JOURNAL OF BACTERIOLOGY, Oct. 1989, p. 5362–5375
0021-9193/89/105362-14$02.00/0

Vol. 171, No. 10

Sequence and Transcription Mapping of *Bacillus subtilis* Competence Genes *comB* and *comA*, One of Which Is Related to a Family of Bacterial Regulatory Determinants

YVETTE WEINRAUCH, NANCY GUILLEN,† AND DAVID A. DUBNAU*

Department of Microbiology, The Public Health Research Institute, New York, New York 10016

Received 9 March 1989/Accepted 26 June 1989

The complete nucleotide sequences of the *comA* and *comB* loci of *Bacillus subtilis* were determined. The products of these genes are required for the development of competence in *B. subtilis* and for the expression of late-expressing competence genes. The major 5′ termini of both the *comA* and *comB* transcripts were determined. The inferred promoters of both *comA* and *comB* contained sequences that were similar to those found at the −10 and −35 regions of promoters that are used by σ^A-RNA polymerase, the primary form of this enzyme in vegetative cells. The *comB* gene was located approximately 3 kilobase pairs upstream of the *comA* gene and encoded a 409-amino-acid protein with a predicted molecular weight of 46,693. The *comA* locus contained two open reading frames (ORFs) and *comB* contained one ORF. The predicted amino acid sequence of the *comA* ORF1 gene product consisted of 214 amino acids, with an aggregate molecular weight of 24,132. The ORF1 product was required for competence, while ORF2, which was cotranscribed with ORF1 and encoded a predicted protein of 126 amino acids, was not. The predicted protein sequence of the *comA* ORF1 gene product was found to be similar to that of several members of the effector class of procaryotic signal transducers. The C-terminal portion of the predicted *comA* sequence contained a possible helix-turn-helix motif, which is characteristic of DNA-binding proteins. *comA* ORF1 was cloned on a multicopy plasmid and was shown to complement the competence-deficient phenotype caused by the *comA124* insertion of Tn*917lac*. Also, the presence of *comA* ORF1 in multiple copies interfered with sporulation. Anti-peptide antibodies raised to the predicted product of *comA* ORF1 reacted strongly with a single protein band of about 24,000 daltons in immunoblots. The possible roles of multiple signal transduction systems in triggering the development of competence are discussed.

The use of Tn*917lac* (61, 63, 64) has allowed the identification and isolation of several genes in *Bacillus subtilis* that are required for the development of genetic competence. Of these, *comA* and *comB* have been shown to be expressed throughout growth and to be expressed to the same extent in all media tested (1, 19). Another class of competence genes is expressed at a sharply increased rate during the transition to the stationary phase and only in growth medium that leads to the development of competence. These two groups of genes are referred to as early and late, respectively.

comA and *comB* have been isolated by molecular cloning, and physical as well as low-resolution transcription maps have been obtained (18). The two genes are located about 3 kilobase pairs (kbp) from one another on the *B. subtilis* chromosome and are transcribed in the same direction, but from independent promoters. The products of *comA* and *comB* are required for the expression of the late *com* genes, suggesting that they play regulatory roles during the development of competence.

In this report we present the complete DNA nucleotide sequences of *comA* and *comB* and identify their probable promoter sequences. The similarity of the *comA* product to other bacterial regulatory proteins involved in signal transduction is noted, and the possible roles of the two genes in the development of competence are discussed.

* Corresponding author.
† Present address: Institut de Microbiologie, Université Paris-Sud, 91405 Orsay, Cedex, France.

MATERIALS AND METHODS

Materials. Restriction endonucleases were obtained from several sources (Boehringer Mannheim Biochemicals, Indianapolis, Ind.; New England BioLabs, Inc., Beverly, Mass.; and Bethesda Research Laboratories, Gaithersburg, Md.) and were used as specified by the manufacturers. The Cyclone Biosystem (International Biotechnologies Inc., New Haven, Conn.) was used to make overlapping deletions.

Strains and plasmids. All *B. subtilis* strains were isogenic with BD630 (*leu metB5 hisH2*). *Escherichia coli* JM109 was used as a host for recombinant M13 molecules. *E. coli* HB101 was used as a host for all in vitro-generated recombinant plasmids by use of the pBR322 vector. The plasmids used in this study are listed in Table 1. pBD438 is a plasmid that replicates from the temperature-sensitive *ori* of pE194 *cop-6* and that confers erythromycin resistance (Em^r) and kanamycin resistance (Km^r). Plasmid pBD439 was constructed by cloning the 2-kb *Bgl*II-*Sst*I fragment from pED4 (18) into the *Bcl*I-*Sst*I sites in the *ermC* determinant of pBD385 (the *Bgl*II site was derived from the Tn*917lac* transposon). A chloramphenicol resistance (Cm^r) cassette was isolated on a 1.3-kb *Sau*3A fragment from pED1 and inserted into one of the two *Bcl*I sites of pBD439 in order to inactivate the *comA* open reading frame 2 (ORF2) (see below), creating pBD443. The location of the Cm^r cassette in the proper *Bcl*I site was confirmed by restriction site mapping. The 1.4-kb *Hin*dIII fragment containing *comA* ORF1 was inserted into the *Hin*dIII site of pBD347 to create pBD440.

