JOURNAL OF BACTERIOLOGY, Nov. 1995, p. 6263–6275 Vol. 177, No. 21
0021-9193/95/$04.00+0

Adjacent and Divergently Oriented Operons under the Control of the Sporulation Regulatory Protein GerE in *Bacillus subtilis*

STEVEN ROELS† AND RICHARD LOSICK*

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

Received 12 May 1995/Accepted 28 August 1995

The DNA-binding protein GerE is the latest-acting regulatory protein in the mother cell line of gene expression during sporulation in *Bacillus subtilis*. GerE directs the transcription of several genes that encode structural components of the protein coat that encases the mature spore. We report on the identification and characterization of a cluster of additional genes whose transcription is dependent on GerE. These genes, which are located in the replication terminus region of the chromosome (181° on the genetic map), are arranged in adjacent and divergently oriented operons called *cgeAB* and *cgeCDE*, which consist of two and at least three genes, respectively. CgeD, the product of the second member of the *cgeCDE* operon, is strikingly similar to the product of a *B. subtilis* gene (*ipa-63d*) of unknown function and is similar at its amino terminus to certain glycosyl transferases involved in polysaccharide biosynthesis. Strains with mutations in the *cgeAB* and *cgeCDE* operons produce spores with altered surface properties, on which basis we propose that proteins encoded by these operons influence maturation of the outermost layer of the spore, perhaps by glycosylation of coat proteins at the spore surface.

Spore formation in the gram-positive bacterium *Bacillus subtilis* is governed by a highly ordered program of gene expression (reviewed in references 14, 34, and 58). At the septation stage of sporulation, when the sporangium is partitioned into mother cell and forespore compartments, the program of gene expression bifurcates into two distinct lines of gene expression. The forespore line of gene expression is relatively simple, being governed by the successive actions of two forespore-specific, RNA polymerase sigma factors, σ^F and σ^G. The mother cell line of gene expression, in contrast, is a hierarchical regulatory cascade involving the successive actions of two sigma factors and two auxiliary regulatory proteins in the alternating order σ^E, SpoIIID, σ^K, and GerE (66). Gene expression in the earliest step of the hierarchy is governed by σ^E, which directs transcription of several morphogenetic genes as well as the gene for the DNA-binding protein SpoIIID (3, 13, 29, 45, 56, 66). Next, SpoIIID, acting in conjunction with σ^E-containing RNA polymerase, stimulates the transcription of additional sporulation genes, including genes involved in the appearance of σ^K (22, 28, 31, 52). The σ^K factor, in turn, governs the expression of the third gene set in the regulatory cascade (7, 9, 49, 64–66), one member of which encodes GerE, the subject of the present communication. Acting in conjunction with σ^K-containing RNA polymerase, this 72-amino-acid DNA-binding protein stimulates transcription of the last temporal class of mother cell-expressed genes (6, 64–66). The GerE-controlled regulon includes several genes (*cotB*, *cotC*, *cotG*, and *cotX*) that encode structural components of the coat, the thick proteinaceous shell that encases the mature spore (12, 47, 63). In the cases of the P_x promoter of *cotX* and the *cotB* and *cotC* promoters, GerE has been shown to bind to specific sequences that overlap or lie upstream of the −35 regions of the promoters and to stimulate transcription by σ^K-containing RNA polymerase in vitro (64–66).

Cells with mutations in the gene (*gerE*) for GerE produce misformed spores that lack several coat proteins, are sensitive to lysozyme, and fail to germinate properly (15, 26, 38). On the other hand, spores from cells with multiple mutations in *cotB*, *cotC*, and *cotG* do not exhibit the *gerE* mutant phenotype (12, 47), nor do cells with mutations in *cotX* (63). These findings indicate the existence of additional, yet-to-be discovered GerE-controlled genes. Here we report the discovery and characterization of two divergently oriented operons, called *cgeAB* and *cgeCDE*, whose transcription is under GerE control. As is the case for the other known GerE-controlled genes, the effects of mutations in *cge* do not mimic the effects of a mutation in *gerE*. However, *cge* mutations do cause alterations in spore surface properties, suggesting that the *cge* products are involved in modification of the outer layer of the coat.

MATERIALS AND METHODS

Bacterial strains. Standard cloning procedures were with *Escherichia coli* TG1 or DH5α (48). The *B. subtilis* strains used in this study are listed in Table 1.

General methods. Selection for drug resistance was done on Luria-Bertani plates (48) containing ampicillin (100 μg/ml), chloramphenicol (5 μg/ml), or kanamycin (5 to 10 μg/ml). *B. subtilis* strains were induced to sporulate by growth in Difco sporulation medium, without antibiotics, in liquid cultures or on agar plates as described previously (49). Transformation and bacteriophage PBS1-mediated transduction of *B. subtilis* cells were carried out as described by Cutting and Vander Horn (10). Assays for heat and lysozyme resistance of spores as well as plate assays for germination were those described by Nicholson and Setlow (39).

DNA sequencing analysis was carried out by the method of Sanger et al. (51) with Sequenase 2.0 (U.S. Biochemical Corp.) according to the manufacturer's instructions. Analysis of DNA and protein sequence data was done by using the programs in the Sequence Analysis Software Package (version 7.3) from Genetics Computer Group, Inc. (11, 17).

Cloning of the *cge* locus. For the initial step in cloning *cge*, we took advantage

* Corresponding author. Mailing address: The Department of Molecular and Cellular Biology, The Biological Laboratories, Harvard University, 16 Divinity Ave., Cambridge, MA 02138. Phone: (617) 495-1774. Fax: (617) 495-9300. Electronic mail address: losick@biosun.harvard.edu.

† Present address: Myco Pharmaceuticals Inc., Cambridge, MA 02139.

TABLE 1. *B. subtilis* strains[a]

Strain	Genotype or description	Source and/or reference(s)
PY79	Prototrophic, SPβ^S	Laboratory stock (62)
SR293	*cgeD*::*kan* SPβ^S	This work
SR294	Δ(*cgeD-cgeB*)::*kan* SPβ^S	This work
SR297	Δ*cgeAB*::*kan* SPβ^S	This work
SR270	*cgeA*::pSR99(*cat*) SPβ^S	This work
SR402	*amyE*::(*cgeAB cat*) SPβ^S	This work
SR404	Δ(*cgeD-cgeB*)::*kan* *amyE*::(*cgeAB cat*) SPβ^S	This work
SR406	Δ*cgeAB*::*kan* *amyE*::(*cgeAB cat*) SPβ^S	This work
KS215(RL142)	*spoIVCB*::Tn*917*ΩHU215 SPβ^S	31, 50
KS450(RL325)	*gerE36* SPβ^S	49
PS138	*chr*(*sspC*)::pHC101(*cat*)	P. Setlow (5)
QB922	*gltA trpC2*	BGSC[b]

[a] All strains except strains PS138 and QB922 were derived from the prototrophic wild-type strain PY79 (62).
[b] BGSC, *Bacillus* Genetic Stock Center.

of the ability of *cotC* Pr1 (5′-aAcCCATATaTaCTCCTCC) to prime cDNA synthesis from *cgeA*-containing mRNA (65) (see Results). The sequence near the 5′ end of the *cotC* Pr1-derived *cge* extension product was determined as described below. We then designed and synthesized an oligonucleotide primer, the *cge* 30-mer (5′-GcTGTGTgAATGAGCTCTGAAAATGCACAG), corresponding to a portion of that sequence (see Fig. 3). (Note that the oligonucleotide does not match the reported sequence of *cge* at two positions because of inaccuracies in determining the sequence of the extension product.) The *cge* 30-mer and *cotC* Pr1 were then used to amplify a 106-bp fragment of *cge* by PCR with chromosomal DNA from PY79 as a template. One PCR produced a product of the expected size which was eluted from an agarose gel and used as the template for a second PCR, which produced only the 106-bp fragment. The product from this second reaction was purified by use of the Mermaid kit (Bio 101), and its ends were rendered blunt by treatment with T4 DNA polymerase (48). Finally, the fragment was ligated to *Sma*I-digested pER19 (43) to yield pSR99. The *cge* fragment in pSR99 was oriented such that the sequence corresponding to the *cge* 30-mer oligonucleotide was nearest the *Bam*HI site of the vector.

Chromosomal DNA adjacent to sequences represented in the PCR clone were obtained as follows. Competent cells of PY79 were transformed with pSR99, with selection for resistance to chloramphenicol. Transformants were expected to arise by integration of the plasmid by a single reciprocal (Campbell-like) recombination event between the insert in the plasmid and corresponding sequences in the chromosome. Chromosomal DNA was prepared from one such transformant, SR270, and was digested with *Hin*dIII, liberating a fragment that contained vector DNA and 2.3 kb of *cge* DNA extending from sequence present in pSR99 to the *Hin*dIII site within *cgeE* (at the left end of the region illustrated in Fig. 1). The DNA in the digest was diluted to 25 μg/ml, circularized by treatment with T4 DNA ligase, and used to transform *E. coli*, with selection for ampicillin resistance. The resulting plasmid, pSR104, was digested with *Pst*I and *Sac*I (to remove a fragment of *cge* DNA including sequence represented in the original PCR-derived clone), treated with T4 DNA polymerase to render the ends blunt, and recircularized with T4 DNA ligase to generate pSR127.

The 4.2-kb *cge* clone discussed in Results was obtained as follows. pSR127 was integrated into the chromosome of PY79 as described above to yield strain SR274. Chromosomal DNA from SR274 was digested with *Eco*RI, liberating a fragment that contained vector DNA and 4.2 kb of *cge* DNA extending from the *Hin*dIII site in *cgeE* to the *Eco*RI site in *orf181* (Fig. 1), and this fragment was recovered by circularizing and transforming *E. coli* as described above. The resulting *cge*-containing plasmid was designated pSR129.

Isolation of RNA. *B. subtilis* PY79 (*spo*$^+$), KS215 (*spoIVCB*::Tn*917*), KS450 (*gerE36*), and SR294 [Δ(*cgeD-cgeB*)::*kan*] were sporulated in liquid Difco sporulation medium, and samples (4 to 8 ml) were taken at the onset of sporulation and at 3, 5, 7, and 9 h of sporulation. In addition, PY79 was grown in liquid Luria-Bertani medium, and cells (20 ml) were harvested during exponential growth at an optical density of 0.7 at 600 nm. Cells were harvested by centrifugation in a microcentrifuge for 2 min or at 8,000 × *g* for 5 min. The cells were washed by suspension in 0.1 volume of cold STE (100 mM NaCl, 10 mM Tris-HCl [pH 8.0], 1 mM EDTA). The cells were pelleted in a microcentrifuge for 2 min, and the cell pellets were frozen in liquid nitrogen and stored at −70°C until needed. RNA was purified as described previously (8) with the following modifications. Cells were suspended in 0.8 ml of LETS buffer (10 mM Tris-HCl [pH 8.0], 50 mM LiCl, 10 mM EDTA, 1% sodium dodecyl sulfate) prewarmed to 75°C and then quickly transferred to 15-ml conical tubes containing 1 ml of acid-washed glass beads (<150 μm in diameter; Sigma) and 0.4 ml of phenol (pH 5.5) prewarmed to 75°C. The tubes were then vortexed vigorously for 3 min to lyse the cells. Chloroform (0.4 ml) was then added, the samples were vortexed again for 30 s, and the aqueous phase was separated by centrifugation at 8,000 × *g* for 10 min at 4°C. All subsequent manipulations were done at 4°C unless otherwise noted. The aqueous layer was extracted twice with 1 volume of a mixture (1:1, vol/vol) of phenol (pH 5.5) and chloroform. The RNA was precipitated by the addition of NaCl (to 0.3 M) and then 2.5 volumes of ethanol followed by incubation at −20°C overnight. The precipitate was collected by centrifugation and suspended in 300 μl of diethylpyrocarbonate-treated water. Next, the samples were extracted twice with a mixture of 100 μl of phenol (pH 4.3; Amresco) and 100 μl of chloroform and then once with 100 μl of chloroform. The RNA was precipitated by the addition of KCl (to 0.3 M) and then 2.5 volumes of ethanol followed by incubation at −20°C overnight. Finally, the precipitate was collected by centrifugation and suspended in 150 μl of diethylpyrocarbonate-treated water.

