72. R. Kolesnick, *J. Biol. Chem.* **262**, 16759 (1987).
73. S. Hakomori, in *Sphingolipid Biochemistry*, J. N. Kanfer and S. Hakomori, Eds. (Plenum, New York, 1983), p. 327.
74. J. Thurin, *FEBS Lett.* **208**, 17 (1986).
75. C. Honsik, G. Jung, R. Reisfeld, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 7893 (1986); W. Dippold, A. Knuth, K-H Buschenfelde, *Cancer Res.* **44**, 806 (1984); I. Hellstrom, V. Brankovan, K. Hellstrom, *Proc. Natl. Acad. Sci. U.S.A.* **82**, 1499 (1985).
76. V. Larraga and M. Edidin, *Proc. Natl. Acad. Sci. U.S.A.* **76**, 2912 (1979).
77. W. Jozwiak and J. Koscielak, *Eur. J. Can. Clin. Oncol.* **18**, 617 (1982).
78. M. Fukuda *et al.*, *J. Biol. Chem.* **261**, 2376 (1986).
79. G. Matyas *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 6065 (1987).
80. T. Taki *et al.*, *J. Biol. Chem.* **261**, 3075 (1986).
81. M. Coulen-Morelec and M. Buc-Caron, *Dev. Biol.* **83**, 278 (1981).
82. R. Glickman and J. Bouhours, *Biochim. Biophys. Acta* **424**, 17 (1976).
83. T. Feizi, R. Childs, K. Watanabe, S. Hakomori, *J. Exp. Med.* **149**, 975 (1979).
84. R. Whatley, S. Ng, J. Roger, W. McMurray, B. Sanwal, *Biochem. Biophys. Res. Commun.* **70**, 180 (1976).
85. J. Moskal, D. Gardner, S. Basu, *ibid.* **61**, 751 (1974).
86. G. Schwarting and A. Summers, *J. Immunol.* **124**, 1691 (1980).
87. W. J. van Blitterswijk, in *Membrane Fluidity in Biology*, R. C. Aloia and J. M. Boggs, Eds. (Academic Press, New York, 1985), vol. 3, p. 85.
88. F. Sharom and C. Grant, *J. Supramol. Struct.* **6**, 249 (1977).
89. R. Ledeen, *J. Neurosci. Res.* **12**, 147 (1984); M. Byrne, R. Ledeen, F. Roisen, G. Yorke, J. Scalafani, *J. Neurochem.* **41**, 1214 (1983).
90. E. Bremer, S. Hakomori, D. Bowen-Pope, E. Raines, R. Ross, *J. Biol. Chem.* **259**, 6818 (1984).
91. D. Marcus, *Mol. Immunol.* **21**, 1083 (1984); R. Robb, *J. Immunol.* **136**, 971 (1986).
92. R. Katok-Semba, L. Facci, S. Skaper, S. Varon, *J. Cell. Physiol.* **126**, 147 (1986).
93. K. Chan, *J. Biol. Chem.* **262**, 2415 (1987).
94. W. Van Heyningen, *J. Gen. Microbiol.* **20**, 291 (1959).
95. L. Thompson *et al.*, *J. Biol. Chem.* **261**, 5209 (1986).
96. D. Roberts, C. Rao, L. Liotta, H. Gralnick, V. Ginsburg, *ibid.*, p. 6872; D. Roberts, S. Williams, H. Gralnick, V. Ginsburg, *ibid.*, p. 3306.
97. S. Hakomori and W. Young, in *Sphingolipid Biochemistry*, J. N. Kanfer and S. Hakomori, Eds. (Plenum, New York, 1983), p. 381.
98. K. Watanabe, S. Hakomori, R. Childs, T. Feizi, *J. Biol. Chem.* **254**, 3221 (1979). D. Markus, M. Naiki, S. Kundu, *Proc. Natl. Acad. Sci. U.S.A.* **73**, 3263 (1976).
99. Experimental allergic encephalomeyelitis, J. M. Fry, S. Weissbarth, G. M. Lehrer, M. Bornstein, *Science* **183**, 540 (1974); paroxysmal cold hemoglobinuria, P. Levine, M. Celano, F. Falkowski, *Ann. N.Y. Acad. Sci* **124**, 456 (1965); and multiple sclerosis, H. Hirsch and M. Parks, *Nature* **264**, 785 (1976). For experimental ganglioside syndrome, see (*104*).
100. S. Wenzi and M. Sumpe, *Cell* **46**, 633 (1986).
101. E. Okuyama and M. Yamazaki, *Chem. Pharm. Bull.* **31**, 2269 (1983).
102. G. Kawai, M. Ohnishi, Y. Fujino, Y. Ikeda, *J. Biol. Chem.* **261**, 779 (1986).
103. S. Spiegel, J. Handler, P. Fishman, *ibid.*, p. 15775.
104. Y. Nagai, T. Momoi, M. Saito, E. Mitsuzawa, S. Ohtani, *Neurosci. Lett.* **2**, 107 (1976).
105. C. Gahmberg and S. Hakomori, *J. Biol. Chem.* **250**, 2438 (1975).
106. M. Civan, K. Peterson-Yantorno, T. O'Brien, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 963 (1988).
107. J. Connor, W. Wadman, P. Hockberger, R. Wong, *Science* **240**, 649 (1988).
108. W. Stoffel, G. Heimann, B. Hellenbroich, *Z. Physiol. Chem.* **354**, 562 (1973); R. Keenan and A. Maxam, *Biochim. Biophys. Acta* **176**, 348 (1969).
109. H. Kondoh, H. Kanoh, T. Ono, *ibid.* **753**, 97 (1983).
110. Supported by NIH grant GM38737, and by NCI grant CA46738, and a physicians scientist award (to Y.H.). Y.H. is a Pew Scholar. We thank C. R. Loomis and J. E. Niedel for careful review of this manuscript.