TABLE 1. Plasmids used in this study

Plasmid	Characters	Comments	Source or reference
pBD347	Cm^r	pIM13 replicon	30
pBD438	Em^r Km^r	pE194*cop-6* replicon	
pBD439	Km^r	*Bgl*II-*Sst*I fragment from pED4	This work
pBD440	Cm^r *comA* ORF1	pIM13 replicon	This work
pBD443	Cm^r Km^r	Chloramphenicol determinant inserted in *comA* ORF2	This work
pED3	Cm^r Amp^r	*comB*, pBR322 replicon	18
pED4	Cm^r Amp^r	*comA* ORF1 and ORF2, pBR322 replicon	18

Measurement of sporulation. Strains to be tested for sporulation were grown for 24 h in nutrient sporulation medium (46) at 37°C and plated for viable counts before and after incubation of the viable count dilution tubes at 80°C for 20 min.

Isolation of plasmid DNA. *B. subtilis* strains were screened for plasmid by the rapid plasmid isolation procedure described by Contente and Dubnau (8). Large-scale preparation of *B. subtilis* plasmid DNA was done by sodium dodecyl sulfate-NaCl treatment of plasmid-bearing strains as described by Guerry et al. (17), followed by CsCl-ethidium bromide purification. *E. coli* strains were screened for plasmid by the alkaline lysis protocol (3). Large-scale preparation of *E. coli* plasmid DNA was done by the alkaline lysis method followed by purification in CsCl-ethidium bromide gradients, as described by Maniatis et al. (29).

Transduction and transformation. Preparation of phage PBS1 transducing lysates and transduction of *B. subtilis* strains were carried out as described previously (12). *B. subtilis* strains were made competent and transformed as described previously (19) by using the one-step competence regimen. Competent *E. coli* cells were prepared as described by Maniatis et al. (29).

S1 nuclease protection mapping. RNA was prepared from the wild-type *com*$^+$ strain (BD630), which was grown in competence medium. Cells were harvested 1 h before time zero (the transition from the exponential to the stationary growth phase). RNA was extracted as described by Ulmanen et al. (57). S1 nuclease mapping was carried out as described previously (2, 14, 18).

Primer extension mapping. Two primers were used; the first (5′-CCGTTTCCAAAATTGTCTTG-3′) was complementary to a region of *comA* mRNA 74 bases from the presumed start site of transcription (inferred from S1 nuclease analysis). The second primer (5′-GCGCTATACACGTT CTCTTC-3′) was complementary to a region of the *comB* mRNA 113 bases from the presumed start site of transcription. Both primers were 5′-end-labeled with polynucleotide kinase as described previously (29). A total of 5 ng of labeled primer was mixed with 50 μg of total RNA (prepared as described above) from wild-type strain BD630. The RNA was dissolved in 16 μl of buffer containing 50 mM Tris (pH 8.0), 30 mM KCl, 8 mM $MgCl_2$, 1 mM dithiothreitol, and 1 μl of RNasin (Promega Biotech). The mixtures were heated at 80°C for 10 min and slowly cooled to 42°C. A total of 4 μl of a mixture of four deoxynucleotide triphosphates (each at 10 mM) and 1 μl of avian myeloblastosis virus reverse transcriptase (Life Sciences, Inc., St. Petersburg, Fla.) were added to the annealed samples, which were then incubated at 42°C for 30 min. Analysis of the extended products was carried out by electrophoresis by using 6% polyacrylamide–urea gels.

DNA sequencing. Sequencing of DNA was by the dideoxynucleotide chain-termination method (45) with [α-^{32}S]dATP (specific activity, 500 Ci/mmol; Dupont, NEN Research Products, Boston, Mass.). Sequencing reactions were performed with modified T7 DNA polymerase kits (Sequenase; U.S. Biochemical Corp.). The DNA was entirely sequenced on both strands by the strategy of Dale et al. (9) to generate a nested series of overlapping clones with deletions. The start points of the deletions obtained were separated, on average, by about 200 bases, permitting the unambiguous ordering of DNA sequences because of substantial overlapping. Sequencing reactions were performed with M13 templates with the M13 universal primer. In some cases in which overlapping clones were not available, 17-mer primers were made (Operon Technologies) based on known sequences.

The sequences of *comA* and *comB* were submitted to the Genetic Sequence Data Bank (GenBank) and were assigned the respective accession numbers M22856 and M22855.

Computer analyses of DNA sequences. The deduced amino acid sequences were analyzed by the TFASTA, FASTP, and RDF programs (26, 40) to determine their similarities to other proteins in a translated version of the GenBank.

Immunological procedures. A custom synthetic peptide (CQKTKPAPSSQKEQD) coupled to keyhole limpet hemocyanin was purchased from Biosearch Corp. and was used to raise antibodies in New Zealand White rabbits. The peptide sequence was selected as a particularly hydrophilic segment, based on analysis by the method of Engelman et al. (13). For Western blot (immunoblot) analysis, *B. subtilis* cells were grown in competence medium, harvested late in the exponential phase, and lysed by incubation at 37°C in 10 mM Tris hydrochloride (pH 7.5)–1 μg of DNase per ml–2 mM phenylmethylsulfonyl fluoride–1 mg of lysozyme per ml. The cells were then boiled in 10% sodium dodecyl sulfate. Solubilized proteins were resolved by electrophoresis through sodium dodecyl sulfate–15% polyacrylamide gels and transferred to nitrocellulose as described by Towbin et al. (54). Immunodetection was done as described by Blake et al. (4) by using alkaline phosphatase-conjugated goat anti-rabbit immunoglobulin G (Calbiochem-Behring, La Jolla, Calif.).