The RNA used in sequencing of the 5′ termini of *cgeAB* transcripts was purified as described previously (8, 45).

Northern (RNA) blot hybridization analysis. Northern blot analysis was carried out as described by Roels et al. (45) with the following modifications. The gels contained 1.25% agarose, and the hybridization and washing solutions were those described by Virca et al. (61). Two single-stranded probes were generated as follows. First, the *cge*-containing plasmid pSR129 was used as a template in

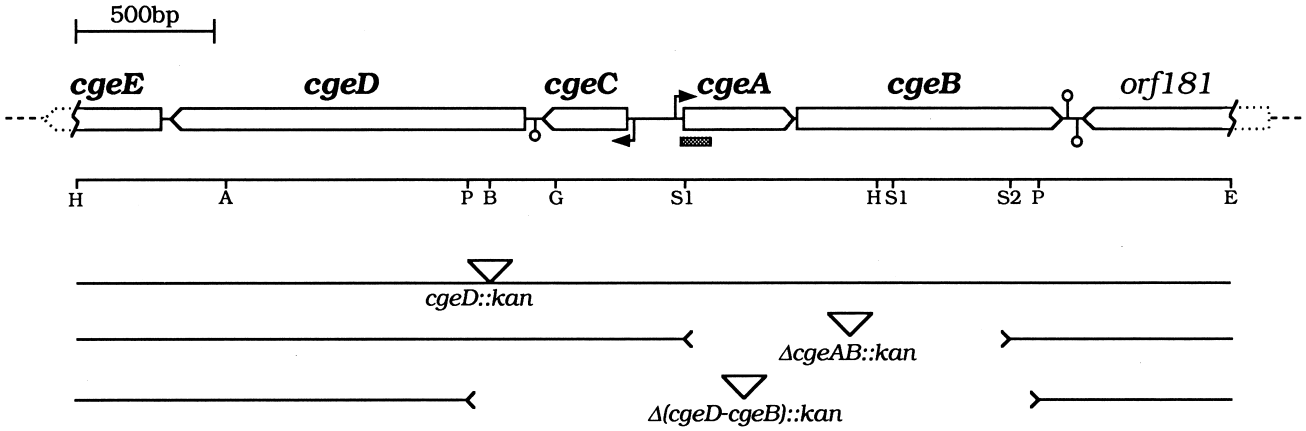


FIG. 1. Physical map of the *cge* region of the chromosome. A partial restriction map of the 4,162-bp *cge* clone is shown below an illustration depicting coding regions and other features in this region. The positions and orientations of open reading frames are indicated by boxes. The positions of the proposed transcription start sites for the *cgeAB* and *cgeCDE* operons are indicated by arrows. The bar near the beginning of *cgeA* corresponds to sequences cloned initially by PCR (see text), and potential transcription terminators are denoted by lollipop symbols. The bottom portion of the figure depicts the chromosome structures of three *cge* mutants (see Materials and Methods). Abbreviations: H, *Hin*dIII; A, *Aat*II; P, *Pst*I; B, *Bst*BI; G, *Bgl*II; S1, *Sac*I; S2, *Sac*II; E, *Eco*RI.

standard PCRs to amplify fragments internal to *cgeC* (with primers ol179 [5′-CCCCAGCTTACAACACTTGAGAG] and ol180 [5′-AAGGGAGAGGGAT GATTTGG]) or *cgeA* (with primers *cge* 30-mer and ol129 [5′-AATGCTGATT GTAGTGCC]). The PCR product from each reaction was purified by ultrafiltration with Microcon-100 concentrators (Amicon) and by agarose gel electrophoresis. The PCR products were used in turn as templates in labeling reactions with only a single oligonucleotide primer (ol179 for *cgeC* or ol129 for *cgeA*) to generate strand-specific probes. The components of these reaction mixtures were as follows: 50 mM KCl; 10 mM Tris-HCl (pH 9.0); 1.5 mM $MgCl_2$; 0.1% Triton X-100; 16.6 μM each dATP, dGTP, and dTTP; 0.83 μM [α-^{32}P]dCTP (50 μCi); 20 ng of template DNA; 50 pmol of primer; and 5 U of *Taq* polymerase (Promega) (final volume, 20 μl). The following cycles were performed: 1 cycle of 95°C for 2 min and then 20 cycles of 95°C for 30 s, 45°C for 1 min, and 72°C for 1 min. The probe DNA was then ethanol precipitated and suspended in water. About half of a probe reaction mixture was used in each Northern hybridization.

Mapping and sequencing the 5′ termini of *cge* transcripts. The 5′ termini of *cge* transcripts were identified by primer extension analysis with the primers ol179 and ol125 (5′-CATCTTGGTTTTCTTCTCGATCC) for *cgeC* and ol124 (5′-GTTGCAGTTGGAAATAAAGGGC) and ol129 for *cgeA*. The primers were labeled at their 5′ ends with [γ-^{32}P]ATP and T4 polynucleotide kinase (Bethesda Research Laboratories) as described by Sambrook et al. (48). One picomole of each labeled primer was incubated with 5 μg of RNA in 5 μl of hybridization buffer (50 mM Tris-HCl [pH 7.7], 100 mM KCl) at 90°C for 1 min and then at 47°C for 20 min. The samples were chilled on ice, and 4 μl of modified 2.5× reverse transcriptase buffer (62.5 mM Tris-HCl [pH 7.7], 25 mM KCl, 25 mM $MgCl_2$, 2.5 mM dithiothreitol, 1 U of RNAguard [Pharmacia] per μl, 2.5 mM each deoxynucleoside triphosphate [dNTP]) was added. The samples were then warmed to 47°C, and 1 μl (20 U) of Moloney murine leukemia virus reverse transcriptase (Pharmacia) or avian myeloblastosis virus reverse transcriptase (Pharmacia) was added. Extension reaction mixtures were incubated at 47°C for 1 h, after which 6.7 μl of loading dye containing 95% formamide and 20 mM EDTA was added. The products of the extension reactions were resolved on 6% polyacrylamide gels containing 8 M urea.

A partial sequence of the *cotC* Pr1-derived *cgeAB* extension product was obtained by a modification of the above-described protocol. Twenty micrograms of RNA was incubated with 20 pmol of 5′-end-labeled *cotC* Pr1 in 15 μl of hybridization buffer at 90°C for 2 min and then at 45°C for 20 min. Three microliters of this solution was used in 10-μl extension reaction mixtures containing (final concentrations) 22 mM Tris-HCl [pH 7.7]; 55 mM KCl; 10 mM $MgCl_2$; 1 mM dithiothreitol; 1 U of RNAguard per μl; 0.4 mM each dNTP; 1 mM ddGTP, ddATP, ddTTP, or ddCTP; and 1 U of avian myeloblastosis virus reverse transcriptase (Life Sciences) per μl. Extension reaction mixtures were incubated at 45°C for 30 min, and then 0.5 μl of a solution of each dNTP at 25 mM was added and the reaction mixtures were incubated for an additional 10 min at 45°C.

Construction of *cge* mutant strains. Three insertion and deletion mutations in *cge*, illustrated in Fig. 1, were generated as follows. A 1.4-kb *Hae*II fragment of pUK19 that confers resistance to kanamycin was purified, and its ends were rendered flush with T4 DNA polymerase. Plasmid pUK19 (a gift from Bill Haldenwang) was constructed by cloning a 1.5-kb *Cla*I fragment from pJH1 (33, 60) containing a kanamycin resistance gene into *Nar*I-digested pUC19 such that the *kan* gene was oriented toward the polylinker. The *kan*-containing fragment was then ligated to blunted backbone fragments of pSR129 resulting from digestion with *Bst*BI, *Pst*I, or both *Sac*I and *Sac*II (Fig. 1), yielding pSR153, pSR157, and pSR160 respectively. These plasmids were then linearized and used to transform competent PY79 cells, with selection for kanamycin resistance. Southern analysis demonstrated that in transformants SR293, SR297, and SR294, the wild-type *cge* region had been replaced by double (marker replacement) recombination with the disrupted *cge* sequences present on pSR153, pSR157, and pSR160, respectively. The *kan* gene in SR293 (and pSR153) is oriented in the same direction as *cgeCDE*, whereas the *kan* gene in SR297 and SR294 (and pSR157 and pSR160) is oriented toward *orf181*.

Construction of strains harboring *cgeAB* at the *amyE* locus. Strains harboring a copy of *cgeAB* at the *amyE* locus were constructed as described by Cutting and Vander Horn (10). A 2.4-kb *Bgl*II-*Eco*RI fragment of pSR129 containing the *cgeAB* operon was cloned into *Bam*HI-*Eco*RI-cut plasmid pDG364 (a gift from Patrick Stragier) between the two halves of the *amyE* gene to generate plasmid pSR179. Competent cells of strains PY79, SR294, and SR297 were then transformed with linearized pSR179, with selection for resistance to chloramphenicol, to generate SR402, SR404, and SR406, respectively. The *cgeAB* sequences were expected to integrate at the *amyE* locus as a consequence of a double recombination event between *amyE* sequences of pSR179 and the corresponding sequences in the chromosome. Disruption of the chromosomal *amyE* gene in SR402, SR404, and SR406 was verified by demonstrating that these strains had an Amy$^-$ phenotype.

Nucleotide sequence accession number. The DNA sequence of the *cge* region of the chromosome as reported here has been deposited in the GenBank database with the accession number U18421.

RESULTS

Identification and cloning of the *cge* locus. The existence of an additional gene under the control of GerE was discovered fortuitously during the course of experiments designed to map the 5′ terminus of the transcript for the coat protein gene *cotC* (65). A synthetic oligonucleotide primer, *cotC* Pr1, was used to generate extension products with RNA from late-sporulating cells as a template. Nucleotide sequence analysis of the extension products showed that one of them, an extension product of 127 nucleotides (nt), did not correspond to *cotC* but was instead copied from a distinct transcript that, like the *cotC* transcript, appeared late during sporulation and was present at only a very low level in *gerE* mutant cells (46, 65). We refer to the locus that specifies this transcript as *cge*, for controlled by GerE.

To clone *cge*, we first determined the partial sequence of the 127-nt Pr1-generated *cge* extension product and synthesized a 30-nt primer corresponding to a portion of the extension product sequence (see Materials and Methods). The 30-nt primer and Pr1 were then used to amplify a 106-bp fragment of *cge* DNA by PCR with *B. subtilis* DNA as a template. The PCR product was cloned, and the resulting plasmid was used, in turn, to clone a 4.2-kb fragment of DNA in the *cge* region of the chromosome by a two-step chromosomal walk as described in Materials and Methods. Figure 1 shows a restriction map of the 4.2-kb *cge*-containing clone. The position of the sequence represented in the PCR-derived clone is also indicated.