Research Articles

Chromosomal Rearrangement Generating a Composite Gene for a Developmental Transcription Factor

PATRICK STRAGIER,* BARBARA KUNKEL, LEE KROOS,† RICHARD LOSICK‡

Differential gene expression in the mother cell chamber of sporulating cells of *Bacillus subtilis* is determined in part by an RNA polymerase sigma factor called σ^K (or σ^{27}). The σ^K factor was assigned as the product of the sporulation gene *spoIVCB* on the basis of the partial amino-terminal amino acid sequence of the purified protein. The *spoIVCB* gene is now shown to be a truncated gene capable of specifying only the amino terminal half of σ^K. The carboxyl terminal half is specified by another sporulation gene, *spoIIIC*, to which *spoIVCB* becomes joined in-frame at an intermediate stage of sporulation by site-specific recombination within a 5–base pair repeated sequence. Juxtaposition of *spoIVCB* and *spoIIIC* need not be reversible in that the mother cell and its chromosome are discarded at the end of the developmental cycle. The rearrangement of chromosomal DNA could account for the presence of σ^K selectively in the mother cell and may be a precedent for the generation of cell type–specific regulatory proteins in other developmental systems where cells undergo terminal differentiation.

THE METAMORPHOSIS OF CELLS OF THE GRAM-POSITIVE bacterium *Bacillus subtilis* into endospores involves the formation of a sporangium consisting of two cell types known as the mother cell and the forespore (*1, 2*). The mother cell and forespore arise by an asymmetric septation, which partitions the sporangium into two unequal compartments. These compartments each receive a chromosome generated by the last round of vegetative DNA replication, but then undergo divergent developmental fates as a consequence of differential gene expression. The forespore can be thought of as a germline cell, because it ultimately becomes the mature spore and gives rise to subsequent progeny. The mother cell, however, is a terminally differentiating cell because it and its chromosome are discarded by lysis when maturation of the spore is complete.

The authors are in the Department of Cellular and Developmental Biology, Biological Laboratories, Harvard University, Cambridge, MA 02138.

*Permanent address: Institut de Microbiologie, Bat. 409, Université Paris-Sud, 91405 Orsay Cedex, France.
†Present address: Department of Biochemistry, Michigan State University, East Lansing, MI 48824.
‡To whom correspondence should be addressed.

Differential gene expression in the two compartments is determined in part by compartment-specific RNA polymerase sigma factors known as σ^G (*3*) and the newly discovered σ^K (or σ^{27}) described in an accompanying report (*4*). The σ^G factor is present in the forespore compartment (*5*) where it directs the transcription of forespore-expressed genes, such as *sspA-E* (*3*), which encode a family of small basic proteins found in the core of the mature spore (*6*). In contrast, σ^K is present in the mother cell (*4*) and directs the transcription of genes encoding the structural components of the coat, the tough protein shell that encases the mature spore (*1, 7*). The factor σ^G is the product of the forespore regulatory gene *spoIIIG* (*3, 5, 8*), whereas σ^K is specified by the mother cell regulatory gene *spoIVCB* (*9*), as deduced from the partial amino-terminal sequence determination of the sigma factor (*4*).

The selective synthesis of σ^G in the forespore and of σ^K in the mother cell is at least in part the basis for compartmentalized gene expression in sporulating cells. As a result of our studies on the mechanisms that govern the differential localization of these transcription factors in the two-compartment sporangium, we report the surprising discovery that *spoIVCB* is an incomplete structural gene that is capable of encoding only the amino terminal half of σ^K. The remainder (that is, the carboxyl terminal half) of σ^K is specified by another regulatory gene that is specific for the mother cell; this gene, known as *spoIIIC* (*10, 11*), is separated by at least 10 kilobases from *spoIVCB* on the chromosome. We show that *spoIVCB* becomes joined in-frame to *spoIIIC* to create a composite structural gene for σ^K by a chromosomal rearrangement occurring at an intermediate stage of sporulation. We suggest that the rearrangement occurs only in the mother cell and is in part responsible for the presence of σ^K only in the mother cell. Because the mother cell is a terminally differentiating cell and is discarded at the end of the sporulation cycle, DNA rearrangements altering its chromosome need not be reversible.

The *spoIVCB* gene encodes a truncated sigma factor. The *spoIVC* locus consists of *B* and *A* genes (*12*) that are separately transcribed and convergently oriented (*9*). The nucleotide sequence of the 5′ strand of the *B* gene of the locus (henceforth *spoIVCB*) and the predicted amino acid sequence of its product are shown in Fig. 1. The underlined residues correspond to the NH_2-terminal sequence of σ^K as determined by sequential Edman degradation (*4*). As noted by Kroos *et al.* (*4*), the NH_2-terminus of mature σ^K is preceded in the predicted coding sequence by 20 amino acids, from which we infer that the primary gene product is a pro-protein that is subsequently processed to the mature sigma. The size of the *spoIVCB* open reading frame (155 codons) is inadequate to encode a polypeptide of the apparent molecular size (27,000 daltons) of σ^K, even allowing for the fact that sigma factors often migrate more slowly than expected during SDS–polyacrylamide gel electrophoresis (PAGE). Thus, the product of the entire open reading frame (that is, the predicted pro-protein) is only 17.3 kD in size, and the deduced product of the mature form (135 codons) is only 15.2 kD in size.

The validity of the surprisingly short open reading frame is substantiated by the following considerations:

1) The open reading frame is based on sequence determinations of both DNA strands with the use of two independently cloned copies of the gene (*13*). Therefore the position of the termination codon is unlikely to be due to a sequencing error or the consequence of a mutation generated during the course of cloning in *Escherichia coli*.

2) Integration into the chromosome by single-reciprocal (Campbell) recombination of a plasmid bearing a 0.4-kb fragment extending from the Sac I site within the open reading frame to the Sac I site located 60 bp downstream of the open reading frame (Fig. 1) yielded Spo^+ transformants. If the true *spoIVCB* sequence extended a sufficient distance (at least 200 bp) downstream to encode a 27-kD protein, then the Sac I–Sac I segment would have been internal to the open reading frame and hence would have been expected to disrupt the gene upon integration. The fact that the transformants were Spo^+ confirms that this Sac I–Sac I segment is not internal to *spoIVCB*, and hence that the σ^K coding sequence must be interrupted.

