CAG CAT GGA TAT TAT CCG TCT TAC
Ile Gln Leu His Ala Gln Val Tyr Arg Tyr Met Tyr Gln His Gly Tyr Tyr Pro Ser Tyr

AAC CTT TCT GAA CTG TTG AAA AAC GAT GTC AGA AAC GCC AAC AGA GCC ATT TCA ATG AAA
Asn Leu Ser Glu Leu Leu Lys Asn Asp Val Arg Asn Ala Asn Arg Ala Ile Ser Met Lys

                                                  Sau3A
TAA AACACAAGGCAAAGGAGTTCCCTTTGCTGCTATTTCTGCTGATC
***
```

FIG. 1. Nucleotide sequence of the nontranscribed strand of *cotF*. The nucleotide sequence of both DNA strands was determined as described in the legend to Fig. 2. The NH_2-terminal sequences of the 5- and 8-kDa coat polypeptides as determined by sequential Edman degradations are shown by the thick underlinings in rows 3 and 6, respectively. The putative ribosome binding site is shown in boldface. The double underlining identifies the nucleotide sequence of oligonucleotide Pr3. The single underlining identifies the nucleotide sequences complementary to the oligonucleotides Pr1, Pr2, and Pr4. The 5′ terminus of *cotF* mRNA is shown by an arrow. The vertical arrowheads identify sites of proteolytic cleavage.

and that of other protein sequences in the Genbank/EMBL and NBRF data bases.

Absence of 5- and 8-kDa polypeptides in the coat of spores from a *cotF* null mutant. To verify that *cotF* specifies the 5-kDa polypeptide, we constructed a deletion-insertion mutation in the cloned gene in vitro, which we then used to replace the wild-type gene in the chromosome of strain PY17 as described previously for *cot* genes *A* to *E* (4, 12). Figure 2 shows the boundaries of the deleted DNA (representing more than 75% of the *cotF* open reading frame), which was replaced with a cassette (11) bearing a chloramphenicol resistance gene (*cat*), to create a *cotF*Δ::*cat* deletion-insertion mutant. The structure of the chromosome-borne mutation in the *cotF*Δ::*cat* mutant was confirmed by Southern hybridization analysis (data not shown). As with *cotA*, *-B*, and *-C* mutants (4), the *cotF*Δ::*cat* mutant produced normal-looking (phase-bright) spores that were resistant to treatment with heat, chloroform, and lysozyme and that germinated normally (as judged by their ability to reduce the dye tetrazolium in the presence of a mixture of germinants).

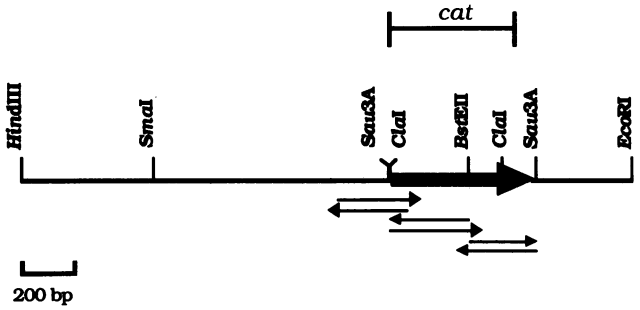

FIG. 2. Endonuclease restriction map of the cloned *Hin*dIII-*Eco*RI fragment containing the *cotF* gene. The figure is a complete map of all indicated restriction sites except *Sau*3A, for which only two sites are shown. The thick arrows indicate the orientation and length of the *cotF* open reading frame. The *Cla*I-*Cla*I fragment that was replaced with the *cat* cassette to construct *cotF*Δ::*cat* is identified above the restriction map. Thin arrows indicate the regions of nucleotide sequence determination and the DNA strand sequenced. The sequence was determined by the dideoxy-chain termination method (8), using as templates phage M13 clones bearing *Sau*3A-*Sau*3A and *Bst*EII-*Sau*3A segments of subcloned DNA (see restriction sites identified in the figure) and a phage M13 clone bearing the entire 2.3-kb *Hin*dIII-*Eco*RI fragment of the figure and the oligonucleotide primers Pr2 and Pr3 depicted in Fig. 1.