Mapping of the *cge* locus. The chromosomal location of *cge* was determined by two methods. First, radioactively labeled *cge* DNA was used as a hybridization probe against an ordered yeast artificial chromosome (YAC) library of the *B. subtilis* 168 chromosome (53), an analysis that was kindly carried out for us by Pascale Serror. The probe was found to hybridize specifically to YAC clone 15-19, and not to either of the overlapping YAC clones 10-131 and 15-37, indicating that *cge* was located at about 181° on the genetic map (Fig. 2), near the position of the SPβ prophage and the replication terminus (1, 53). Other genes in this region include *gltA* (177°), which is also located within the region specific to YAC 15-19, and *sspC* (182°), which is located outside SPβ and within the region of overlap between YAC clones 15-19 and 10-131. This positions *sspC* between SPβ and *cge*.

In confirmation and extension of the YAC library mapping, phage PBS1-mediated generalized transduction demonstrated that *gltA* and sequences adjacent to *sspC* were cotransduced at frequencies of 52 and 98%, respectively, with a drug resistance gene inserted at *cge* (*cgeD*::*kan*). These results, together with the relatively weak genetic linkage between *gltA* and *sspC* reported previously (5% cotransduction) (5), are consistent with the gene order *sspC-cge-gltA*, with *cge* being located somewhat closer to *sspC* than to *gltA*.

Nucleotide sequence of the *cge* region of the chromosome. The nucleotide sequence of the 4.2-kb (precisely 4,162-bp) *cge* clone is shown in Fig. 3. Inspection of the sequence revealed four complete and two partial open reading frames which, for reasons indicated below, were designated *cgeE*, *cgeD*, *cgeC*, *cgeA*, *cgeB*, and *orf181* (Fig. 1). *cgeA* and *cgeB* are oriented to the right as drawn in Fig. 1, whereas the remaining open reading frames are oriented to the left. The 5′ terminus of the *cge* transcript identified by primer extension with *cotC* Pr1 is located just upstream of *cgeA* (nucleotide 2166 in Fig. 3), and this transcript extends into *cgeA*. A pair of inverted repeat sequences (possible transcription terminators) can be found between *cgeB* and *orf181*, and another inverted repeat is located just downstream of *cgeC* (Fig. 1 and 3). The sizes of the

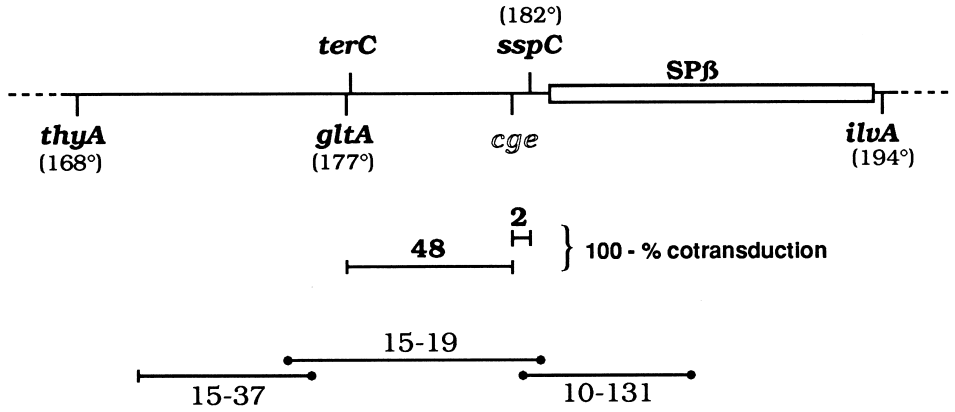


FIG. 2. Map position of *cge* on the *B. subtilis* 168 chromosome. The approximate position of the SPβ prophage and genes near 180° on the genetic map (1) are indicated. Also shown are the genetic distances, on the basis of phage PBS1-mediated generalized transduction, between *cge* and the nearby genes *sspC* and *gltA*. The lines at the bottom of the figure depict the inserts in YAC clones in the *cge* region (adapted from reference 53).

cge open reading frames and the predicted molecular masses of the polypeptides they are capable of encoding are listed in Table 2. *cgeB* is unusual in that none of the potential initiation codons near the beginning of the open reading frame are preceded by a compelling match to a Shine-Dalgarno sequence. One possibility is that translation of *cgeB* is facilitated by its close proximity (a separation of 6 nucleotides) to the end of *cgeA*. Alternatively, translation may initiate at one of two downstream sites as indicated in Table 2. Similarly, the proposed translation start site for *cgeE* is preceded, at a distance of 14 codons, by an alternative initiation codon which lacks a compelling Shine-Dalgarno sequence but which overlaps with the end of *cgeD* (Fig. 3). This raises the possibility that translation of *cgeE* originates at the upstream site and is facilitated by translation through *cgeD*. One noteworthy feature of CgeC is the presence of a stretch of 12 hydrophobic amino acids, mostly leucines, near its amino terminus (Fig. 3).

None of the amino acid sequences of the predicted *cge* gene products exhibited significant similarity to protein sequences represented in GenBank and related databases, except for CgeD, whose sequence was highly similar to that of the predicted product of a recently identified *B. subtilis* open reading frame of unknown function called *ipa-63d* (18) (see Discussion). The *ipa-63d*-encoded protein and the first 258 residues of CgeD were 41% identical at the amino acid level, with the similarity extending to 62% when conservative changes were considered (Fig. 4A).

Interestingly, CgeD and the product of *ipa-63d* are also similar in sequence to several bacterial proteins that are known or are inferred to be nucleotide sugar transferases involved in the biosynthesis of extracellular polysaccharides. The most compelling similarities involve residues in two closely spaced regions within the amino-terminal 111 amino acids of CgeD. A partial alignment between these regions of CgeD and the corresponding regions of proteins exhibiting the greatest degree of similarity to CgeD is shown in Fig. 4B. The sequences shown include those for the ExoO, ExoW, and ExoU glycosyl transferases of *Rhizobium meliloti*, which are required for biosynthesis of the repeating octasaccharide subunit of succinoglycan, the exopolysaccharide involved in host nodule invasion (19, 20). The regions shown in Fig. 4 are also those in which the ExoO, -W, and -U proteins (as well as the other proteins shown) exhibit the greatest similarity to each other, and these regions are thus postulated to be important for enzymatic activity (19). Also included in the alignment are partial amino acid sequences of the following proteins: (i) KfaB, a putative glycosyl transferase from *E. coli* (44); (ii) VirB, a protein from *Vibrio anguillarum* that is involved in the production of a lipopolysaccharide antigen found on the surface of the flagellar sheath (40); (iii) the product of ORF5 in the *lsg* locus of *Haemophilus influenzae*, a locus required for lipopolysaccharide synthesis (37); (iv) the product of a partially sequenced *E. coli* gene of unknown function called *yibD* (54); (v) the product of an open reading frame downstream of the *hetA* gene in an *Anabaena* sp. (23), whose similarity to ExoO had been noted previously (19); (vi) the product of another recently identified *B. subtilis* gene of unknown function called *ipa-56d* (18); and (vii) NodC, a protein from *R. meliloti* that is involved in production of Nod factors, which are oligomers of *N*-acetylglucosamine that are required for the initiation of host nodule development (2, 25). As shown in Fig. 4, a somewhat degenerate consensus sequence, involving 56 residues, was deduced from examination of the 11 aligned sequences (excluding CgeD). The sequence of CgeD was found to agree with the consensus sequence at 40 (71%) of those positions, and at four of the remaining positions at least two other proteins had the same amino acid as CgeD. Moreover, the sequence of CgeD agreed with the consensus sequence at 12 of 15 positions where a particular amino acid was found in at least 9 of the 11 aligned sequences.

***cge* consists of two adjacent and divergent GerE-dependent transcription units, *cgeAB* and *cgeCDE*.** To determine the distribution and regulation of transcription units in the *cge* region, we carried out Northern hybridization analysis with blots containing RNA harvested from wild-type and mutant cells during vegetative growth and at various times during development. As noted above, the *cge* transcript identified by primer extension with *cotC* Pr1 originates upstream of *cgeA*, indicating that *cgeA* is transcribed under the control of GerE. When an internal fragment of *cgeA* was used to generate a single-stranded probe, two transcripts of approximately 1.4 and 1.0 kb were detected which accumulated late during sporulation (beginning at about 5 h) but which were present at greatly reduced levels in *gerE* mutant cells (Fig. 5A). Both transcripts appear to originate from a promoter just upstream of *cgeA* (46) (see below), and the size of the larger transcript is consistent with that expected if *cgeA* and *cgeB* were cotranscribed. (The 1.0-kb transcript, then, presumably results from transcription termination within *cgeB* or from processing/degradation of the more abundant, and presumably full-length, 1.4-kb transcript.) Moreover, *cgeB*