3) The *spoIVCB* open reading frame is similar to the NH_2-terminal part of sigma factors. It contains a region of substantial similarity to a highly conserved feature of sigma factors thought to be the binding site to core RNA polymerase (*14*) (Fig. 2). This region is known as the core binding site or region 2.2 (*14*). Downstream of region 2.2 is an element that is thought to mediate contact with the −10 region of cognate promoters (*15, 16*). In other sigma factors, this element displays three highly conserved hydrophobic residues (generally isoleucines) at intervals of four amino

```
5′ AATATTTTTAACATTTGAAGTTAGTATGCTGCTTACCAAAGCCGGACTCCCCCGCGAGAAATTTCCCGGTACAGAC

ACAGACAGCCTCCCGGTCACATACATTTACATATAGGCTTTTGCCTACATACTTTTGTGGAGGTGACGATG GTG ACA
                                                                        M   T      2

GGT GTT TTC GCA GCG CTC GGC TTT GTT GTT AAA GAG CTT GTC TTT TTA GTA TCT TAC GTG
 G   V   F   A   A   L   G   F   V   V   K   E   L   V   F   L   V   S   Y   V     22
                                                                    SacI
AAA AAC AAT GCC TTT CCA CAA CCG CTC TCA AGC AGC GAA GAA AAA AAA TAC TTA GAG CTC
 K   N   N   A   F   P   Q   P   L   S   S   S   E   E   K   K   Y   L   E   L     42

ATG GCT AAA GGG GAT GAA CAT GCC AGA AAC ATG CTG ATT GAG CAT AAT CTT CGC TTG GTC
 M   A   K   G   D   E   H   A   R   N   M   L   I   E   H   N   L   R   L   V     62

GCC CAT ATT GTG AAA AAG TTC GAA AAT ACA GGT GAG GAT GCA GAG GAC TTA ATC TCC ATC
 A   H   I   V   K   K   F   E   N   T   G   E   D   A   E   D   L   I   S   I     82

GGA ACG ATC GGG CTT ATT AAA GGA ATT GAA AGC TAT TCC GCT GGA AAA GGG ACA AAG CTG
 G   T   I   G   L   I   K   G   I   E   S   Y   S   A   G   K   G   T   K   L    102

                                    *** **
GCG ACG TAT GCA GCG AGG TGT ATT GAA AAT GAG ATT GTA ATT ACA AAA GGG GGG TGC ATA
 A   T   Y   A   A   R   C   I   E   N   E   I   V   I   T   K   G   G   C   I    122

CAC CCC TCT TTA ATA CGT TTC AAT ATA TAT GGT GTC AGA ATC CAC AAT GGT AAC TTC TTT
 H   P   S   L   I   R   F   N   I   Y   G   V   R   I   H   N   G   N   F   F    142

CAC GAT AAA GTT AAC AAT TGT TTT TTT ATC TTC AAG AGT TAA GTTATCTGCACCGATTGATTGAA
 H   D   K   V   N   N   C   F   F   I   F   K   S   -                           155
                                SacI
AATAGTCGATGGCTCTTTTTAGAGCATTTTCACTTGAGCTC 3′
```

Fig. 1. Nucleotide sequence of the 5′ strand of the *spoIVCB* gene. Nucleotide sequence analysis was carried out as described (*13*). The deduced amino acid sequence is shown beneath the nucleotide sequence. The column of numbers at the right indicates the positions in the amino acid sequence of predicted residues at the end of each line. The predicted initiation codon is shifted downstream by one codon from that proposed previously (*9*) in order to achieve a more appropriate spacing relative to the ribosome binding site (underlined). The position of the NH_2-terminal sequence of σ^K as determined by sequential Edman degradation (*4*) is shown by the underlined amino acids. The previously determined (*9*) transcription start site is at the base of the arrow. The asterisks above the sequence AATGA identify the site of recombination with *spoIIIC*. Single letter abbreviations for the amino acid residues are: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.

acids (Fig. 2). The homology of the *spoIVCB* product to other sigma factors ends within this element, with the third highly conserved hydrophobic residue being replaced by a lysine (at codon 118) (Fig. 1), and with the open reading frame terminating 37 codons downstream of this lysine codon.

4) Analysis of sequences downstream of the stop codon at position 156 did not reveal any potential amino acid sequence with significant similarity to the characteristic α helix–β turn–α helix region found at the COOH-terminus of bacterial sigma factors, an element that is believed to be responsible for the recognition of the −35 sequences of promoters (*16, 17*).

5) Another open reading frame of more than 145 codons in length of converging orientation, which is likely to be the *A* gene of the *spoIVC* locus, extends all the way to and actually overlaps with the last 29 codons of the *B* open reading frame. It seems unlikely that the overlap of the coding sequences on the two DNA strands could extend much further.

The *spoIVCB* gene is joined to *spoIIIC* by a chromosomal rearrangement during sporulation. The *spoIIIC* gene is a sporulation gene whose predicted product (*11*) is similar to the COOH-terminus of sigma factors, as noted by Errington (*10*). [The name *spoIIIC* is actually a misnomer as *spoIIIC* mutants are blocked at stage IV and their phenotype is similar to that of *spoIVCB* mutants (*2*).] The similarity of the *spoIIIC* product to the COOH-terminus of sigma factors begins close to and partially overlaps with the end of the similarity of the *spoIVCB* product with the NH_2-terminus of sigma factors (Fig. 2). We postulated that *spoIVCB* is joined to *spoIIIC* by a site-specific recombinational event to create a composite coding sequence for σ^K. If the recombinational event occurred within the repeated sequence of AATGA (as shown in Figs. 3 and 4A), which is present near the terminus of *spoIVCB* and close to the beginning of *spoIIIC*, the resulting composite gene (241 codons) would specify after processing a 221-residue polypeptide with a molecular size of 25.4 kD.