RESULTS

Nucleotide sequences of *comA* and *comB*. We have isolated plasmid clones carrying *comA*(pED4) and *comB*(pED3) and have established physical maps of these determinants (18). Based on these data, restriction fragments containing *comA* and *comB* sequences were cloned into M13mp19*cat*, M13mp19, or M13mp18; and a sequential series of overlapping deletant bacteriophages was isolated by the procedure of Dale et al. (9). By using these derivatives, the nucleotide sequences of *comA* (Fig. 1) and *comB* (Fig. 2) were determined completely on both strands and across all the restriction sites used for cloning. To identify the precise points of insertion of the Tn*917lac* transposons that were used to isolate the original *comA124* and *comB138* mutations, restriction fragments were isolated from plasmids carrying the chromosomal DNA-transposon junctions (18) and cloned into M13mp18 or M13mp19, and the nucleotide sequences were determined. These junction points are indicated in Fig. 1 and 2. The Tn*917lac* insertion into *comB* occurred very near the transcriptional start site (see below) and adjacent to a sequence that was identical to that at the end of the

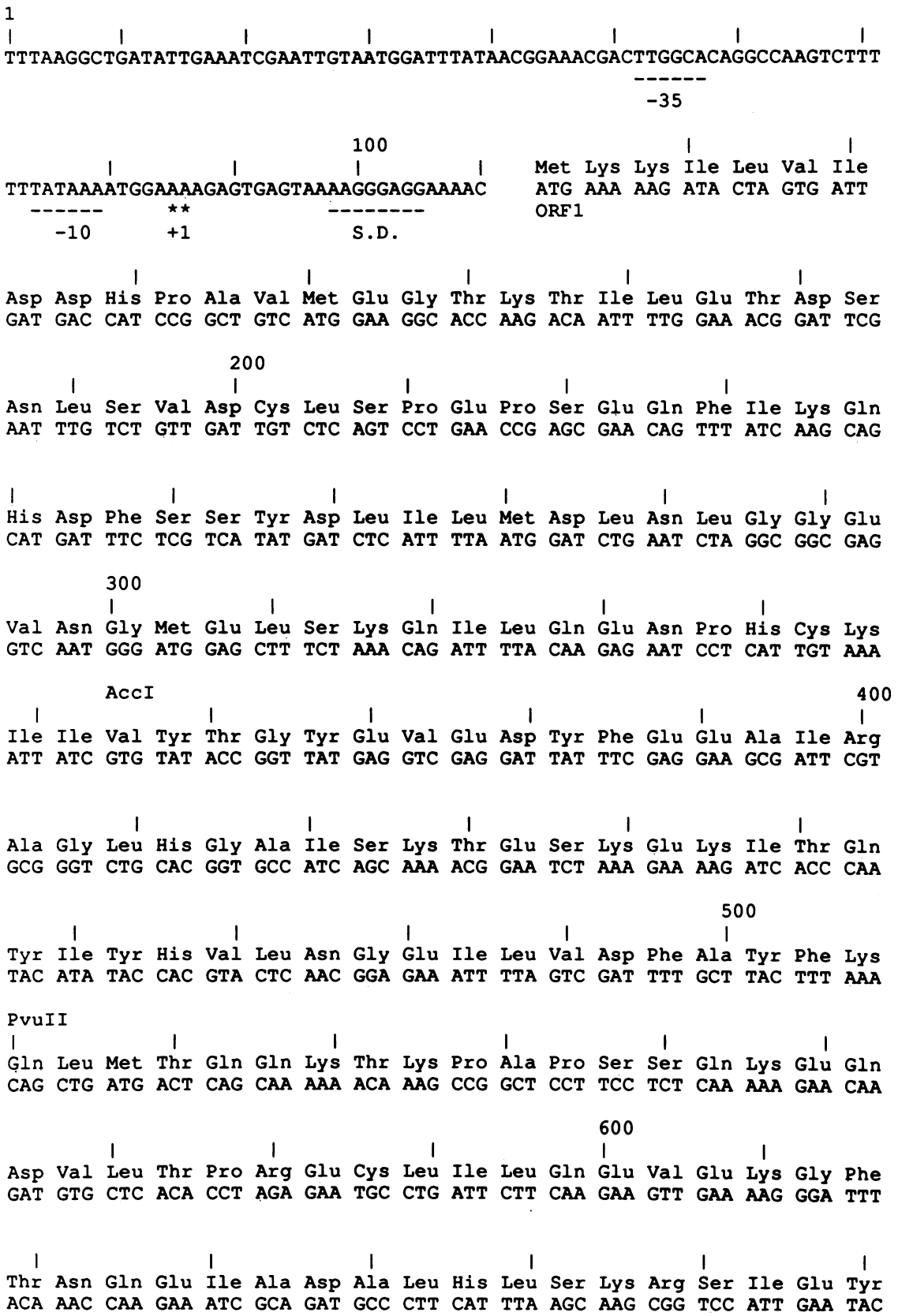

FIG. 1. Nucleotide sequence of *comA*. The complete nucleotide sequence of *comA* is shown together with the predicted amino acid sequences of two protein products (ORF1 and ORF2). Putative −10 and −35 promoter sequences and Shine-Dalgarno (S.D.) sequences (50) are indicated by underlining. The transcriptional start site inferred from S1 nuclease mapping (asterisks) and the site of the Tn*917lac* insertion (arrowhead at position 694) are also marked. Converging arrows indicate dyad symmetry.