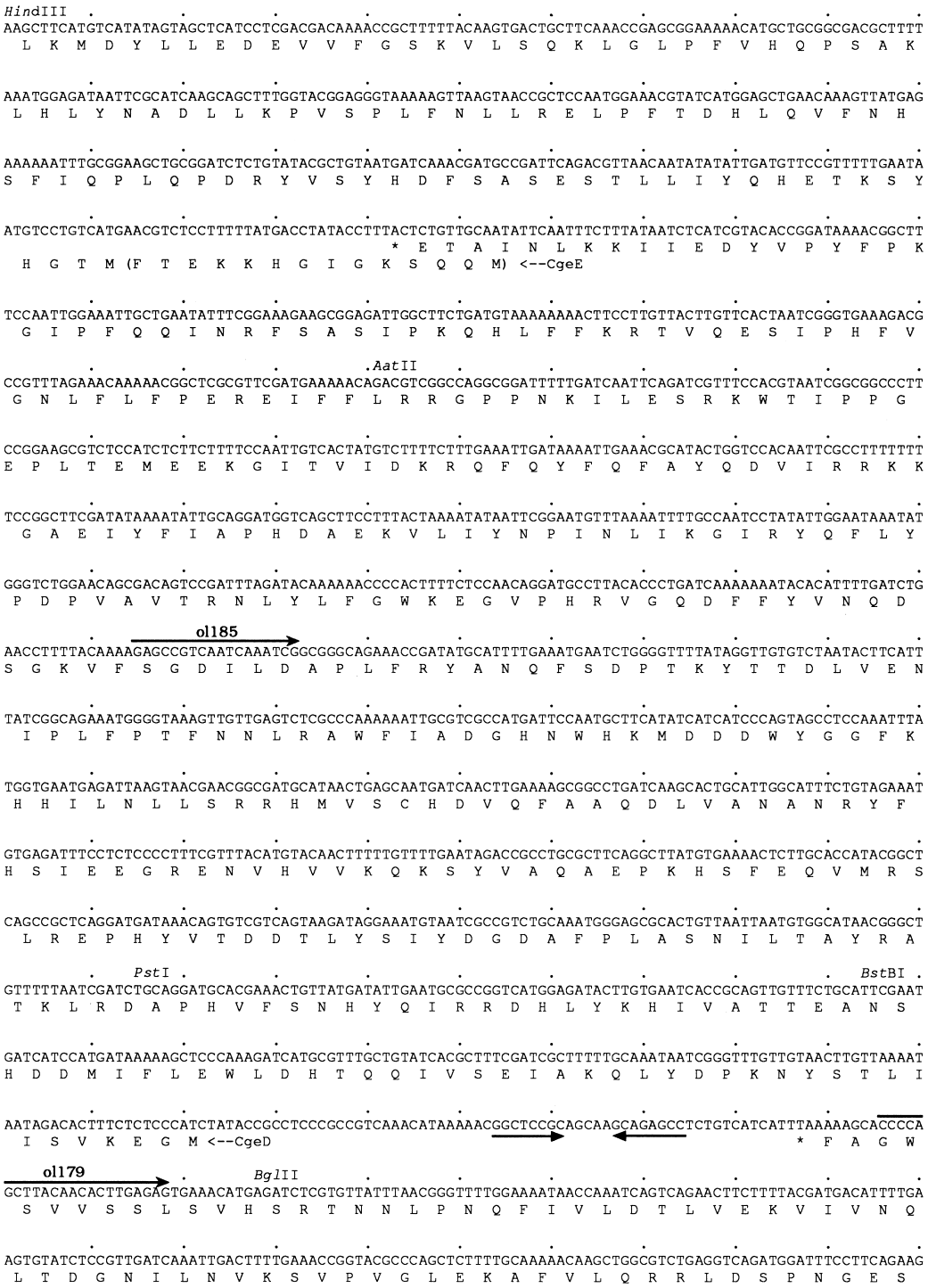


FIG. 3. Nucleotide sequence of the *cge* locus. The nucleotide sequence of the cloned *cge* fragment, with the first nucleotide position corresponding to the left end of the fragment as drawn in Fig. 1, is shown. The sequence shown is that of the nontranscribed strand with respect to the *cgeAB* open reading frames, but the sequences of both strands in the vicinity of the *cge* promoters are shown. The amino acid sequences of the predicted translation products are shown above (for *cgeA* and *cgeB*) or below (for the remaining genes) the DNA sequence. A stretch of 14 amino acids shown in parentheses at the amino terminus of CgeE would result from the use of an alternative initiation codon which precedes the predicted initiation codon but for which there is no compelling Shine-Dalgarno sequence. The positions of sequences corresponding to oligonucleotide primers are indicated by arrows. Inverted repeat sequences (possible transcription terminators) between *cgeB* and *orf181* and downstream of *cgeC* are indicated by pairs of arrows. The proposed transcription start sites for *cgeAB* and *cgeCDE* are denoted by large arrows, and regions corresponding to the −10 and −35 regions of the *cge* promoters are also indicated. Sequences between the proposed *cge* transcription start sites which are similar to the consensus sequence for GerE binding (RWWTRGGY--YY, where R refers to purine, W refers to T or A, and Y refers to pyrimidine) (64) are delimited by brackets. The positions of the restriction sites shown in Fig. 1 are also indicated.

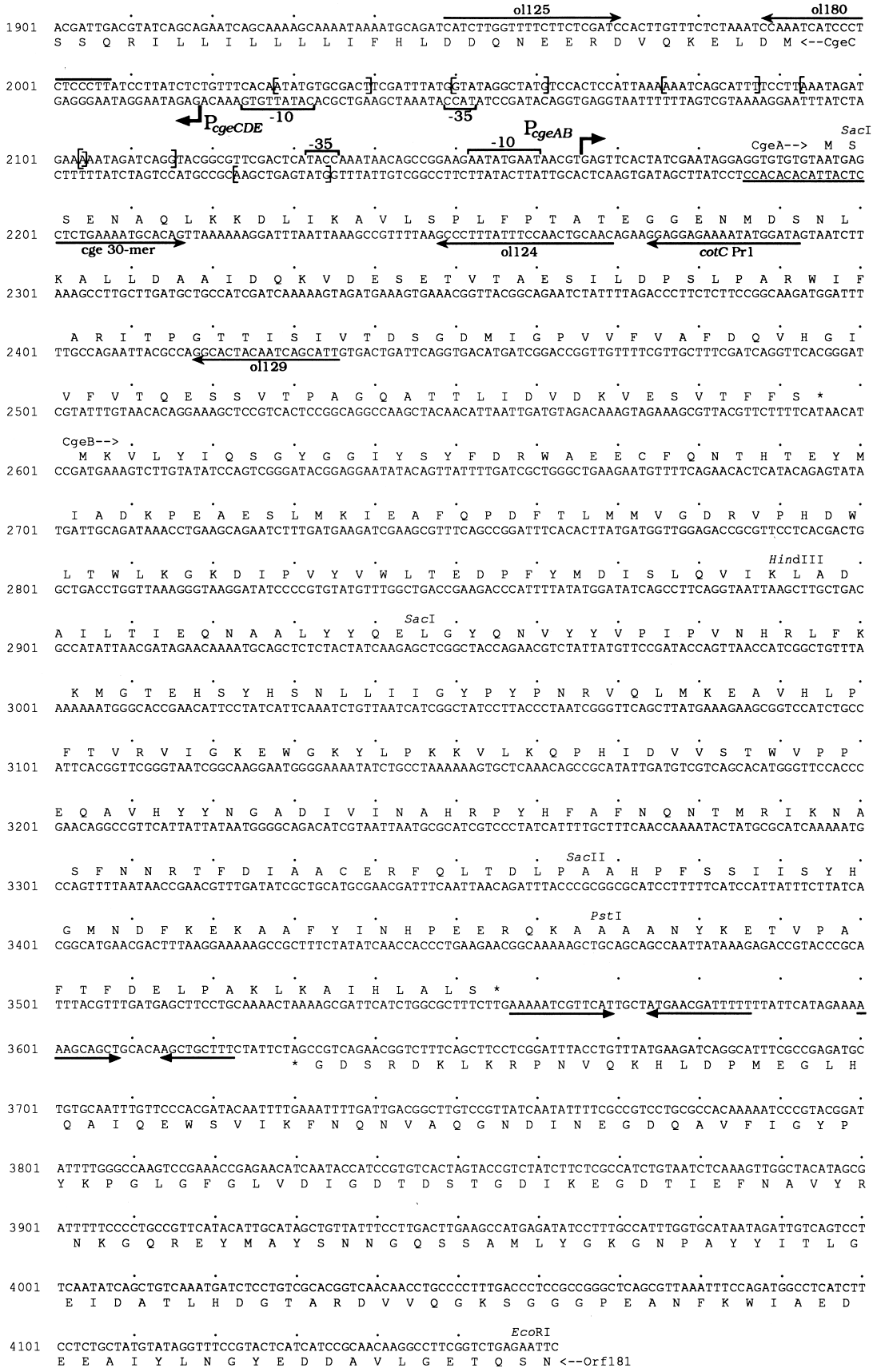


FIG. 3—*Continued.*

TABLE 2. Genes in the *cge* region of the chromosome

Gene or open reading frame	No. of codons	Predicted size (kDa)
cgeA	133[a]	14.2
cgeB	317[b]	36.9
cgeC	101	11.4
cgeD	426	50.1
cgeE	≥104[a]	
orf181	>177	

[a] *cgeA* and *cgeE* have alternative translation start sites (initiation codons) which precede the proposed start sites by 9 and 42 nucleotides, respectively, but for which there is no compelling Shine-Dalgarno sequence.

[b] The start codon that would correspond to the beginning of the *cgeB* open reading frame listed here (and depicted in Fig. 3) is not preceded by a compelling Shine-Dalgarno sequence. Two alternative start codons, preceded by sequences with successively better matches to a Shine-Dalgarno sequence, would give rise to polypeptides of 274 or 241 amino acids, respectively. Also, the *cgeB* open reading frame ends in a UGA stop codon which is read at low frequency as a Trp codon in *B. subtilis* (35). Translation through that stop codon would extend the open reading frame by 13 codons.

does not appear to have an additional mode of expression, because no additional transcripts were detected when a fragment of DNA extending from the *Sac*I site at the beginning of *cgeA* to the *Sac*I site within *cgeB* was used as a probe in Northern hybridization analysis (46). Thus, *cgeA* and *-B* constitute an operon, hereafter referred to as *cgeAB*, whose expression is under the control of GerE.

Inspection of sequences near the proposed *cgeAB* promoter (see below) led us to suspect that a second GerE-controlled promoter was located upstream of, and directed toward, *cgeC*. To investigate this possibility, Northern hybridization was carried out with a single-stranded probe generated from an internal fragment of *cgeC*. Two principal transcripts, an abundant 0.4-kb transcript and a much less abundant 2.7-kb transcript, were detected. Like the *cgeAB* transcripts, these transcripts accumulated late during sporulation (beginning at about 5 h) and were present at only very low levels in *gerE* mutant cells (Fig. 5B). The size of the smaller transcript coincides approximately with the size (303 bp) of the *cgeC* open reading frame. Assuming that the two transcripts have the same transcription start site (see below), the larger transcript would extend about 0.7 kb beyond the region represented in the *cge* clone and could encompass *cgeC*, *cgeD*, and presumably *cgeE*. Thus, *cgeC*, *-D*, and *-E* constitute a second GerE-controlled operon, hereafter referred to as *cgeCDE*, located just upstream of, and oriented divergently from, *cgeAB*.

The relatively low abundance of the 2.7-kb *cgeCDE* transcript may be due to transcription termination between *cgeC* and *cgeD*, a possibility suggested by the presence of an inverted repeat in this region. If this is the case, then the lower level of GerE-dependent expression of *cgeD* and *cgeE* could be due to attenuation of transcription originating from the promoter upstream of *cgeC*. Alternatively, the 3′ portion of the *cgeCDE* polycistronic mRNA could be less stable than the 5′ region containing *cgeC*. In either case, it would appear that the distal *cgeCDE* products are needed at lower levels or that the principal mode of *cgeD* and *cgeE* expression may not be related to expression from the promoter upstream of *cgeC*. To investigate whether *cgeD* is expressed at other times from one or more additional promoters, Northern blots were hybridized with a single-stranded *cgeD* probe. The results were the same as those obtained with the *cgeC* probe except that the 0.4-kb transcript, which presumably encompasses only *cgeC*, was no longer detected (46). Thus, *cgeD* does not appear to have an alternative mode of expression, at least not at the times or under the conditions examined. It remains possible, however, that *cgeE* is expressed from another promoter in addition to the promoter upstream of *cgeC*.

Because GerE affects promoter recognition by RNA polymerase containing the sporulation sigma factor σ^K, and because the *gerE* gene is itself transcribed by σ^K RNA polymerase (7, 65), it was expected that the expression of *cgeAB* and *cgeCDE* would also be dependent on σ^K. To test this, we examined *cge* transcript levels during sporulation in a strain containing a mutation in the gene, *spoIVCB*, that encodes the amino-terminal half of σ^K (*spoIVCB* and *spoIIIC* are joined by a developmentally regulated DNA rearrangement to generate *sigK*, the composite gene for σ^K [30, 57]). The *cge* transcripts, which were greatly reduced in abundance in *gerE* mutant cells, were absent from *spoIVCB* mutant cells, in agreement with expectations (Fig. 5).

Mapping of the 5′ termini of *cge* mRNAs. Figure 6 shows the results of high-resolution mapping of the 5′ termini of the *cgeAB* and *cgeCDE* transcripts. In each case, a pair of oligonucleotide primers was used to prime cDNA synthesis from RNA isolated from sporulating wild-type cells or, as a control, from sporulating mutant cells in which much of the *cge* region (from the *Pst*I site in *cgeD* to the *Pst*I site in *cgeB*) had been deleted. The *cgeA* Pr1 and Pr2 primers yielded predominantly extension products of 104 and 270 nt, respectively, which corresponded to a 5′ terminus located 30 nt upstream of the proposed start codon of *cgeA*, in agreement with the position suggested by the 127-nt extension product obtained previously with the *cotC* Pr1 primer. The *cgeC* Pr1 and Pr2 primers yielded predominantly extension products of 68 and 321 nt, respectively, which corresponded to a 5′ terminus located 25 nt upstream of the proposed start codon for *cgeC*. (The extension products of greater than 68 nt that were obtained with *cgeC* Pr1 are attributable to a lack of primer specificity. This is demonstrated by their presence in a control reaction with RNA harvested from a strain in which the region corresponding to the primer had been deleted, as well as by the absence of corresponding extension products when *cgeC* Pr2 was used [Fig. 6B].) Consistent with the finding that accumulation of *cge* transcripts was dependent on σ^K, the sequences centered approximately 10 and 32 nt upstream of the proposed transcription start sites for both *cgeAB* and *cgeCDE* were similar to those found at those positions in promoters for genes known to be transcribed by σ^K RNA polymerase (Fig. 7).