We next investigated the effect of the *cotF*Δ::*cat* mutation on the pattern of alkali-soluble coat polypeptides by extracting mutant spores with NaOH. The polyacrylamide gel of Fig. 3 shows that the mutant spores were specifically lacking coat polypeptides of 5 and 8 kDa. When coat polypeptides were solubilized from *cotF* mutant spores by treatment with SDS and dithiothreitol instead of with NaOH, the pattern of coat polypeptides was indistinguishable from that of the wild-type strain, PY17 (data not shown).

These results confirm the assignment of the 5-kDa polypeptide to *cotF* and indicate that the 5-kDa species arises by proteolysis of the predicted primary product (19 kDa) of *cotF*. Because the 5-kDa protein is contained within the NH_2-terminal portion of CotF, we infer that cleavage occurs at a site corresponding to the COOH terminus of the 5-kDa species. We also infer the existence of a second cleavage at a site corresponding to the NH_2 terminus of the 5-kDa

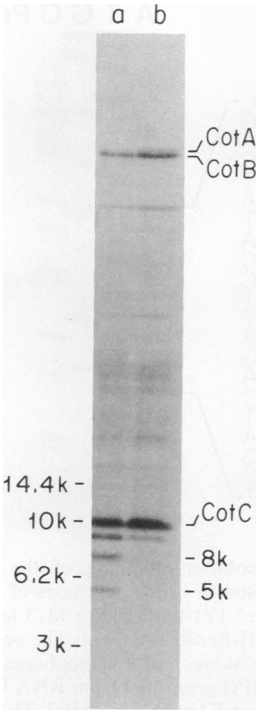

FIG. 3. SDS-polyacrylamide gel electrophoresis of alkali-soluble coat polypeptides from mutant and wild-type spores. Spores of a *cotF*Δ::*cat* mutant (lane b) and its isogenic wild-type parent (PY17; lane a) were purified as described previously (6). Coat proteins were extracted with 0.1 M NaOH at 0°C for 15 min (4) and subjected to electrophoresis in an SDS-high-Tris polyacrylamide gel containing 18% acrylamide as described by Thomas and Kornberg (10). The sizes of selected proteins are indicated in kilodaltons (k).

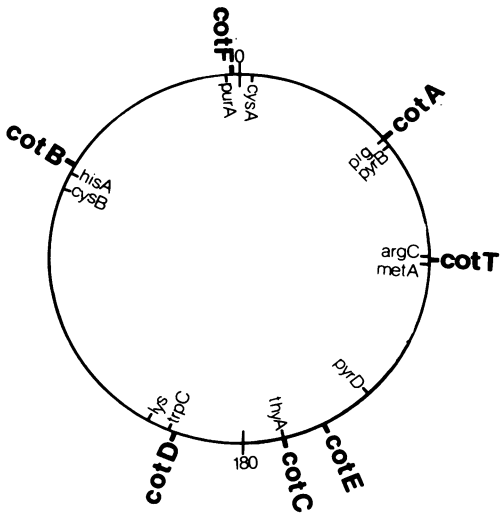

FIG. 4. Location of spore coat protein genes on the *B. subtilis* chromosome. The map positions are taken from the following references: *cotA* to *-D*, Donovan et al. (4); *cotE*, Zheng et al. (12); *cotT*, Aronson et al. (2); and *cotF*, this study.

protein because the NH_2-terminal amino acid of the 5-kDa polypeptide is located five residues downstream from the NH_2 terminus of the predicted primary product of *cotF*.