To test this hypothesis, we extracted DNA from cells early (hour 1) and late (hour 5) in development, and subjected the DNA's to DNA hybridization analysis with radioactively labeled *spoIVCB*- and *spoIIIC*-containing DNA's as probes. Hybridization to DNA from early sporulating cells (Fig. 4B) shows that radioactive *spoIVCB* DNA reacted with a 7.5-kb Sac I fragment, a 1.7-kb Pvu II fragment, a 3.6-kb Eco RI fragment, a 7.7-kb Hind III fragment and Stu I fragments of about 1.3 and 0.7 kb, a hybridization pattern that was fully consistent with the known endonuclease restriction map of the *spoIVCB* region of the chromosome. Similarly, the *spoIIIC* probe (Fig. 4C) reacted with Eco RI and Hind III fragments of 4.7 and 1.3 kb, respectively, and Stu I fragments of

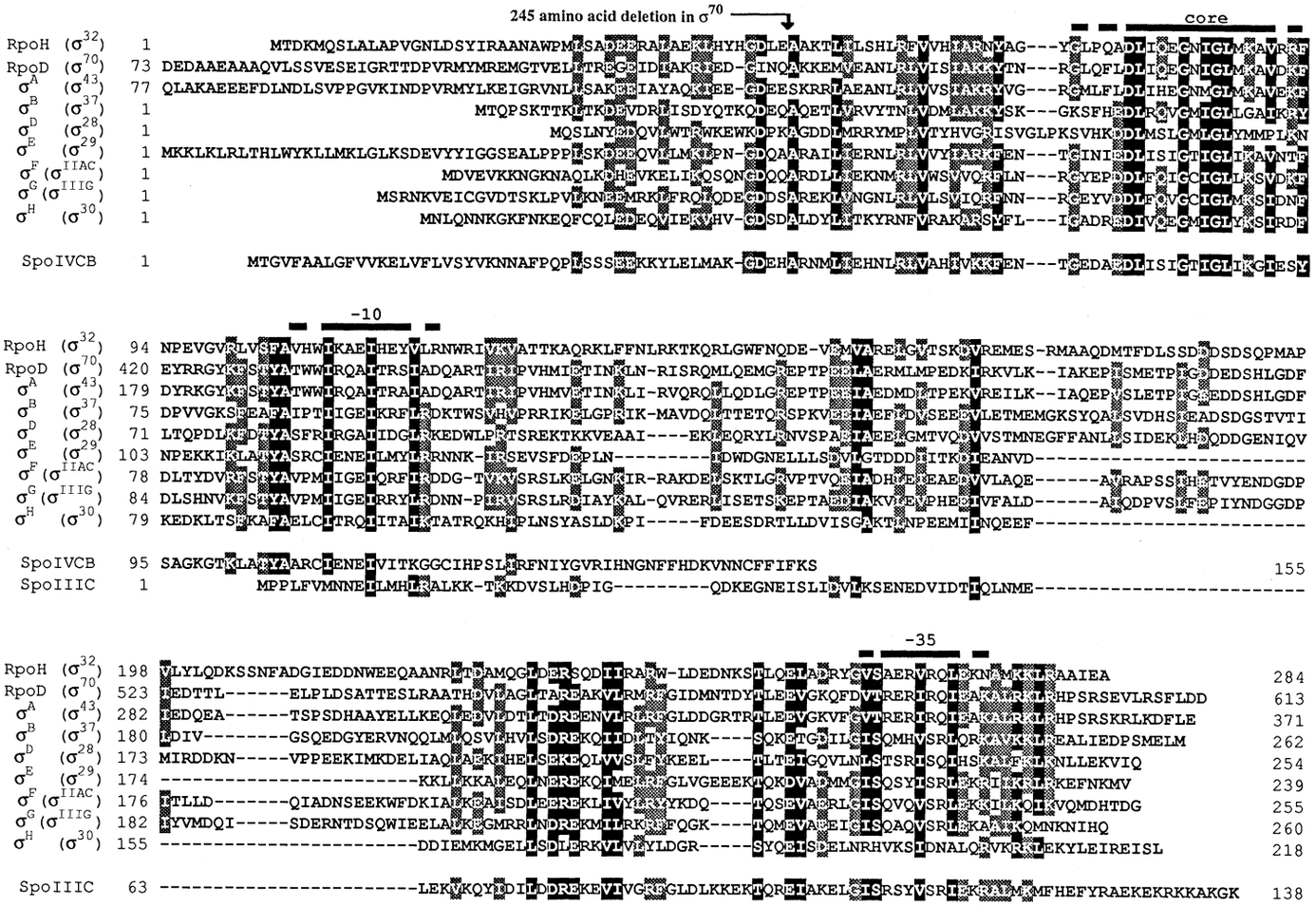

Fig. 2. Alignment of the predicted products of *spoIVCB* and *spoIIIC* with bacterial sigma factors. Two sigma factors of *E. coli* (σ^{32} and σ^{70}) and seven sigma factors of *B. subtilis* are aligned according to Karmazyn-Campelli *et al.* (*5*). A 245–amino acid deletion corresponding to residues 130 to 374 was introduced in the σ^{70} sequence at the position shown by the arrow. Chemically similar amino acids are defined as the following groups: D and E; K and R; S and T; F and Y; I, L, and V. A black background indicates positions occupied by identical or similar amino acids in eight or nine sigma factors. A grey background indicates positions occupied by identical or similar residues in six or seven sigma factors. The locations of functional domains are indicated. The alignment of the *spoIIIC* product corresponds to that suggested by Errington (*10*).

Fig. 3. Alignment of *spoIVCB* and *spoIIIC* nucleotide sequences at the site of recombination. The *spoIIIC* sequence is from Errington *et al.* (*11*). The identical 5-bp sequence within which recombination occurs is identified by the asterisks. The arrows indicate the 21-bp inverted repeat sequence.

```
spoIVCB  GCGACGTATGCAGCGAGGTGTATTGAAAATGAGATTGTAATTACAAAAGGGGGGTGCATACA
         CGCTGCATACGTCGCTCCACATAACTTTTACTCTAACATTAATGTTTTCCCCCCACGTATGT

spoIIIC  TTTATGCCCCCCCTTTTTGTAATGAATAATGAAATCCTCATGCATTTGCGCGCATTGAAAAA
         AAATACGGGGGGGAAAAACATTACTTATTACTTTAGGAGTACGTAAACGCGCGTAACTTTTT
```

approximately 5 and 3.4 kb, a pattern of hybridization that was once again in accord with the known physical map of the *spoIIIC* chromosomal region (*11*).