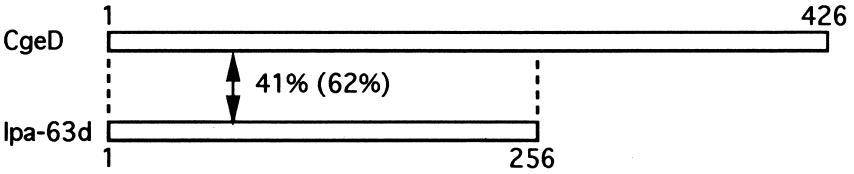
The dependence of *cge* expression on GerE suggests that one or more GerE binding sites would be located near the *cge* promoters, analogous to those found near the promoters for other sporulation genes whose expression is influenced by GerE (64, 65). Indeed, several sequences that are similar to the consensus sequence for a GerE binding site (64) are present in the *cge* promoter region (see Fig. 3 and its legend). However, the degree of degeneracy in the GerE binding site consensus sequence and the absence of sites near the promoters that match the consensus sequence perfectly make it difficult, on the basis of sequence analysis alone, to speculate as to which, if any, of the *cge* sequences serve as binding sites for GerE. Nevertheless, it is worth noting that a pair of sequences similar to the consensus sequence can be found extending from position −32 to −43 and from position −58 to −69 relative to the proposed transcription start site for *cgeCDE* (Fig. 3). The positions of these sequences relative to the transcription start site, and the spacing between them, are reminiscent of those found for sequences within the GerE-binding regions of the *cotB* and *cotX* promoters (64, 65). Also, another potential GerE-binding sequence can be found extending from position −32 to −43 relative to the transcription start site for *cgeAB*, the same

position and orientation as for one of the sequences noted above that overlaps the −35 region of the *cgeCDE* promoter.

Effects of *cge* mutations on spore surface properties. To assess the possible role of *cge*-encoded proteins in spore formation, we constructed strains harboring deletion or insertion mutations in *cge* genes (Fig. 1; see Materials and Methods). All of the mutant strains were able to produce normal yields of optically refractile, heat-resistant spores (46). (This result was expected because *gerE* mutant cells, while exhibiting defects in the composition and appearance of the spore coat as well as in some spore properties, are not blocked in spore formation per se.) The *cge* mutant spores, in contrast to those produced by *gerE* mutants, were resistant to lysozyme treatment and were not substantially impaired in germination (46). However, the spores produced by *cge* mutants did exhibit altered surface properties. In particular, strains harboring a deletion of most of the *cgeAB* operon (Δ*cgeAB*::*kan*) produced spores that had a tendency to clump, formed compact pellets (relative to those formed by wild-type spores) when centrifuged, and adhered to the surface of glass or plastic tubes. In contrast, the spores produced by strains harboring an insertion in *cgeD* (*cgeD*::*kan*) were difficult to harvest by centrifugation, and the pellets formed were much less compact than those formed by wild-type spores under the same conditions. Strains harboring a large deletion extending from within *cgeD* to near the end of *cgeB* [Δ(*cgeD-cgeB*)::*kan*] produced spores that behaved like those produced by Δ*cgeAB* mutants. In each case, these spore properties were observed even after purification of spores (39) by application over a Renografin density gradient or by lysozyme treatment followed by extensive washing.

To verify that the phenotypes observed for the Δ*cgeAB* and Δ(*cgeD-cgeB*) mutants were attributable to absence of one or both *cgeAB* gene products, we carried out complementation analysis. A copy of the *cgeAB* operon was placed at the *amyE* locus of a wild-type strain and two of the *cge* mutants (see Materials and Methods). The presence of a second copy of

A




```
CgeD       1 MGEKVSIILTSYNKPDYLQKAIESVIQQTHDLWELFIMDDHSNAETTAVIHKYLHDRRIQYHNS.FVHPADRLKTARYATLINSALPFADGDYISYLTDD 99
Ipa-63d    1 .MPKVSVIMTSYNKSDYVAKSISSILSQTFSDFELFIMDDNSNEETLNVIRPFLNDNRVRFYQSDISGVKERTEKTRYAALINQAIEMAEGEYITYATDD 99
CgeD     100 TVYHPERLSRMVQEFSHKPEAQAVYSKQKVVHVNERGEEISHFYRNANAVLDQAAFQVDHCSVMHRRSLLNLIHHKFGGYWDDDMKHWNHGDAIFWARLN 199
Ipa-63d  100 NIYMPDRLLKMVRELDTHPEKAVIYSASKTYHLNENRDIVKETVRPAAQVTWNAPCAIDHCSVMHRYSVLEKVKEKFGSYWDESPAFYRIGDARFFWRVN 199
CgeD     200 NFTPFLPINEVLDTTYKTPDSFQNAYRFLPADLIDGSFVKG..SDQNVYFFDQGVRHPVGEKWGFLYLNRTVAVPDPYLFQYRIGKILNIPNYILVKEAD 297
Ipa-63d  200 HFYPFYPLDEELDLNYITDQSIH..FQLF..ELEKNEFVRNLPPQRNCRELRESLKKLGMG..................................... 256
```

B

```
                *S***  YN  D *   A* S**AQT    *E****DD S D          R* **  E N GP  ARN G*    A G ** F*D DD *  P +* +**

   CgeD      5 VSIILTSYNKPDYLQ.KAIESVIQQTHD.LWELFIMDDHSNA  44    66 VHPADRLKTARYATLI.NSALP..FADGDYISYL.TDDTVYHPERLSRMV 111

 Ipa-63d     4 VSVIMTSYNKSDYVA.KSISSILSQTFS.DFELFIMDDNSNE  43    66 SGVKERTEKTRYAALI.NQAIE..MAEGEYITYA.TDDNIYMPDRLLKMV 111
   ExoO     12 VTFVVAAYNSADTIV.RAIESALAQEGV.TVEVVVVDDCSAD  51    64 RVRLIALDRNRGPGGARNAGIG..AARGRWIAVLDSDDTV.RPDRLRRMI 110
   ExoW      4 LTVVIPYYQKEPGILRRALASVFAQTLE.DFHVLVIDDESPY  44    60 RITVIR.QPNGGPGGARNTGLDNVPADSDFVAFLDSDDVW.TPDHLLNAY 107
   ExoU      9 VCIIISAKNAADTIA.RAVASALA.EPE.AAEVVVIDDGSTD  47    62 RLNVVRFEENRGPAAARNHAIA..ISHSPLIGVLDADDFF.FPGRLGQLL 108
   KfaB    266 VSIIMSVFNSEDTIA.YSLHSLLNQTYE.NIEILVCDDCSSD 305   320 RVKVYSSRKNQGPYNIRNELIK..KAHGNFITFQDADDLS.HPERIQRQV 366
   VirB     11 VSLCILTFNHEKYIT.KAINSCLAQSYS.NIEIIIVDNNSSD  50    68 EIKLFDLEHNTYPSHGFNYALK..KSQGEYVSLFSGDDTL.CLNKVERQI 114
Lsg_Orf5     2 LSIIVPSYNRKAEVP.ALLESLTQQTSS.NFEVIIVDDYSKE  41    52 PVTVIRNETNQGAAESRNIGAR..ASKGDWLLFLDDDDRF.MPEKCEKIL  98
   YibD      8 LSVIIPLYNAGDDFR.TCMESLITQTWT.ALEIIIINDGSTD  47    61 PHVRLLHQANAGASVARNRGIE..VATGKYVAFVDADDEV.YPTMYETLM 107
Anab.Orf    49 ISVIISNYNYARYLS.RAINSVLAQTHS.DIEIVIVDDGSTD  88   103 DKIKPIFQANQGQGGAFNAGFA..AATGEVVAFLDADDVW.KPHKLQRIV 149
 Ipa-56d     3 ISIVIVTYNRIPALC.ELLESISRQTLM.PYEIIIVNDAGES  42    54 PIAVINLEKNSGHVAARNAGVK..EASGDCIMLCDDDDFF.TPGHIEKMA 100
   NodC     51 VDVIVPSFNEDPGILSACLASIADQDYPGELRVYVVDDGSRN  92   108 RFSFILLPENVGKRKAQIAAIG..QSSGDLVLNVDSDSTI.AFDVVSKLA 154
```

FIG. 4. Sequence comparison of CgeD with the product of *B. subtilis ipa-63d* and with regions of certain bacterial glycosyl transferase enzymes. (A) Alignment, generated by using the GAP program of the Genetics Computer Group sequence analysis package (see Materials and Methods), between the predicted *ipa-63d* product and the first 258 amino acids of CgeD. Identical residues are connected by vertical lines, and similar residues are connected by two dots. Over the region of comparison, 41% of the paired residues were identical in the two proteins, and an additional 21% were similar. (B) Alignment between two regions at the amino terminus of CgeD and corresponding regions of high similarity within several bacterial proteins, most of which are known or are believed to be glycosyl transferases (see text). Shown above the sequences is a consensus sequence derived from comparison of the 11 sequences listed below that for CgeD. An amino acid was included in the consensus sequence if it was found at that position in at least five sequences. Asterisks denote positions where one of five hydrophobic residues (I, L, V, M, or F) was found in at least eight sequences, and an asterisk was used even if an I, L, or V residue was found in five or more sequences. Plus signs denote positions where a basic residue (R, K, or H) was found in at least eight sequences. A dot denotes a position where an aromatic residue (F, Y, or W) was found in seven sequences. The numbers refer to the positions within each protein of the amino acids bordering the two regions shown. The database accession numbers corresponding to the sequences shown are as follows: *B. subtilis* Ipa-63d and Ipa-56d, X73124; *R. meliloti* ExoO, Z22636; *R. meliloti* ExoU, P33700; *R. meliloti* ExoW, P33702; *Anabaena* Orf, P22639; *E. coli* KfaB, X77617; *V. anguillarum* VirB L08012; *H. influenzae* Lsg Orf5, M94855; *E. coli* YibD, U00039; and *R. meliloti* NodC, P04341.