The discovery that *cotF*Δ::*cat* mutant spores also lack a 8-kDa coat polypeptide suggested that the 8-kDa species is an additional cleavage product of the primary product of *cotF*. To investigate this possibility, we purified the 8-kDa protein and subjected it to sequential Edman degradation. The resulting partial NH_2-terminal amino acid sequence (Ala Asp Asn Pro Phe; as indicated by the thick underlining in the sixth row of Fig. 1) confirmed the assignment of the 8-kDa species to the *cotF* coding sequence and indicated that it is derived from the COOH-terminal region of the *cotF* primary product.

Because of the uncertainty in determining the molecular weights of small polypeptides on the basis of their electrophoretic mobility, our estimation of the sizes of the 5- and 8-kDa polypeptides could be inaccurate. This uncertainty allows us to propose that the 5- and 8-kDa polypeptides are generated by as few as two cleavage events of CotF: one after the arginine at residue 4 and one after the arginine at residue 76. In this event, the true molecular size of the 5- and 8-kDa species would be 8.6 and 9.6 kDa, respectively. Another coat species that arises by proteolysis is the SDS/dithiothreitol-soluble, 8-kDa CotT protein, which is generated by cleavage of the 12-kDa primary product of the *cotT* gene at its NH_2 terminus (2). Interestingly, like the cleavages of CotF described here, cleavage of CotT to generate the 8-kDa species also occurs just after an arginine residue.

Map position of *cotF*. The position of *cotF* on the chromosome was determined by phage PBS1-mediated transduction and by DNA-mediated transformation, using the chloramphenicol resistance gene in the *cotF*Δ::*cat* mutant as a selectable marker. *cotF* (*cat*) was 90 and 12% cotransduced with the auxotrophic mutations *purA16* and *cysA14*, respectively, and 50% cotransformed with the *purA16* mutation. These results and the results of three-factor crosses (data not shown) indicate that the order of genes is *purA-cotF-cysA* and that the position of *cotF* is 349° on the genetic map of Henner and Hoch (5). Interestingly, as with the genes (*sspA* to *-E*; 9) that encode the principal sporulation-specific proteins of the spore core, the seven genes so far identified as encoding components of the spore coat are located at widely scattered sites on the chromosome (Fig. 4).

Mapping of the 5′ terminus of *cotF* mRNA. The 5′ terminus of *cotF* mRNA was mapped by extension of 18-nucleotide-long synthetic primers called Pr1, Pr2, and Pr4, which were designed to anneal to the proximal part of *cotF* mRNA (Fig. 1). In separate experiments, the primers were labeled at their 5′ termini and incubated under hybridization conditions with total RNA isolated from wild-type (PY79) cells harvested at h 9 after the onset of sporulation (T_9; when *cotF* mRNA is abundant; see below). Reverse transcriptase was used to generate cDNA primer extension products that were separated by electrophoresis in a 5% polyacrylamide gel alongside radioactively labeled, *Hpa*II-cut fragments of pBR322 that were denatured and used as size standards. The results were that primers Pr1 (data not shown), Pr2, and Pr4 (Fig. 5A) yielded extension products of 57, 91, and 152 bases, respectively. No primer extension product was obtained when *E. coli* RNA was used instead of *B. subtilis* RNA. These low-resolution mapping experiments indicate that the 5′ terminus of *cotF* transcripts is about 30 bp upstream of the initiation codon of the *cotF* open reading frame.