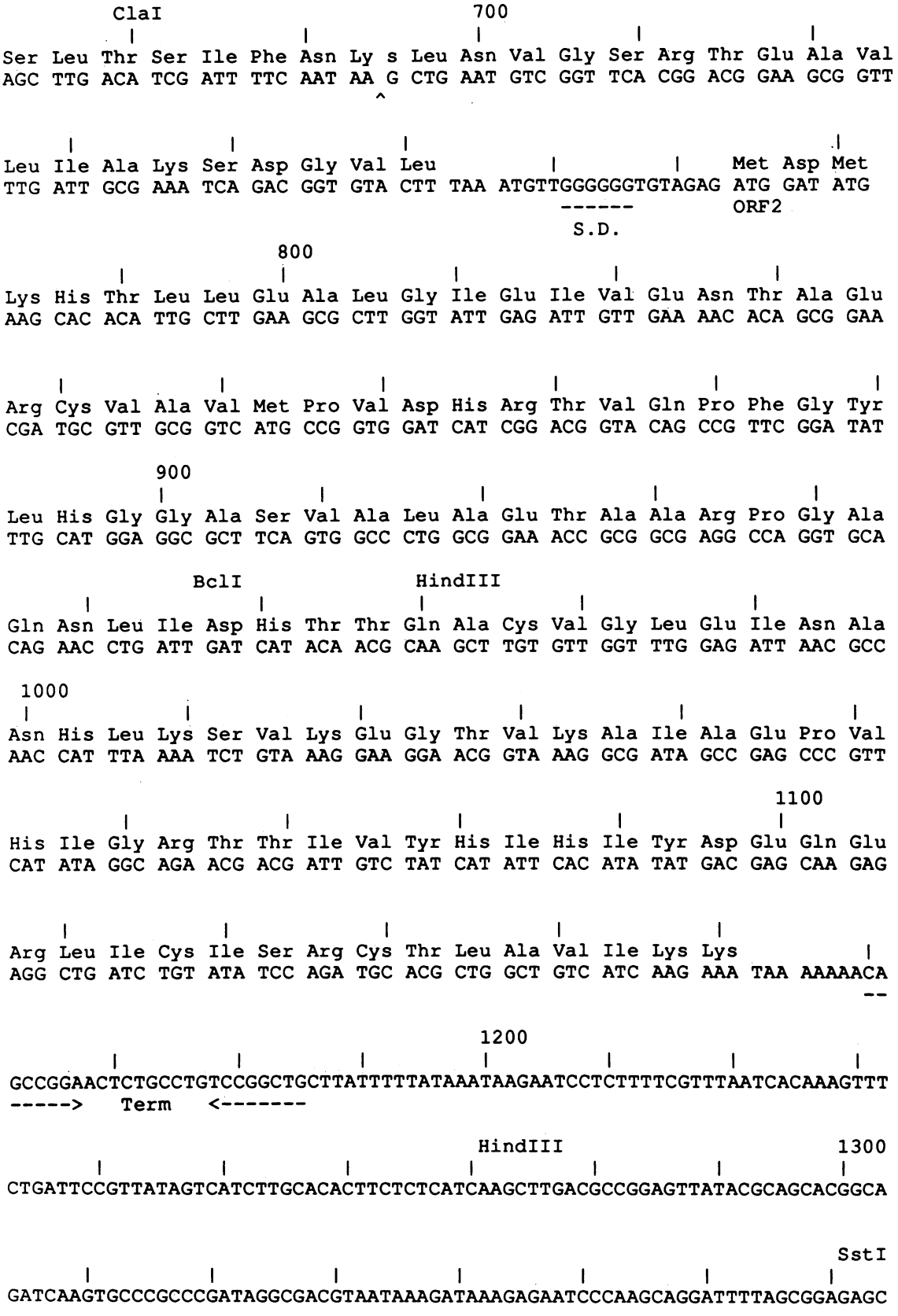

FIG. 1—*Continued*

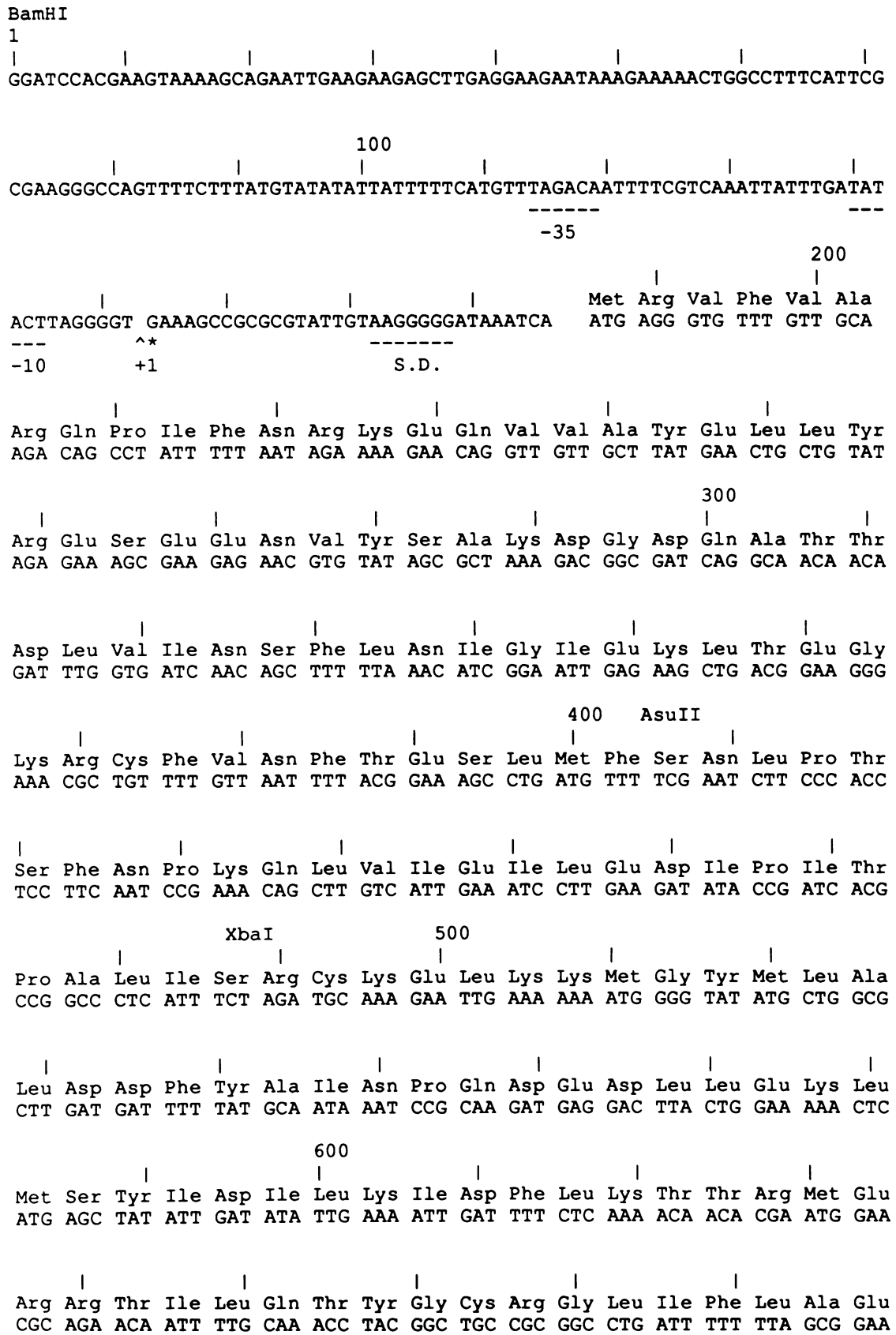

FIG. 2. Nucleotide sequence of *comB*. The complete nucleotide sequence of *comB* is shown together with the predicted amino acid sequence of the *comB* product. Putative −10 and −35 promoter sequences and Shine-Dalgarno (S.D.) sequences (50) are indicated by underlining. The transcriptional start site inferred from S1 nuclease and reverse transcriptase mapping (asterisk) and the site of Tn*917lac* insertion (arrowhead at position 153) are also marked. Converging arrows indicate dyad symmetry.

```
                700
                |
Lys Val Glu Thr Arg Lys Glu Tyr Lys Gln Ala Val Gln Asp Gly Phe Gln Leu
AAA GTA GAG ACC CGA AAA GAA TAT AAA CAG GCG GTC CAA GAC GGC TTT CAA TTA

Phe Gln Gly Tyr Phe Phe Ser Glu Pro Arg Ile Ile Ser Gly His Asp Leu Ser
TTC CAA GGC TAC TTT TTC AGC GAG CCG CGC ATC ATC AGC GGG CAT GAC CTA TCG

  800
  |
Thr His Phe Tyr Ser Tyr Tyr Glu Leu Leu Asn Glu Leu Ser Lys Glu Gln Pro
ACT CAT TTC TAT TCT TAC TAT GAA CTG CTT AAC GAA CTG AGT AAA GAG CAG CCT

                                                        900
                                                        |
Asn Ile Lys Arg Val Thr Glu Tyr Ile Glu Arg Asp Leu Ser Leu Ser Tyr Gln
AAC ATA AAG CGT GTG ACA GAG TAC ATA GAG CGG GAT TTA TCA CTG TCC TAT CAA

Ile Leu Lys Phe Leu Asn Ser Ser His Ser Arg Leu Ser Gln Lys Ile Glu Ser