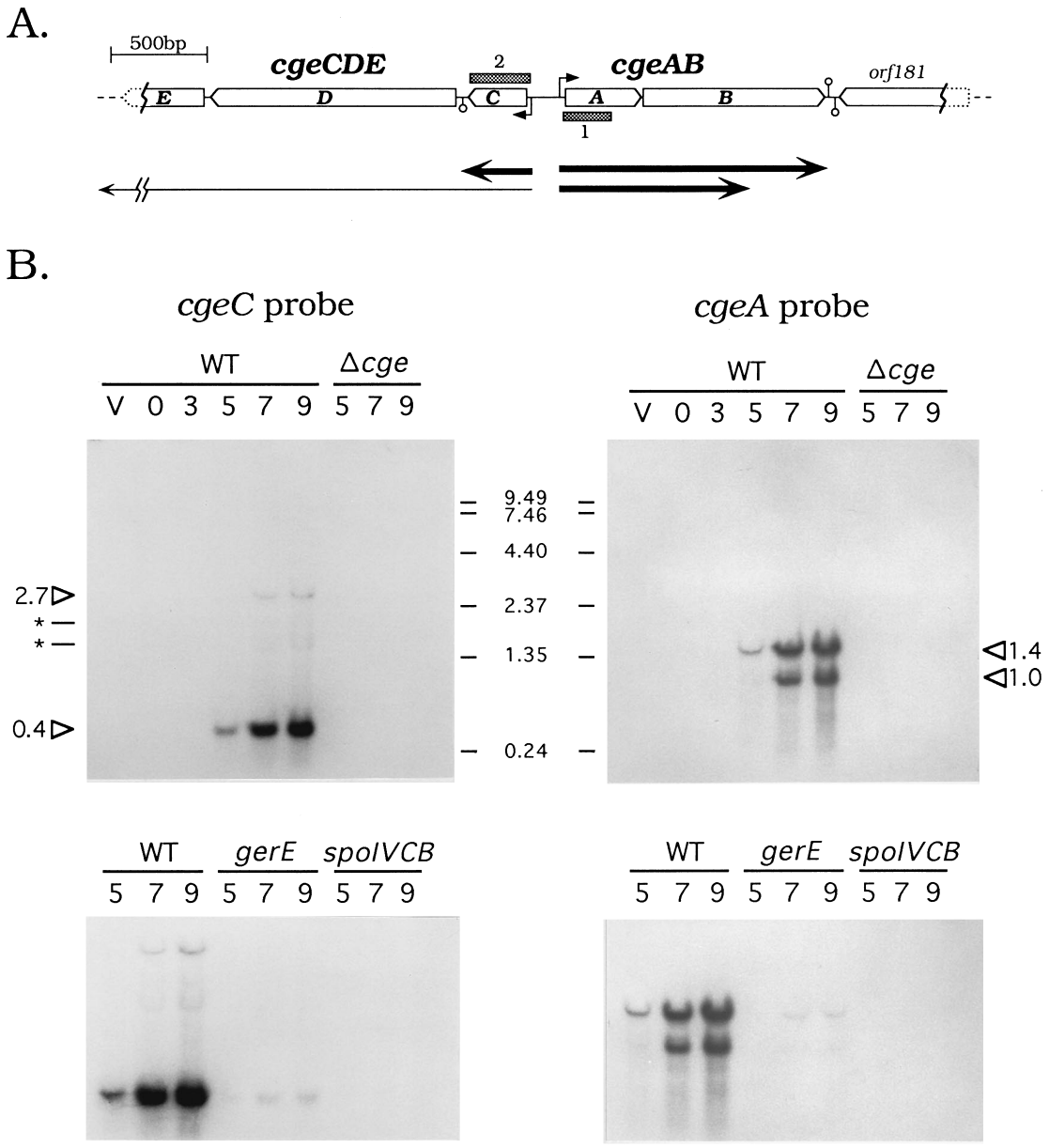


FIG. 5. Characterization of *cge* transcription units by Northern blot analysis. (A) Illustration of the *cge* region of the chromosome. Shaded bars denote sequences used as single-stranded probes (see Materials and Methods). Probes 1 and 2 are specific for *cgeA*-containing transcripts and *cgeC*-containing transcripts, respectively. Arrows indicate the directions and extents of the transcripts detected. (B) Detection of *cge* transcripts by Northern blot analysis. RNA was purified from wild-type (WT) or mutant cells harvested during vegetative growth (V) and at intervals during sporulation (0, 3, 5, 7, and 9 h). The RNA was then electrophoresed through agarose gels containing formaldehyde, transferred to nylon membranes, and hybridized to the single-stranded probes illustrated in panel A. Autoradiographic images of the resulting blots are shown. The positions and estimated sizes (in kilobases) of prominent transcripts are indicated. Asterisks denote the positions of a few minor transcripts not discussed in the text. The positions of RNA molecular size standards (Bethesda Research Laboratories) are shown between the top images. The strains used were PY79 (*spo*$^+$), SR294 [Δ(*cgeD-cgeB*)::*kan*], RL325 (*gerE36*), and KS215 (*spoIVCB*::Tn*917*Ω215).

cgeAB in a *cge*$^+$ strain had no effect, whereas a copy of *cgeAB* at *amyE* in a Δ*cgeAB* mutant strain resulted in spores that were indistinguishable from those produced by the wild-type strain. Finally, complementation of the large deletion [Δ(*cgeD-cgeB*)] resulted in a strain that produced spores exhibiting the properties noted above for the *cgeD* insertion mutant.

The timing of *cge* expression and the effects of *cge* mutations on the surface properties of spores suggested that *cge* might encode one or more structural components of the coat. However, none of the *cge* mutations visibly altered the spore ultrastructure as judged by electron microscopy, nor did they noticeably alter the pattern of coat polypeptides extracted from purified spores, as judged by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (46).

DISCUSSION

We have identified and characterized a cluster of GerE-controlled (*cge*) genes that are located at 181° on the genetic map of the *B. subtilis* 168 chromosome. The five *cge* genes are organized into two adjacent and divergent transcription units designated *cgeAB* and *cgeCDE*. The *cgeAB* operon consists of

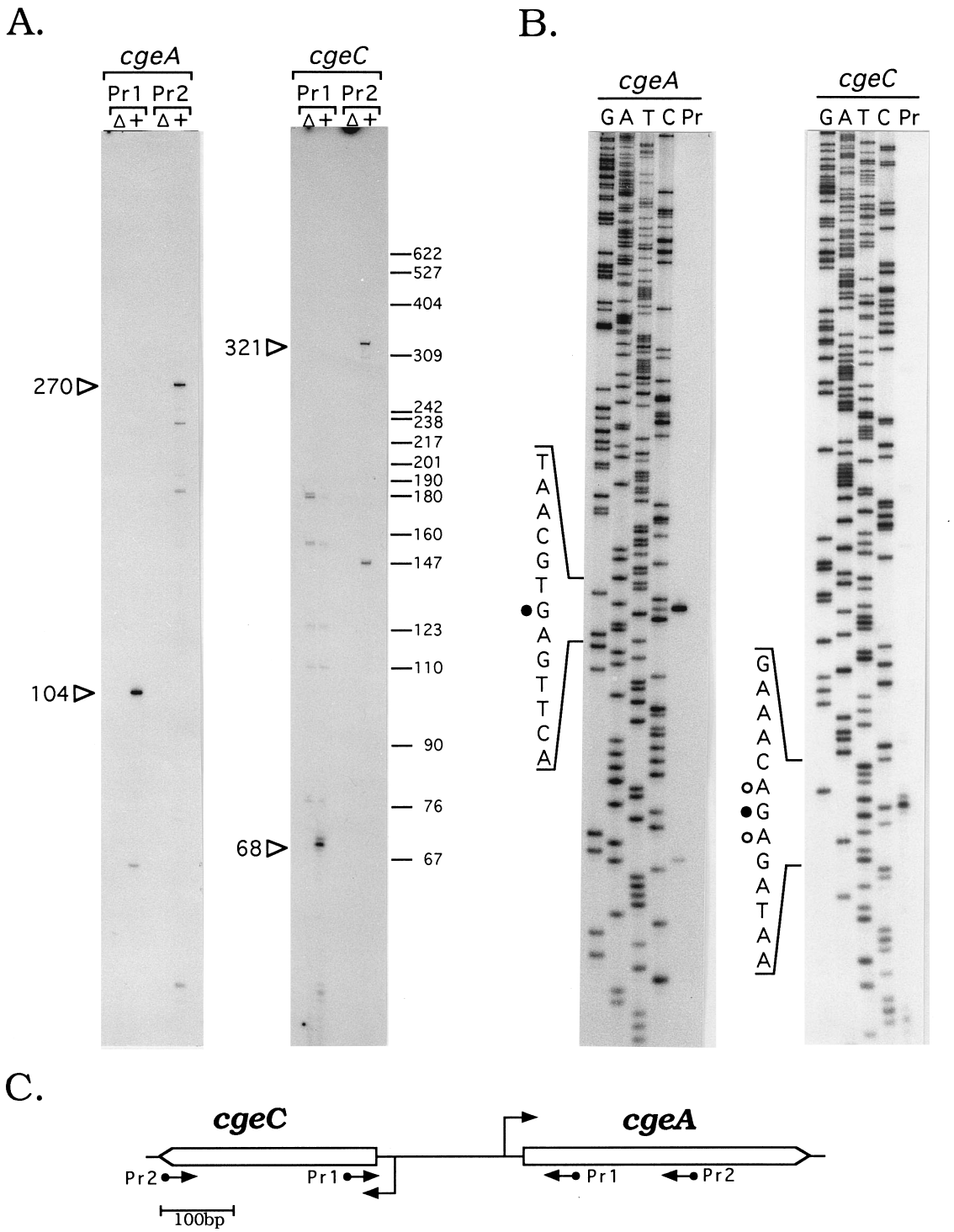


FIG. 6. Mapping of the 5′ termini of *cge* transcripts. (A) Low-resolution mapping of the 5′ termini of *cge* transcripts. RNA was purified from cells of PY79 (*spo*$^+$) (+) or SR294 [Δ(*cgeD-cgeB*)::*kan*] (Δ) cells harvested at 7 h of sporulation, and this RNA was used as a template in primer extension reactions with primer Pr1 or Pr2 to *cgeA* or *cgeC* (see panel C and Materials and Methods). The products of the extension reactions were resolved by electrophoresis through a polyacrylamide gel, and autoradiographic images of the gel are shown. The positions and sizes (in nucleotides) of principal extension products are indicated. Less abundant extension products are attributable to a lack of specificity (in the case of *cgeC* Pr1) or to premature termination of extension reactions (demonstrated by an absence or reduced abundance of these extension products when avian myeloblastosis virus reverse transcriptase was used instead of Moloney murine leukemia virus reverse transcriptase [46]). The positions and lengths of DNA fragments used as size markers are indicated on the right. (B) High-resolution mapping of the 5′ termini of *cge* transcripts. RNA from PY79 (see above) was used as a template for primer extension reactions with *cgeA* Pr1 or *cgeC* Pr1, and the products of the extension reactions were resolved on polyacrylamide slab gels alongside sequence ladders generated with the same 5′-end-labeled primers. The nucleotide sequences near the termini are shown, and nucleotides corresponding to the termini are indicated by circles. In the case of *cgeC*, the Pr1 extension reaction produces three extension products which differ in length by a single nucleotide. The filled circle denotes the nucleotide at the 5′ terminus of the principal extension product. (C) *cgeC* and *cgeA* Pr1 and Pr2 primers.