To determine whether *spoIVCB* becomes joined to *spoIIIC* at late developmental stages, we compared the pattern of hybridization observed with DNA fragments from early sporulating cells with that obtained by hybridization to DNA from late sporulating cells.

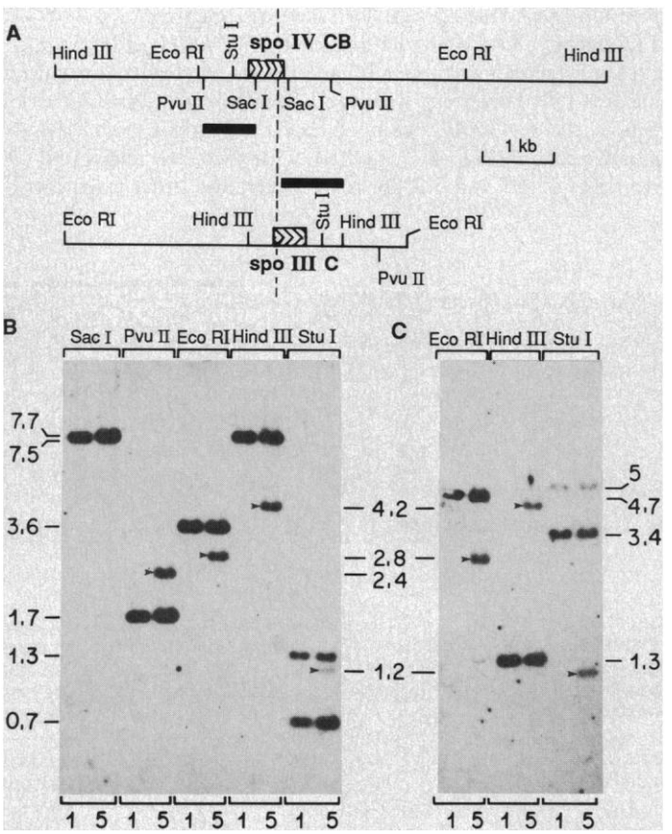

Fig. 4. Southern hybridization analysis of DNA from sporulating cells. (**A**) Partial endonuclease restriction map of the *spoIVCB* and *spoIIIC* regions of the chromosome based on (*9*, *11*), and our own mapping experiments. Only previously known restriction sites relevant to the DNA hybridization analysis are shown, and certain sites (Stu I) revealed in the DNA hybridization that had not been previously mapped in cloned DNA are not shown in the map. The cross-hatched boxes show the location and orientation of *spoIVCB* and *spoIIIC*. The horizontal bars show the sequences used as probes in the DNA hybridizations. These are the 0.74 kb Pvu II–Sac I segment for *spoIVCB* and 0.92-kb Bss HII–Hind III segment for *spoIIIC*. (The Bss HII site is not labeled in the map.) The vertical dotted line indicates the proposed site of recombinational joining of the two genes. (**B** and **C**) Autoradiographs of filters containing electrophoretically separated restriction fragments that had been incubated under hybridization conditions with radioactively labeled plasmid subclones containing, respectively, the *spoIVCB* and *spoIIIC* inserts identified by the thick bars in (A). The filters were prepared by transfer from a 0.7 percent agarose gel in which restriction fragments generated by cleavage with the indicated endonuclease restriction enzymes had been subjected to electrophoresis. The lanes alternately contain restriction fragments of DNA from cells harvested early (hour 1) and late (hour 5) in sporulation (*28*). The arrowheads identify hybridized fragments that occur only in the DNA from cells harvested at hour 5.

Reaction of the *spoIVCB* probe (Fig. 4B) to Sac I fragments revealed no difference between early- and late-purified DNA's. This was the expected result as Sac I cuts *spoIVCB* upstream of the anticipated rearrangement site and the radioactive probe was homologous only to sequences upstream of Sac I. However, hybridization of the *spoIVCB* probe to Pvu II, Eco RI, Hind III, and Stu I fragments revealed, in addition to the fragments that were present in DNA from the early developing cells, new fragments of 2.4, 2.8, 4.2, and 1.2 kb, respectively, which occurred only in the DNA from late developing cells (Fig. 4B). The sizes of the new fragments were in each instance completely consistent with those expected as a result of a recombinational event between a site (the vertical dotted line in Fig. 4A) near the terminus of the *spoIVCB* open reading frame and a corresponding site near the beginning of the *spoIIIC* open reading frame.

As further evidence that the new fragments were the consequence of a juxtaposition of *spoIVCB* and *spoIIIC*, hybridization with *spoIIIC*-containing probe (Fig. 4C) revealed the same sized, rearranged Eco RI, Hind III, and Stu I fragments (2.8, 4.2, and 1.2 kb, respectively) as had been detected with the *spoIVCB*-bearing probe. That the *spoIVCB* and *spoIIIC* probes had reacted with identical-sized fragments in DNA from late developing cells was confirmed in an experiment in which the filter of Fig. 4B was washed under denaturing conditions to melt off the *spoIVCB* probe and then hybridized again with radioactive *spoIIIC* DNA.