ATT TTA AAA TTT TTA AAC TCA TCC CAC AGC CGT TTG AGC CAG AAA ATT GAA AGC

                                                1000
                                                |
Ile Gln Gln Ala Ile Met Leu Leu Gly Phe Asn Glu Ile Lys Arg Trp Ile Tyr
ATT CAA CAG GCC ATT ATG CTG CTG GGA TTT AAT GAA ATC AAA CGG TGG ATA TAT

Ile Leu Ser Phe Lys Asp Leu Ser Arg Lys Gly His Ser Ser Lys His Glu Ile
ATT CTT TCC TTT AAG GAT TTA AGC AGG AAA GGG CAT TCC AGC AAG CAC GAA ATC

                                HindIII   1100
                                |         |
Ile Lys Ile Ser Leu Ile Arg Ala Lys Leu Cys Glu Leu Leu Ala Arg Lys Thr
ATT AAG ATT TCT CTG ATA AGA GCA AAG CTT TGC GAA CTG CTG GCG AGA AAG ACA

Ser Arg Pro Gln Pro Ala Ser Tyr Met Leu Ile Gly Met Phe Ser Leu Ile Asp
TCC CGG CCG CAG CCT GCT TCT TAT ATG CTG ATC GGA ATG TTT TCT CTC ATA GAC

                                1200
                                |
Thr Leu Leu His Arg Glu Ile Glu Glu Ile Val Gln Glu Leu Pro Leu Lys Asp
ACC CTC CTG CAT AGA GAA ATA GAG GAA ATT GTT CAA GAA TTG CCT TTA AAA GAT

Glu Val Gly Gln Ala Leu Leu Gly His Gln Asn Asp Tyr Tyr Gln Met Leu Glu
GAA GTC GGG CAA GCA TTA TTA GGC CAT CAA AAC GAC TAC TAC CAA ATG CTT GAG

                        1300
                        |
Leu Val Lys Leu Ile Glu Ser Asn Asn Trp Asp Thr Cys Ser Glu Leu Gly Asn
CTT GTG AAA TTA ATT GAA AGC AAC AAC TGG GAC ACT TGC TCA GAA TTA GGC AAT
```