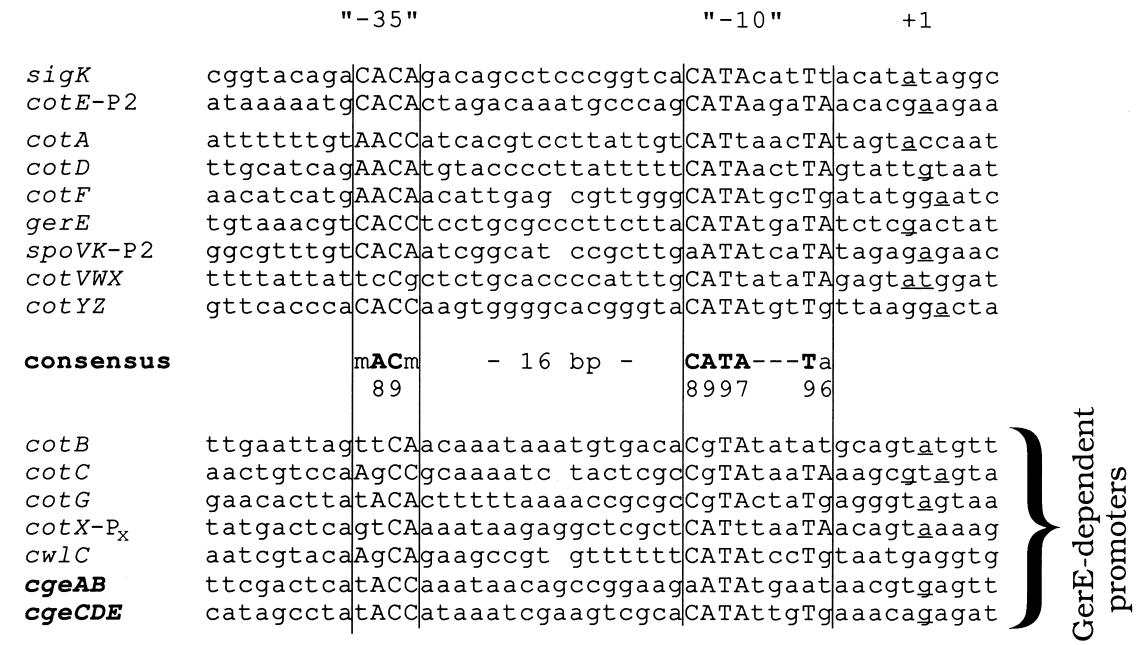


FIG. 7. Alignment of promoter regions of genes transcribed by σ^K RNA polymerase. Shown are the sequences near the transcription start sites of genes transcribed by RNA polymerase containing σ^K. Promoters for nine transcription units that are transcribed efficiently in the absence of GerE are shown above consensus sequences derived from inspection of these promoters. A nucleotide was included in the consensus sequence if it was found at that position in at least six of the nine promoters (and is shown in uppercase if found in at least seven). Also indicated are cases, in the −35 region, in which one of two nucleotides was found at a particular position in at least eight promoters (m means C or A). The promoter regions of seven transcription units that require GerE in addition to σ^K RNA polymerase for efficient expression are shown below the consensus sequence. Note that our assignment of the *cwlC* promoter to the GerE-dependent class is tentative and is based only on the relatively late appearance (7.5 h) of *cwlC* transcripts (32). Nucleotides in each promoter that match the consensus sequence are shown in uppercase. Nucleotides corresponding to the 5′ ends of transcripts are underlined. References for the sequences of these promoters are as follows: *sigK* (*spoIVCB*), 31; *cotA*, 49; *cotB*, *cotD*, and *cotE*, 66; *cotC*, 65 and 66; *cotF*, 9; *gerE*, 7; *spoVK*, 16; *cotVXW*, *cotX*, and *cotYZ*, 63; *cotG*, 47; and *cwlC*, 32.

two genes, *cgeA* and *cgeB*, whereas the *cgeCDE* operon comprises the remaining three genes, *cgeC*, *cgeD*, and *cgeE*. *cge* transcripts were shown to accumulate relatively late during sporulation and in a manner that was strongly dependent on GerE. We infer that the *cge* operons are transcribed by RNA polymerase containing σ^K, with GerE acting as an auxiliary, positively acting regulatory protein (see below), as has been shown to be the case for the GerE-controlled genes *cotB*, *cotC*, and *cotX* (64, 66).

The *cgeAB* and *cgeCDE* operon promoters bring to seven the number of promoters that are known (or inferred in the case of *cwlC*) to be under the control of GerE. The alignment shown in Fig. 7 compares the sequences of these seven promoters with those of nine σ^K-controlled promoters that are not strongly dependent on GerE. Updating previous analyses (16, 65) is the finding that the −35 and −10 regions of the GerE-independent promoters conform to the consensus sequences mACm (where m is C or A) and CATA---Ta, respectively. The −35 regions of these nine promoters match the consensus sequence perfectly (with one exception), and the −10 sequences conform with no more than one mismatch. In contrast, the GerE-dependent promoters (shown at the bottom of Fig. 7), including those for *cgeAB* and *cgeCDE*, conform imperfectly to the consensus sequence. The mismatches were most apparent in the −35 regions, where none of the sequences matched the consensus sequence at all four positions, but some of the GerE-dependent promoters also lack highly conserved nucleotides in their −10 regions. The deviation of GerE-dependent promoters from the σ^K consensus sequence is consistent with the view that GerE influences transcription by compensating for otherwise weak interactions between σ^K and the promoter (64).

Our interest in the *cge* locus was based in part on the fact that none of the known GerE-dependent genes accounted for the dramatic defects exhibited by *gerE* mutant spores. For example, mutations in *cotB* or *cotC* resulted only in the absence of the corresponding polypeptide from the spore coat and had no discernible effect on the appearance of the coat or on resistance properties or germination of the spores (12). In extension of these previous findings, none of the *cge* mutations, alone or in combination with mutations in the GerE-controlled coat protein genes *cotB*, *cotC*, and/or *cotG*, resulted in the production of spores exhibiting any of the dramatic defects associated with mutations in *gerE* (46). Moreover, *cotX* has a GerE-independent mode of expression (Fig. 7) (64), and neither *cotX* nor *cwlC* mutations affect spore resistance properties or impair germination (32, 63). Interestingly, the predicted product of *cgeD* is very similar to that of another *B. subtilis* gene, *ipa-63d*, which is also part of a large, developmentally regulated operon that is under the control of σ^K (18, 24). This raises the possibility that the operon encompassing *ipa-63d* is another member of the GerE regulon (although this has not yet been determined). However, a deletion of the *ipa-63d* operon, alone or in combination with mutations in *cge*, does not result in spores that exhibit the *gerE* mutant phenotype of lysozyme sensitivity and defective germination (46). It is likely, then, that a full understanding of the GerE regulon and the effects of *gerE* mutations on spore properties will require the discovery and characterization of additional genes whose transcription is dependent on this late-acting transcription factor.

Despite not appreciably mimicking the effects of a *gerE* mutation, *cge* mutations nevertheless resulted in the production of spores with altered surface properties. These observations, combined with the timing and mother cell localization of

cge expression, suggested that *cge* encodes one or more coat components or a protein that is required for the proper assembly of part of the coat. However, mutations in *cge* genes did not affect the pattern of coat proteins that could be extracted from purified spores, nor did they affect the appearance of the coat as viewed by electron microscopy (46). It remains possible that *cge* does encode one or more coat proteins but that these proteins are not sufficiently abundant to allow detection, are not resolved adequately under the conditions used, or are components of the insoluble fraction (about 30%) of the spore coat. Interestingly, a deletion of the gene for CotX (which is found in the insoluble coat fraction) or a deletion that prevents the production of both CotY and CotZ results in the production of spores which readily clump upon purification (63), a phenotype that is reminiscent of that observed for strains harboring a deletion of *cgeAB*.

Alternatively, *cge*-encoded proteins may be involved in modification of the spore surface without being structural components or affecting coat protein assembly per se. Possible clues are suggested by the similarity between the amino terminus of the predicted CgeD protein and conserved regions of several nucleotide sugar transferases that are involved in polysaccharide biosynthesis, including glycosyl transferases from *R. meliloti*. This region of similarity to nucleotide sugar transferases is also present in the proposed product of *ipa-63d*, which is highly similar along its entire length to CgeD (Fig. 4). In support of the view that this similarity to the transferase enzymes is indicative of the function of these two proteins, other genes in the *ipa-63d* operon encode proteins that are homologous to proteins involved in lipopolysaccharide biosynthesis in gram-negative bacteria (18, 24). This, together with the observed effect of an insertion in *cgeD* on spore surface properties, led us to suspect that CgeD might be involved in glycosylation of one or more structural components of the coat. In keeping with this idea, purified coat fractions contain carbohydrate material (21), and at least one extractable coat polypeptide appears to be a glycoprotein (27, 42).

Another possibility is that *cge*-encoded proteins, and CgeD in particular, are involved in biosynthesis of an outermost spore layer referred to as the exosporium. The exosporium, found in several *Bacillus* and *Clostridium* species, has been characterized most extensively for *Bacillus cereus* (59), in which it exists as a loose-fitting, membranous layer consisting of protein, lipid, and polysaccharide which can self-assemble to an extent from solubilized components in vitro (4, 36, 41). The existence of a corresponding outer layer in *B. subtilis* is uncertain, however. Nevertheless, a tight-fitting exosporium-like envelope that is visible by electron microscopy after partial extraction of coat proteins has been described (55). The effects of *cge* mutations, then, may reflect a role in biosynthesis or modification of this exosporium layer, presumably affecting its polysaccharide component.

With regard to either of these latter models, it is interesting that disruption of the *B. subtilis ipa-63d* operon results in the production of spores that are phenotypically similar to those produced by Δ*cgeAB* mutants (24, 46). Thus, the products of the *cge* and *ipa-63d* operons may constitute a new class of sporulation proteins that are involved in glycosylation of the spore surface.

In summary, we have described the cloning and sequence analysis of two adjacent and divergently oriented GerE-controlled operons, *cgeAB* and *cgeCDE*. Mutations in *cge*, as is the case for mutations in other known *gerE*-controlled genes, did not substantially affect lysozyme resistance or germination of spores, properties that are strongly affected by mutations in *gerE*. This indicates that not all the members of the GerE regulon have as yet been identified. Nevertheless, *cge* mutations do affect properties of the spore surface, and we have postulated, on the basis of similarity between CgeD and nucleotide sugar transferases, that *cge* may be involved in the attachment of sugar residues to the spore surface.

ACKNOWLEDGMENTS

We thank Marie-Françoise Hullo and Frank Kunst for sharing their results prior to publication and members of the Losick laboratory for many helpful discussions.

This work was supported by Public Health Service grant GM18568 from the National Institutes of Health to R.L.