Joining of *spoIVCB* in-frame to *spoIIIC* generates a composite open reading frame. As a direct test of the rearrangement hypothesis, we cloned the putative *spoIVCB-spoIIIC* gene fusion from DNA extracted from late sporulating cells. To do this, we cloned into phage M13mp19 size-selected fragments of DNA, 1 to 1.5 kb in length, that had been digested with both Sac I and Hind III. (The juxtaposition of the two genes was expected to generate a Sac I–Hind III fragment of about 1.1 kb.) Clones bearing the expected insert were identified by plaque hybridization with radioactively labeled DNA containing *spoIIIC* sequences as a probe. A nucleotide sequencing ladder was obtained across the expected junction of *spoIVCB* and *spoIIIC* DNA of one such clone (Fig. 5). It can be seen (i) that the clone contains both *spoIVCB* and *spoIIIC* sequences; (ii) that the junction lies within the 5-bp repeated sequence postulated (above) to be the site at which the recombinational event would take place; and (iii) that the two genes had been joined in-frame so as to generate a composite open reading frame (whose predicted length is 241 codons).

The *spoIVCB-spoIIIC* joining (as monitored by the appearance of the rearrangement-specific 4.2-kb Hind III fragment) was not detected until the third hour (stage III) of sporulation (Fig. 6A). The rearrangement was prevented in sporulation mutants *spoIIG* (Fig. 6B, lane 7) and *spoIIID* (lane 8), but it was not prevented in mutants *spoIID* (lane 6) and *spoIIIG* (lane 9).

Implications for the role of chromosomal rearrangement in developmental regulation. The chromosomal DNA rearrangement that generates a composite gene for the mother cell sigma factor σ^K is developmental-specific because it occurs at the third hour of sporulation, and it is prevented in certain (for example, *spoIIG* and *spoIIID*) but not all (for example, *spoIID* and *spoIIIG*) mutants blocked at stages II and III of sporulation. The rearrangement is probably restricted to the mother cell chromosome because it is prevented by a mutation in *spoIIID*, a regulatory gene for the

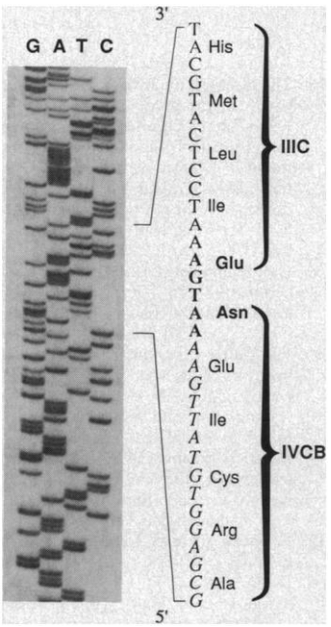

Fig. 5. Nucleotide sequence across the expected junction of *spoIVCB* and *spoIIIC* in the rearranged, composite gene for σ^K. A Sac I–Hind III fragment of 1.1 kb spanning the junction of *spoIVCB* and *spoIIIC* was cloned into phage M13mp19 as described (*29*) from DNA extracted from cells at a late stage of sporulation. Single-stranded phage DNA from the clone was subjected to sequencing by the procedure of Sanger *et al.* (*30*). Primer extension DNA synthesis began from the Sac I end of the fragment (representing *spoIVCB* DNA) and continued to and across the junction site with *spoIIIC*. A region of the sequencing ladder containing the junction of *spoIVCB* sequences (italicized bases) with *spoIIIC* sequences (nonitalicized bases) is shown. The boldface sequence (AATGA) represents a 5-bp sequence common to both genes at the site at which recombination occurs.

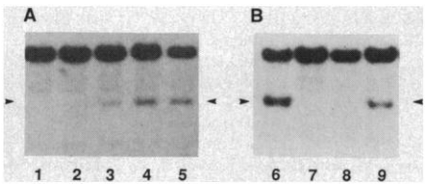

Fig. 6. The rearrangement of chromosomal DNA is developmentally regulated. *Bacillus subtilis* chromosomal DNA was digested with Hind III and subjected to DNA hybridization analysis as described for Fig. 4B, with radioactively labeled *spoIVCB*-containing DNA as a probe. The arrowheads identify the 4.2-kb rearranged fragment. (**A**) The DNA was extracted from wild-type cells (strain PY79) harvested at hours 1 to 5 of sporulation (lanes 1 to 5, respectively). (**B**) The DNA was extracted at hour 5 from the following mutants (*31*): strain 298.4 bearing the mutation *spoIID298* (lane 6); strain 55.3 bearing the mutation *spoIIG55* (lane 7); strain BK395 bearing the mutation *spoIIID83* (lane 8); and strain BK338 bearing the mutation *spoIIG*Δ1 (lane 9).

mother-cell line of gene expression (*4, 9*), but not by a mutation in *spoIIIG*, a gene that governs forespore gene expression (*3*). If the juxtaposition of *spoIVCB* and *spoIIIC* is restricted to the mother cell, then the rearrangement process need not be reversible because the mother cell chromosome is discarded at the end of the developmental cycle.

The rearrangement of chromosomal DNA in the mother cell could be part of the underlying mechanism for compartmentalized gene expression. Thus, transcription that is directed by σ^K-containing RNA polymerase may be limited to the mother cell simply because the *spoIVCB-spoIIIC* composite gene, which we designate *sigK*, is generated only in this compartment. Compartmentalization of σ^K could also be achieved, however, by selective utilization of the *spoIVCB* promoter in the mother cell or by compartment-specific processing of pro-σ^K, the hypothetical primary product of the composite gene. Insofar as *spoIVCB* is autoregulatory (it has a σ^K-utilized promoter) (*4*), the previously observed compartmentalization of *spoIVCB* transcription (*9*) may largely be a secondary consequence of the compartmentalization of the rearrangement process. Thus, if the composite gene and hence σ^K is present only in the mother cell, then transcription from the *spoIVCB* promoter would necessarily be largely confined to the mother cell. The relative contributions of DNA rearrangement, selective transcription, and pro-σ^K processing to the compartmentalization of σ^K and of σ^K-directed gene expression remain to be determined.

The gene *sigK* is a composite that encodes a presumed binding domain to core RNA polymerase; this domain is a region of high amino acid conservation among sigma factors, which is provided by *spoIVCB*. The gene *sigK* also encodes a presumed −35 recognition region, the α helix–β turn–α helix element near the COOH-terminus, which in turn is provided by *spoIIIC* (Fig. 2). Interestingly, recombination between *spoIVCB* and *spoIIIC* occurs within the coding region for a third domain that is similar to the element in other sigma factors that is thought to interact with the −10 region of promoters (*14–16*). Thus, the presumed −10 recognition region of σ^K is encoded in part by *spoIVCB* and in part by *spoIIIC*. Each gene appears therefore to encode one of two functional domains of σ^K, the third one (the −10 recognition element) being a hybrid region resulting from the recombination process.