FIG. 2—*Continued*

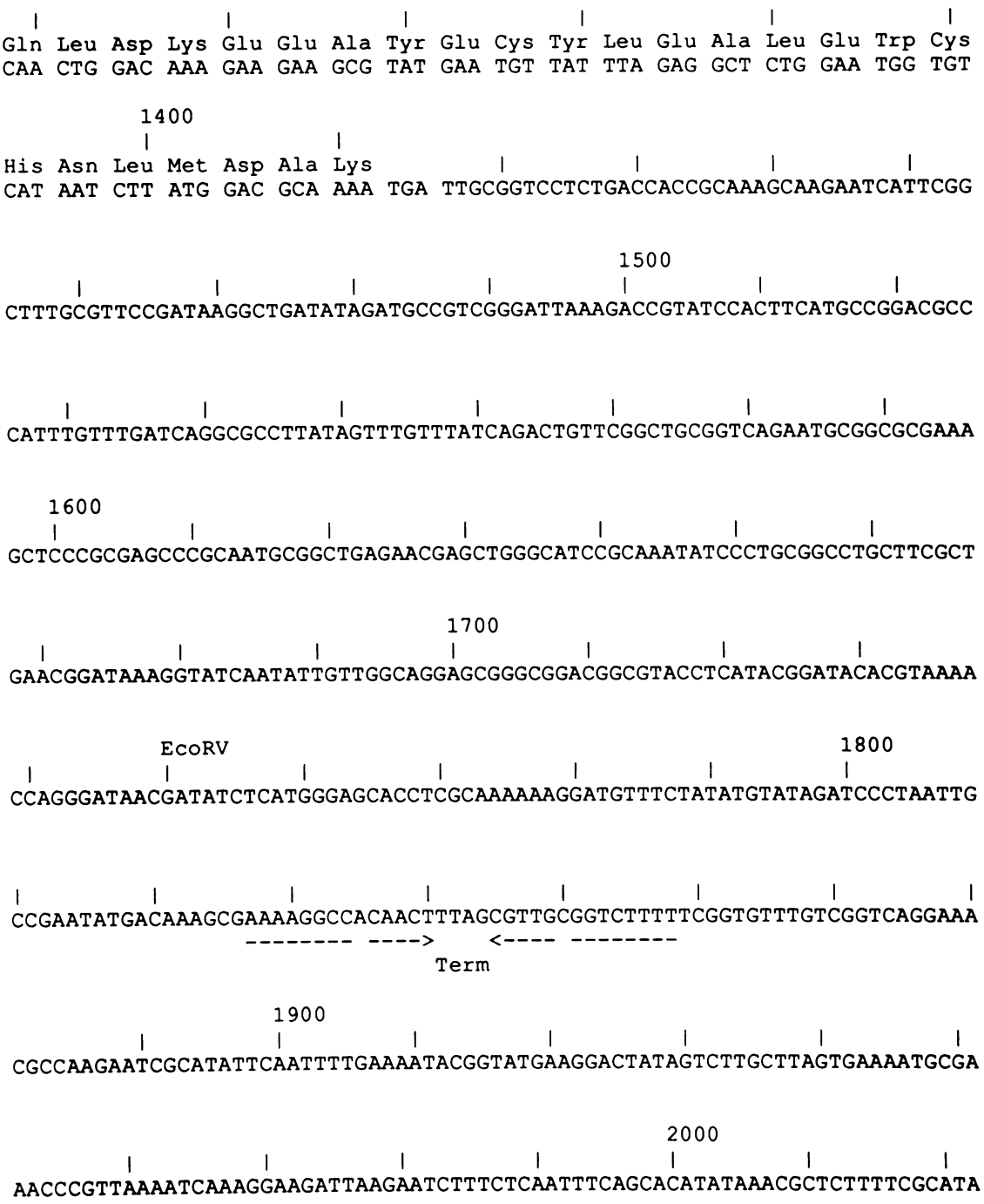

FIG. 2—*Continued*

transposon itself (GGGGT) (42). A predilection for insertion into homologous sequences has been noted for the related transposon Tn*3* (56).