REFERENCES

1. **Anagnostopoulos, C., P. J. Piggot, and J. A. Hoch.** 1993. The genetic map of *Bacillus subtilis*, p. 425–462. *In* A. L. Sonenshein, J. A. Hoch, and R. L. Losick (ed.), *Bacillus subtilis* and other gram-positive bacteria: biochemistry, physiology, and molecular genetics. American Society for Microbiology, Washington, D.C.
2. **Atkinson, E. M., and S. R. Long.** 1992. Homology of *Rhizobium meliloti* NodC to polysaccharide polymerizing enzymes. Mol. Plant Microbe Interact. **5:**439–442.
3. **Beall, B., A. Driks, R. Losick, and C. P. Moran Jr.** 1993. Cloning and characterization of a gene required for assembly of the *Bacillus subtilis* spore coat. J. Bacteriol. **175:**1705–1716.
4. **Beaman, T. C., H. S. Pankratz, and P. Gerhardt.** 1971. Paracrystalline sheets reaggregated from solubilized exosporium of *Bacillus cereus*. J. Bacteriol. **107:**320–324.
5. **Connors, M. J., S. Howard, J. Hoch, and P. Setlow.** 1986. Determination of the chromosomal location of four *Bacillus subtilis* genes which code for a family of small, acid-soluble spore proteins. J. Bacteriol. **166:**412–416.
6. **Cutting, S., and J. Mandelstam.** 1986. The nucleotide sequence and the transcription during sporulation of the *gerE* gene of *Bacillus subtilis*. J. Gen. Microbiol. **132:**3013–3024.
7. **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.
8. **Cutting, S., S. Roels, and R. Losick.** 1991. Sporulation operon *spoIVF* and the characterization of mutations that uncouple mother-cell from forespore gene expression in *Bacillus subtilis*. J. Mol. Biol. **221:**1237–1256.
9. **Cutting, S., L. Zheng, and R. Losick.** 1991. Gene encoding two alkali-soluble components of the spore coat from *Bacillus subtilis*. J. Bacteriol. **173:**2915–2919.
10. **Cutting, S. M., and P. B. Vander Horn.** 1990. Genetic analysis, p. 27–74. *In* C. R. Harwood and S. M. Cutting (ed.), Molecular biological methods for *Bacillus*. John Wiley & Sons Ltd., Chichester, United Kingdom.
11. **Devereux, J., P. Haeberli, and O. Smithies.** 1984. A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. **12:**387–395.
12. **Donovan, W., L. Zheng, K. Sandman, and R. Losick.** 1987. Genes encoding spore coat polypeptides from *Bacillus subtilis*. J. Mol. Biol. **196:**1–10.
13. **Driks, A., and R. Losick.** 1991. Compartmentalized expression of a gene under the control of sporulation transcription factor σ^E in *Bacillus subtilis*. Proc. Natl. Acad. Sci. USA **88:**9934–9938.
14. **Errington, J.** 1993. *Bacillus subtilis* sporulation: regulation of gene expression and control of morphogenesis. Microbiol. Rev. **57:**1–33.
15. **Feng, P., and A. I. Aronson.** 1986. Characterization of a *Bacillus subtilis* germination mutant with pleiotropic alterations in spore coat structure. Curr. Microbiol. **13:**221–226.
16. **Foulger, D., and J. Errington.** 1991. Sequential activation of dual promoters by different sigma factors maintains *spoVJ* expression during successive developmental stages of *Bacillus subtilis*. Mol. Microbiol. **5:**1363–1373.
17. **Genetics Computer Group.** 1991. Program manual for the GCG package, version 7. Genetics Computer Group, Madison, Wis.
18. **Glaser, P., F. Kunst, M. Arnaud, M.-P. Coudart, W. Gonzales, M.-F. Hullo, M. Ionescu, B. Lubochinsky, L. Marcelino, I. Moszer, E. Presecan, M. Santana, E. Schneider, J. Schweizer, A. Vertes, G. Rapoport, and A. Danchin.** 1993. *Bacillus subtilis* genome project: cloning and sequencing of the 97 kb region from 325° to 333°. Mol. Microbiol. **10:**371–384.
19. **Glucksmann, M. A., T. L. Reuber, and G. C. Walker.** 1993. Family of glycosyl transferases needed for the synthesis of succinoglycan by *Rhizobium meliloti*. J. Bacteriol. **175:**7033–7044.
20. **Glucksmann, M. A., T. L. Reuber, and G. C. Walker.** 1993. Genes needed for the modification, polymerization, export, and processing of succinoglycan by *Rhizobium meliloti*: a model for succinoglycan biosynthesis. J. Bacteriol. **175:**7045–7055.
21. **Goldman, R. C., and D. J. Tipper.** 1978. *Bacillus subtilis* spore coats: complexity and purification of a unique polypeptide component. J. Bacteriol. **135:**1091–1106.
22. **Halberg, R., and L. Kroos.** 1994. Sporulation regulatory protein SpoIIID

from *Bacillus subtilis* activates and represses transcription by both mother-cell-specific forms of RNA polymerase. J. Mol. Biol. **243:**425–436.
23. **Holland, D., and C. P. Wolk.** 1990. Identification and characterization of *hetA*, a gene that acts early in the process of morphological differentiation of heterocysts. J. Bacteriol. **172:**3131–3137.
24. **Hullo, M.-F., and F. Kunst.** 1994. Personal communication.
25. **Jacobs, T. W., T. T. Egelhoff, and S. R. Long.** 1985. Physical and genetic map of a *Rhizobium meliloti* nodulation gene region and nucleotide sequence of *nodC*. J. Bacteriol. **162:**469–476.
26. **Jenkinson, H. F., and H. Lord.** 1983. Protease deficiency and its association with defects in spore coat structure, germination and resistance properties in a mutant of *Bacillus subtilis*. J. Gen. Microbiol. **129:**2727–2737.
27. **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.
28. **Kroos, L., B. Kunkel, and R. Losick.** 1989. Switch protein alters specificity of RNA polymerase containing a compartment-specific sigma factor. Science **243:**526–529.
29. **Kunkel, B., L. Kroos, H. Poth, P. Youngman, and R. Losick.** 1989. Temporal and spatial control of the mother-cell regulatory gene *spoIIID* of *Bacillus subtilis*. Genes Dev. **3:**1735–1744.
30. **Kunkel, B., R. Losick, and P. Stragier.** 1990. The *Bacillus subtilis* gene for the developmental transcription factor σ^K is generated by excision of a dispensable DNA element containing a sporulation recombinase gene. Genes Dev. **4:**525–535.
31. **Kunkel, B., K. Sandman, S. Panzer, P. Youngman, and R. Losick.** 1988. The promoter for a sporulation gene in the *spoIVC* locus of *Bacillus subtilis* and its use in studies of temporal and spatial control of gene expression. J. Bacteriol. **170:**3513–3522.
32. **Kuroda, A., Y. Asami, and J. Sekiguchi.** 1993. Molecular cloning of a sporulation-specific cell wall hydrolase gene of *Bacillus subtilis*. J. Bacteriol. **175:** 6260–6268.
33. **LeBlanc, D. J., J. M. Inamine, and L. N. Lee.** 1986. Broad geographical distribution of homologous erythromycin, kanamycin, and streptomycin resistance determinants among group D streptococci of human and animal origin. Antimicrob. Agents Chemother. **29:**549–555.
34. **Losick, R., and P. Stragier.** 1992. Crisscross regulation of cell-type-specific gene expression during development in *Bacillus subtilis*. Nature (London) **355:**601–604.
35. **Lovett, P. S., N. P. Ambulos, Jr., W. Mulbry, N. Noguchi, and E. J. Rogers.** 1991. UGA can be decoded as tryptophan at low efficiency in *Bacillus subtilis*. J. Bacteriol. **173:**1810–1812.
36. **Matz, L. L., T. C. Beaman, and P. Gerhardt.** 1970. Chemical composition of exosporium from spores of *Bacillus cereus*. J. Bacteriol. **101:**196–201.
37. **McLaughlin, R., S. M. Spinola, and M. A. Apicella.** 1992. Generation of lipooligosaccharide mutants of *Haemophilus influenzae* type b. J. Bacteriol. **174:**6455–6459.
38. **Moir, A.** 1981. Germination properties of a spore coat-defective mutant of *Bacillus subtilis*. J. Bacteriol. **146:**1106–1116.
39. **Nicholson, W. L., and P. Setlow.** 1990. Sporulation, germination and outgrowth, p. 391–450. *In* C. R. Harwood and S. Cutting (ed.), Molecular biological methods for *Bacillus*. John Wiley & Sons Ltd., Chichester, United Kingdom.
40. **Norqvist, A., and H. Wolf-Watz.** 1993. Characterization of a novel chromosomal virulence locus involved in expression of a major surface flagellar sheath antigen of the fish pathogen *Vibrio anguillarum*. Infect. Immun. **61:** 2434–2444.
41. **Ohye, D. F., and W. G. Murrell.** 1973. Exosporium and spore coat formation in *Bacillus cereus* T. J. Bacteriol. **115:**1179–1190.
42. **Pandey, N. K., and A. I. Aronson.** 1979. Properties of the *Bacillus subtilis* spore coat. J. Bacteriol. **137:**1208–1218.
43. **Ricca, E., S. Cutting, and R. Losick.** 1992. Characterization of *bofA*, a gene involved in intercompartmental regulation of pro-σ^K processing during sporulation in *Bacillus subtilis*. J. Bacteriol. **174:**3177–3184.
44. **Roberts, I. S.** Unpublished data; GenBank accession number X77617.
45. **Roels, S., A. Driks, and R. Losick.** 1992. Characterization of *spoIVA*, a sporulation gene involved in coat morphogenesis in *Bacillus subtilis*. J. Bacteriol. **174:**575–585.
46. **Roels, S., and R. Losick.** Unpublished data.
47. **Sacco, M., E. Ricca, R. Losick, and S. Cutting.** 1995. An additional GerE-controlled gene encoding an abundant spore coat protein from *Bacillus subtilis*. J. Bacteriol. **177:**372–377.
48. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
49. **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.
50. **Sandman, K., R. Losick, and P. Youngman.** 1987. Genetic analysis of *Bacillus subtilis spo* mutations generated by Tn*917*-mediated insertional mutagenesis. Genetics **117:**603–617.
51. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74:**5463–5467.
52. **Sato, T., K. Harada, Y. Ohta, and Y. Kobayashi.** 1994. Expression of the *Bacillus subtilis spoIVCA* gene, which encodes a site-specific recombinase, depends on the *spoIIGB* product. J. Bacteriol. **176:**935–937.
53. **Serror, P., V. Azevedo, and S. D. Ehrlich.** 1993. An ordered collection of *Bacillus subtilis* DNA segments in yeast artificial chromosomes, p. 473–474. *In* A. L. Sonenshein, J. A. Hoch, and R. L. Losick (ed.), *Bacillus subtilis* and other gram-positive bacteria: biochemistry, physiology, and molecular genetics. American Society for Microbiology, Washington, D.C.
54. **Sofia, H. J., V. Burland, D. L. Daniels, G. Plunkett, and F. R. Blattner.** 1994. Analysis of the *Escherichia coli* genome. V. DNA sequence of the region from 76.0 to 81.5 minutes. Nucleic Acids Res. **22:**2576–2586.
55. **Sousa, J. C., M. T. Silva, and G. Balassa.** 1976. An exosporium-like outer layer in *Bacillus subtilis* spores. Nature (London) **263:**53–54.
56. **Stevens, C. M., and J. Errington.** 1990. Differential gene expression during sporulation in *Bacillus subtilis*: structure and regulation of the *spoIIID* gene. Mol. Microbiol. **4:**543–552.
57. **Stragier, P., B. Kunkel, L. Kroos, and R. Losick.** 1989. Chromosomal rearrangement generating a composite gene for a developmental transcription factor. Science **243:**507–512.
58. **Stragier, P., and R. Losick.** 1990. Cascades of sigma factors revisited. Mol. Microbiol. **4:**1801–1806.
59. **Tipper, D. J., and J. J. Gauthier.** 1972. Structure of the bacterial endospore, p. 3–12. *In* H. O. Halvorson, R. Hanson, and L. L. Campbell (ed.), Spores V. American Society for Microbiology, Washington, D.C.
60. **Trieu-Cout, P., and P. Courvalin.** 1983. Nucleotide sequence of the *Streptococcus faecalis* plasmid gene encoding the 3′5″-aminoglycoside phosphotransferase type III. Gene **23:**331–341.
61. **Virca, G. D., W. Northemann, B. R. Shiels, G. Widera, and S. Broome.** 1990. Simplified Northern blot hybridization using 5% sodium dodecyl sulfate. BioTechniques. **8:**370–371.
62. **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.
63. **Zhang, J., P. C. Fitz-James, and A. I. Aronson.** 1993. Cloning and characterization of a cluster of genes encoding polypeptides present in the insoluble fraction of the spore coat of *Bacillus subtilis*. J. Bacteriol. **175:**3757–3766.
64. **Zhang, J., H. Ichikawa, R. Halberg, L. Kroos, and A. I. Aronson.** 1994. Regulation of the transcription of a cluster of *Bacillus subtilis* spore coat genes. J. Mol. Biol. **240:**405–415.
65. **Zheng, L., R. Halberg, S. Roels, H. Ichikawa, L. Kroos, and R. Losick.** 1992. Sporulation regulatory protein GerE from *Bacillus subtilis* binds to and can activate or repress transcription from promoters for mother-cell-specific genes. J. Mol. Biol. **226:**1037–1050.
66. **Zheng, L., and R. Losick.** 1990. Cascade regulation of spore coat gene expression in *Bacillus subtilis*. J. Mol. Biol. **212:**645–660.