The rearrangement is the result of site-specific recombination within a 5-bp sequence (AATGA) present identically in *spoIVCB* and *spoIIIC*. A striking inverted repeat sequence (identical in 19 out of 21 bp) is present (Fig. 3) in *spoIVCB* and *spoIIIC* at positions flanking the recombination site. Conceivably, the inverted repeat sequences are recognition elements for a recombinase that has as its substrate chromosomal segments aligned as shown in Fig. 3. The recombinase could be the product of the second member of the *spoIVC* locus because in preliminary experiments we have found that recombination is prevented in a *spoIVCA* mutant.

In that the relative positions and orientations of *spoIVCB* and *spoIIIC* are unknown, several topological possibilities exist for the recombination event. If *spoIIIC* is located downstream of and in the same orientation as *spoIVCB*, then the recombinational event would result in the excision of sequences located between the genes. If *spoIIIC* is located downstream (or upstream) of *spoIVCB* but in the opposite orientation, then site-specific recombination would result in an inversion of intervening sequences. If *spoIIIC* is located upstream and in the same orientation as *spoIVCB*, then recombination would result in the excision of the composite gene and expression of *sigK* from an extrachromosomal element.

Whatever the basis for the juxtaposition, a significant segment of chromosomal DNA is involved in the rearrangement process. Even though they are located in the same general vicinity of the chromosome, *spoIVCB* and *spoIIIC* are at least 10 kb apart, as judged by the failure to detect linkage between them by co-transformation. Nevertheless, the rearrangement mechanism evidently imposes a limit on how far apart from each other (or in what relative orientation) *spoIVCB* and *spoIIIC* can be. Thus, in the course of our work we have found that a *spoIVCB* mutation is complemented very poorly (<0.1 percent efficiency) by a cloned copy of the entire *spoIVC* locus that has been inserted into the chromosome at the distantly located *amy* locus. The same DNA does, however, complement a *spoIVCA* mutation with high efficiency. Errington *et al.* (*11*) have reported complementation of a *spoIIIC* mutation by a specialized Φ105 transducing phage bearing a cloned copy of *spoIIIC* contained in a 1.3-kb Hind III fragment, but the efficiency of complementation was found to be low (*18*).

Chromosomal DNA rearrangements associated with alternative developmental states are well known in prokaryotes and eukaryotes. Examples are (i) flagellar antigen phase variation switching in *Salmonella*, in which gene expression is regulated by recombinational inversion of promoter-bearing DNA (*19*); (ii) host-range variation in phage Mu in which recombinational inversion generates alternative composite coding sequences for phage tail fiber genes (*20*); (iii) mating type interconversion in yeast, in which regulatory information for alternative mating types is transferred (in mother cells but not in daughters) by a kind of site-specific "gene conversion" process to an expression locus (*21*); and (iv) the generation of vertebrate immunoglobulin genes, in which antibodies are generated by the juxtaposition during B cell development of coding

elements for the variable and constant domains of the light and heavy chains of immunoglobulin molecules (*22*).

The joining of *spoIVCB* to *spoIIIC* during differentiation in *B. subtilis* is, however, most reminiscent of the case of the rearrangement of nitrogen fixation genes during heterocyst differentiation in the cyanobacterium *Anabaena*. Golden *et al.* (*23*) have shown that the formation of heterocysts, a terminally differentiated cell type that carries out nitrogen fixation, involves the excision of an 11-kb intervening sequence that interrupts *nifD* in chromosomal DNA of vegetative cells. The excision, which occurs by site-specific recombination between an 11-bp repeated sequence and is catalyzed by a recombinase specified by sequences within the 11-kb intervening sequence (*24*), replaces the terminal 26 codons of the vegetative *nifD* open reading frame with 43 codons juxtaposed as a result of the rearrangement, and creates a complete *nifHDK* operon (*23*). Another rearrangement deletes a 55-kb sequence in the *nifS* region (*25*).

Whereas the rearrangement process in *Anabaena* creates the open reading frame for a nitrogenase gene and restores *nif* operon expression, the chromosomal juxtaposition we have described generates a regulatory gene whose product directs gene expression in the cell type (the mother cell) in which the rearrangement occurs. We suggest that the generation of cell type–specific regulatory genes by chromosomal rearrangements could be a general feature of developmental systems that involve terminally differentiating cell types. In *B. subtilis* itself other rearrangements could activate or inactivate additional regulatory genes in a temporal and compartment-specific way. In *Anabaena* other yet to be identified chromosomal rearrangements could be responsible for generating regulatory genes whose products mediate heterocyst-specific gene expression. Other microbial developmental systems that might be expected to employ such a strategy are those involving cell types that are programmed to undergo lysis, such as aerial mycelium formation in streptomycetes (*26*) and fruiting body formation in myxobacteria (*27*). In higher organisms irreversible rearrangements may create regulatory genes whose products govern gene expression in nongermline cell types.