The *comA* sequence revealed two ORFs. The first extended from position 111 to a TAA stop codon at position 753. ORF1 encoded a 214-amino-acid protein with a predicted molecular weight of 24,132. The ATG start codon was preceded by a potential ribosome-binding site, AAGGG AGG (50). The point of insertion of Tn*917lac* was between positions 794 and 795 in codon 195 of the predicted ORF1 protein. The second ORF of *comA* encoded a predicted 126-amino-acid protein with a molecular weight of 13,778, assuming that translation began at the first of the two ATG codons at positions 774 and 780. ORF2 was preceded by a potential ribosome-binding site (GGGGGG) and ended with a TAA codon. The two ORFs were separated by an intergenic region of 18 base pairs (bp). The predicted ORF1 protein exhibited amino acid sequence similarity to a group of procaryotic regulatory proteins that is involved in signal transduction. These results are discussed below. Just downstream from ORF2 was a hyphenated dyad, followed by a string of T residues. Low-resolution S1 nuclease mapping has placed the likely major *comA* terminator at about 510 bp beyond the *Cla*I site at position 681 (18). The string of T

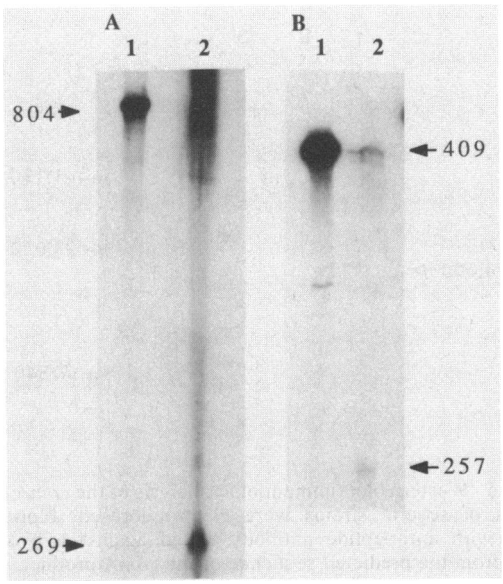

FIG. 3. S1 nuclease protection mapping of the *comA* (A) and *comB* (B) transcripts. For *comA*, an 804-bp *Hin*dIII-*Acc*I fragment that was 5′-end-labeled at the *Acc*I site was used as a probe. For *comB*, a 409-bp *Bam*HI-*Asu*II fragment that was 5′-end-labeled at the *Asu*II site was used. The lengths (in base pairs) of the probe and protected fragments, which were inferred from dideoxynucleotide sequencing reactions run on the same gels as size standards (data not shown), are indicated. Lanes 1, Untreated probe; lanes 2, RNA at 1 h before the transition from the exponential to the stationary growth phase that was hybridized to probe and treated with S1 nuclease.

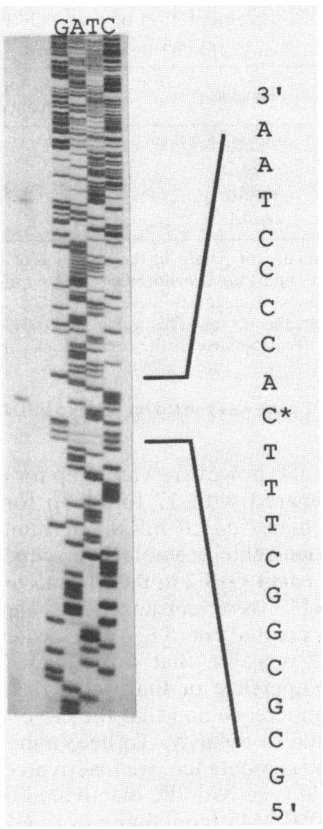

FIG. 4. Primer extension analysis of *comB* transcription. The signals obtained from primer extension analysis of *comB* (performed as described in the text) are shown in the leftmost lane. Also shown are dideoxynucleotide sequencing reactions carried out with the same primer and template that were used for the reverse transcriptase reaction. The position of the major 5′ terminus is indicated with an asterisk on the inferred sequence. The inferred sequence is the complement of the sequence shown in Fig. 2.

residues following this putative terminator began at position 1190, in agreement with the S1 nuclease mapping data.

The *comB* nucleotide sequence contained a single ORF that encoded a predicted protein of 409 amino acids (46,693 daltons). The predicted translational start (ATG) was preceded by an AAGGGGG sequence that probably serves for ribosome recognition. About 450 bases downstream from the *comB* termination codon (TGA) was a dyad symmetry element, followed by a string of 5 T residues. Low-resolution S1 nuclease mapping placed the transcriptional termination site of *comB* at about 1.4 kb downstream from the *Xba*I site (position 486). The predicted termination would therefore occur at position 1886, and the string of T residues would begin at position 1855, which is in good agreement with the mapping data.