To map the 5′ terminus of *cotF* mRNA more precisely, we used a phage M13 clone containing *cotF* in a 2.3-kb *Hin*dIII-*Eco*RI fragment (see above) and primer Pr2 to generate a dideoxy sequencing ladder. The Pr2 primer extension product was then subjected to electrophoresis alongside the

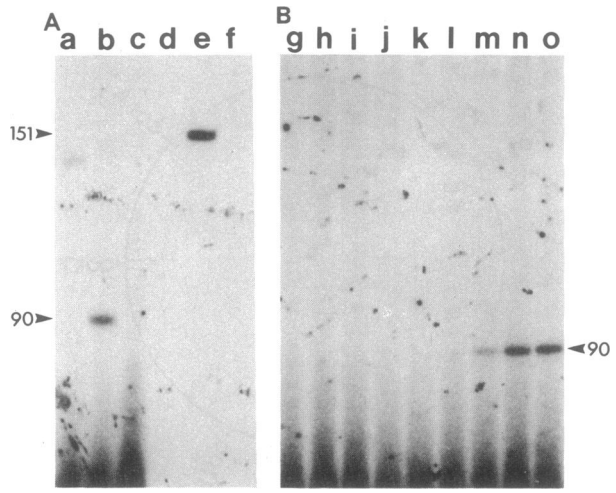

FIG. 5. Mapping of the 5′ terminus of *cotF* mRNA and the appearance of *cotF* mRNA during sporulation. (A) Products of primer extension reactions using oligonucleotides Pr2 and Pr4. Pr2 (lanes a to c) and Pr4 (lanes d to f) were labeled at their 5′ termini with ^{32}P and used to prime cDNA synthesis by reverse transcriptase after hybridization with RNA from *B. subtilis* sporulating cells harvested at T_9 (lanes b and e), RNA from *E. coli* (lanes a and d), and no RNA (lanes c and f). The reaction products were resolved in a 5% urea-polyacrylamide gel alongside appropriate size markers. Arrowheads show the positions (in bases) of the Pr2 and Pr4 extension products. (B) Time course of appearance of *cotF* mRNA during sporulation. Labeled oligonucleotide Pr2 was separately hybridized with 16-μg samples of total *B. subtilis* RNA from cells collected at T_0 to T_7 h (lanes g to n) and at T_9 (lane o) in Difco sporulation medium in the study of Zheng and Losick (13). Reverse transcriptase was used to prime cDNA synthesis, and the reaction products were resolved on a 5% urea-polyacrylamide gel. The methods for carrying out primer extension analysis were as previously described (13).

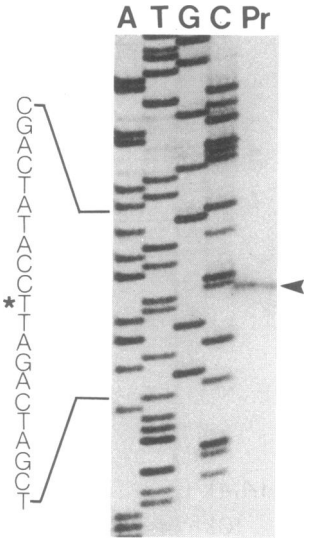

FIG. 6. High-resolution mapping of the 5′ terminus of *cotF* mRNA. Pr2 was used to prime synthesis of a dideoxy sequencing ladder (A, C, G, and T) from a phage M13 template containing the entire 2.3-kb *Hin*dIII-*Eco*RI fragment. The products of the sequencing reactions were subject to electrophoresis alongside a primer extension product (Pr) generated from RNA isolated from sporulating cells (harvested at T_9) with primer Pr2. The arrowhead shows the position of the extension product in the sequencing gel.

sequencing ladder. The results of electrophoresis (Fig. 6) show that the 5′ terminus (+1) of *cotF* mRNA was 28 bp upstream from the translation initiation codon (Fig. 1). An analogous experiment using primer Pr1 gave an identical position for the 5′ terminus (data not shown).