REFERENCES AND NOTES

1. R. Losick and P. Youngman, in *Microbial Development* R. Losick and L. Shapiro, Eds. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1984), pp. 63–88; R. Losick, P. Youngman, P. J. Piggot, *Annu. Rev. Genet.* **20**, 625 (1986).
2. P. Piggot and J. Coote, *Bacteriol. Rev.* **40**, 908 (1976).
3. D. Sun, P. Stragier, P. Setlow, *Genes Dev.* **3**, 141 (1989).
4. L. Kroos, B. Kunkel, R. Losick, *Science* **243**, 526 (1989).
5. C. Karmazyn-Campelli, C. Bonamy, B. Savelli, P. Stragier, *Genes Dev.* **3**, 150 (1989).
6. J. M. Mason, R. H. Hackett, P. Setlow, *J. Bacteriol.* **170**, 239 (1988); P. Setlow, *Annu. Rev. Microbiol.* **42**, 19 (1988).
7. W. Donovan, L. Zheng, K. Sandman, R. Losick, *J. Mol. Biol.* **196**, 1 (1987).
8. E. Masuda, H. Anaguchi, K. Yamada, Y.Kobayashi, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 7637 (1988).
9. B. Kunkel, K. Sandman, S. Panzer, P. Youngman, R. Losick, *J. Bacteriol.* **170**, 3513 (1988).
10. J. Errington, *FEBS Lett.* **224**, 257 (1987).
11. ———, S. Rong, M. S. Rosenkrantz, A. L. Sonenshein, *J. Bacteriol.* **170**, 1162 (1988).
12. R. Farquhar and M. D. Yudkin, *J. Gen. Microbiol.* **134**, 9 (1988).
13. The *spoIVCB* nucleotide sequence was determined from two independently cloned copies of the gene. The first copy was obtained as reported (*9*) by cloning in *Escherichia coli* chromosomal sequences adjacent to a Tn*917* insertion in the gene. Using this DNA as a source, we subcloned into phage M13 a 400-bp Sac I–Sac I fragment containing the 3′ portion of the open reading frame and a 700-bp fragment extending from a Hpa I site located at codon 146 to a Pvu II site downstream of the open reading frame. We sequenced across both strands of the Sac I–Sac I fragment, and in the direction from Hpa I toward Pvu II of the Hpa I–Pvu II fragment. The second copy of *spoIVCB* was obtained by screening a size-selected library of Eco RI–Pst I fragments that had been cloned into the plasmid vector pUC18; a radioactively labeled oligonucleotide of 41 bases in length designed on the basis of the partial NH_2-terminal sequence of σ^K (*4*) was used as a probe. The Eco RI–Pst I DNA (3 kb) was used as a source in phage M13 subcloning experiments of (i) a 300-bp fragment extending from a Ssp I site upstream of the open reading frame to a Bst BI site at codon 69; a 300-bp fragment extending from the Bst BI site at codon 69 to the Sac I site downstream of the open reading frame; and a 700-bp fragment extending from the Hpa I site at codon 146 to a Pst I site located about 200 bp upstream of the open reading frame. The Ssp I–Bst BI fragment was sequenced on both strands; the Bst BI–Sac I fragment was sequenced in the direction from Bst BI toward Sac I; and the Hpa I–Pst I fragment was sequenced in the direction from Hpa I toward Pst I.
14. J. D. Helmann and M. J. Chamberlin, *Annu. Rev. Biochem.* **57**, 839 (1988).
15. P. Zuber, J. Healy, L. H. Carter, S. Cutting, C. P. Moran, R. Losick, *J. Mol. Biol.*, in press.
16. D. A. Siegele, J. C. Hu, W. A. Walter, C. A. Gross, *ibid.*, in press.
17. T. Gardella, H. Moyle, M. M. Susskind, *ibid.*, in press.
18. J. Errington, personal communication.
19. J. Zieg, M. Silverman, M. Hilmen, M. Simon, *Science* **196**, 170 (1977).
20. P. van de Putte, S. Cramer, M. Giphart-Gasslere, *Nature* **286**, 218 (1980).
21. J. Hicks, J. Strathern, I. Herskowitz, in *DNA Insertion Elements, Plasmids and Episomes*, A. I. Bukhari, J. A. Shapiro, S. L. Adhya, Eds. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1977), p. 457; J. Hicks, J. N. Strathern, A. J. S. Klar, *Nature* **282**, 478 (1979); J. N. Strathern and I. Herskowitz, *Cell* **17**, 371 (1979).
22. N. Hozumi and S. Tonegawa, *Proc. Natl. Acad. Sci. U.S.A.* **73**, 3628 (1976).
23. J. W. Golden, S. J. Robinson, R. Haselkorn, *Nature* **314**, 419 (1985).
24. P. J. Lammers, J. W. Golden, R. Haselkorn, *Cell* **44**, 905 (1986).
25. J. W. Golden, C. D. Carrasco, M. E. Mulligan, G. J. Schneider, R. Haselkorn, *J. Bacteriol.* **170**, 5034 (1988).
26. K. F. Chater, in *Microbial Development*, R. Losick and L. Shapiro, Eds. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1984), p. 89.
27. D. Kaiser, *ibid.*, p. 197.
28. Chromosomal DNA was prepared by extraction of cells (20 ml) harvested at the designated times during sporulation in Difco sporulation medium as described [K. Sandman, L. Kroos, S. Cutting, P. Youngman, R. Losick, *J. Mol. Biol.* **200**, 461 (1988)].
29. DNA from wild-type strain PY79 was extracted at hour 5 of sporulation (*28*) and digested with Sac I and Hind III. Fragments (1 to 1.5 kb) were gel-purified and mixed under ligation conditions with phage M13mp19 DNA that had been cut with Sac I and Hind III. (Only rearranged DNA was expected to give rise to a Sac I–Hind III of 1 to 1.5 kb, and hence our cloning strategy should have yielded *spoIVCB-spoIIIC* fusion DNA but not the unrearranged genes.) Next, *E. coli* strain TG1 was transformed with the ligation mixture and plated (~200 plaques per agar plate). Clones carrying the rearranged gene were identified by plaque hybridization with the radioactively labeled *spoIIIC* probe described for Fig. 4.
30. F. Sanger, S. Nicklen, A. R. Coulson, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5463 (1977).
31. Strains 298.4 and 55.3 were obtained from J. Errington [J. Errington and J. Mandelstam, *J. Gen. Microbiol.* **132**, 2967 (1986)]. The remaining strains are described in (*9*).

32. We thank A. L. Sonenshein for help with the manuscript. Supported by an NIH grant (R.L.), an NSF predoctoral fellowship (B.K.), a postdoctoral fellowship from the Helen Hay Whitney Foundation (L.K.), and travel support from the CNRS (P.S.).

22 November 1988; accepted 21 December 1988