***cotF* is induced coincidently with genes of the σ^K regulon.** Genes encoding coat proteins are regulated in a hierarchical regulatory cascade (13). The cascade consists of four coordinately controlled gene sets whose transcription is induced by the successive appearance of regulatory proteins: σ^E, SpoIIID, σ^K, and GerE. The time course experiment of Fig. 5B indicates that *cotF* is a member of the so-called *cotA* regulon of σ^K-controlled genes (*cotA*, *cotD*, and *gerE*; 3, 7, 13), henceforth called the σ^K regulon. The experiment shows that *cotF* transcripts appeared during h 6 after the start of sporulation. These results can be directly compared with the previously determined times of appearance of transcripts of *cot* genes *E* (h 4 from its P1 promoter and h 5 from its P2 promoter), *D* (h 6), and *B* and *C* (h 7), since the same sporulation RNAs used in the experiment of Fig. 5B were used in the time course experiment of Zheng and Losick (13). Thus, *cotF* is induced coincidently with a representative (*cotD*) of the σ^K regulon. Additional considerations assigning *cotF* to the σ^K regulon are that (i) the putative -10 (CATA---T) and -35 (GAACA) regions of its promoter are highly similar in nucleotide sequence, and in position relative to the transcription start site, to the corresponding regions of the *cotD* promoter (13) and (ii) *cotF* transcription is not prevented by a mutation in the regulatory gene (*gerE*) controlling the terminal regulon (*cotB* and *cotC*; 13) of coat gene expression (data not shown).

ACKNOWLEDGMENTS

We thank William Lane of the Harvard Microchemistry Facility for carrying out the amino acid sequence determination and Chris Kroeger and Katrina Cooper for their participation in the early stages of the project.

This work was supported by NIH grant GM18568.

REFERENCES

1. **Aronson, A. I., and P. C. Fitz-James.** 1976. Structure and morphogenesis of the bacterial spore coat. Bacteriol. Rev. **40**:360–402.
2. **Aronson, A. I., H.-Y. Song, and N. Bourne.** 1988. Gene structure and precursor processing of a novel *Bacillus subtilis* spore coat protein. Mol. Microbiol. **3**:437–444.
3. **Cutting, S., S. Panzer, and R. Losick.** 1989. Regulatory studies on the promoter for a gene governing synthesis and assembly of the spore coat in *Bacillus subtilis*. J. Mol. Biol. **207**:393–404.
4. **Donovan, W., L. Zheng, K. Sandman, and R. Losick.** 1987. Genes encoding spore coat polypeptides from *Bacillus subtilis*. J. Mol. Biol. **196**:1–10.
5. **Henner, D. J., and J. A. Hoch.** 1980. The *Bacillus subtilis* chromosome. Microbiol. Rev. **44**:57–82.
6. **Jenkinson, H. F., W. D. Sawyer, and J. Mandelstam.** 1981. Synthesis and order of assembly of spore coat proteins in *Bacillus subtilis*. J. Gen. Microbiol. **123**:1–16.
7. **Sandman, K., L. Kroos, S. Cutting, P. Youngman, and R. Losick.** 1988. Identification of the promoter for a spore coat protein gene in *Bacillus subtilis* and studies on the regulation of its induction at a late stage of sporulation. J. Mol. Biol. **200**:461–473.
8. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
9. **Setlow, P.** 1989. Forespore-specific genes of *Bacillus subtilis*:

function and regulation of expression, p. 211–221. *In* I. Smith, R. A. Slepecky, and P. Setlow (ed.), Regulation of procaryotic development. American Society for Microbiology, Washington, D.C.

10. **Thomas, J. O., and R. D. Kornberg.** 1978. The study of histone-histone associations by chemical cross-linking. Methods Cell Biol. **18**:429–440.
11. **Youngman, P., J. B. Perkins, and R. Losick.** 1984. Construction of a cloning site near one end of Tn*917* into which foreign DNA may be inserted without affecting transposition in *Bacillus subtilis* or expression of the transposon-borne *erm* gene. Plasmid **12**:1–9.
12. **Zheng, L., W. Donovan, P. C. Fitz-James, and R. Losick.** 1988. Gene encoding a morphogenic protein required in the assembly of the outer coat of the *Bacillus subtilis* endospore. Genes Dev. **2**:1047–1054.
13. **Zheng, L., and R. Losick.** 1990. Cascade regulation of spore coat gene expression in *Bacillus subtilis*. J. Mol. Biol. **212**:645–660.