High-resolution mapping of the 5′ end of *comA* and *comB* transcripts. To determine the probable transcriptional start site of *comA* by S1 nuclease protection mapping, an 804-bp *Hin*dIII-*Acc*I probe that was 5′-end-labeled at the *Acc*I site was hybridized to RNA isolated from the *com*$^+$ strain grown in competence medium to minus 1 h before the transition from the exponential to the stationary growth phase. A 409-bp *Bam*HI-*Asu*II probe that was 5′-end-labeled at the *Asu*II site was used to map the *comB* transcript. The locations of the *Acc*I and *Asu*II sites used for labeling are shown in Fig. 1 and 2. The protected fragment shown in Fig. 3 suggests that the 5′ terminus of the *comA* transcript is located at the tandem A residues located at positions 85 to 86 in the sequence (Fig. 1). The *comB* results suggest a 5′ terminus at the T or G residues located at positions 152 to 153 (Fig. 2). These results are in excellent agreement with the results of low-resolution S1 nuclease mapping (18), which predicted start sites for *comA* and *comB* at positions 80 and 150, respectively. However, since the size standards used for these mapping experiments were derived from dideoxynucleotide sequencing reactions, an uncertainty of several bases existed as to the precise 5′ termini of these transcripts, since not all DNA fragments of a given length migrates identically. We therefore attempted to confirm the 5′ termini inferred from S1 nuclease mapping by primer extension mapping. This was unsuccessful in the case of *comA*, since no signal was obtained, perhaps because of the low abundance of this mRNA. In the case of *comB*, however, a clear signal was obtained and is shown in Fig. 4, together with dideoxynucleotide sequencing reactions that were carried out with the same primer and template. The results suggest that the 5′ terminus is at the G residue that was inferred from the S1 nuclease mapping experiment (position 153) (the sequence read from Fig. 4 is the complement of the sequence in Fig. 2). In addition, a minor 5′ terminus was evident at the tandem G residues located at positions 59 to 60.

Sequences preceding the major 5′ termini of both transcripts (Fig. 1 and 2) showed good similarity at the −35 and −10 regions with those promoters recognized by the *B. subtilis* σ^A form of RNA polymerase. There was a 5- of 6-base match at the −35 and the −10 regions of *comA* and *comB* with the consensus sequence (TTGACA and TATAAT, respectively) (32, 34). The spacing between these

TABLE 2. Complementation of *comA124* by multicopy *comA* ORF1

Chromosome	Plasmid	Leu$^+$ transformation (transformants/CFU, normalized)[a]
com$^+$	Vector[b]	1.0
com$^+$	*comA*[c]	2.5
comA124	Vector	1.5×10^{-2}
comA124	*comA*	2.0

[a] Cultures were made competent in the presence of chloramphenicol, to maintain the plasmid, and were transformed with 1 μg of chromosomal DNA per ml. Transformation frequencies were normalized to the value for the *com*$^+$ strain carrying the vector. This value was 0.04%, which was about fivefold lower than usual because of the presence of chloramphenicol in the growth medium.

[b] pBD347.

[c] The *comA* ORF1 gene was carried on plasmid pBD440.

sequence elements, however, was 16 bp for *comA* and 20 bp for *comB*, compared with 17 to 19 bp for the consensus sequence. The minor *comB* mRNA terminus was not preceded by a recognizable promoterlike sequence.

Disruption of *comA* ORF2 in the chromosome. Since *comA* ORF1 and ORF2 were cotranscribed and the Tn*917lac* transposon that created *comA124* was located in ORF1 (Fig. 1), it remained possible that both ORFs 1 and 2 were required for competence or that only ORF1 or ORF2 was required for competence and that the effect of the Tn*917lac* insertion was due to polarity. To determine whether ORF2 was essential for competence, we inactivated it and replaced ORF2 in the wild type with the inactivated ORF2 construct. Plasmid DNA (pBD443) containing both ORF1 and ORF2, with ORF2 inactivated by insertion of a Cm^r cassette, was linearized and used to transform a *com*$^+$ strain, replacing the wild-type ORF2 with the mutant version by a double-crossover event. The resulting transformants exhibited wild-type levels of transformation. We conclude that *comA* ORF1, but not ORF2, is required for competence.

***comA* ORF1 cloned in multiple copies complements the *comA124* defect and partially inhibits sporulation.** A 1.4-kb *Hin*dIII fragment carrying ORF1, about half of ORF2, and about 550 bp upstream of the *comA* transcriptional start site was cloned into pBD347 to create pBD440. This plasmid uses the pIM13 replicon and has a likely copy number of 150 to 200 (30). pBD440 was introduced into isogenic *com*$^+$ and *comA124* strains by transformation. The latter construction was possible since the *comA124* defect is leaky (19). Both strains were grown to competence and transformed with chromosomal DNA. In the experiment for which the results are shown in Table 2, the *comA* defect imparted a 100-fold deficiency in transformation, compared with a usual deficiency of about 400-fold (19). The presence of *comA* ORF1 in the multicopy plasmid restored transformation to a level somewhat higher than that in the *com*$^+$ strain. The ability of ORF1 to complement the *comA124* lesion was consistent with the results of the inactivation experiment, which showed that ORF2 is not required for competence.

We noticed that the strain carrying ORF1 on a multicopy plasmid appeared to exhibit a sporulation-deficient phenotype on solid medium. We therefore tested in liquid medium the frequency of sporulation of strain BD630 (*com*$^+$) carrying ORF1 on a multicopy plasmid and strain BD630 carrying the plasmid vector alone. The strain carrying the vector pBD347 yielded a total of 3.1×10^8 viable CFU, of which 2.8×10^8 (90%) were heat resistant. The strain carrying the *comA* multicopy plasmid yielded 3.5×10^8 CFU, of which 1.7×10^7 (4.9%) were heat resistant, confirming the impression obtained from the inspection of colonies on solid medium.

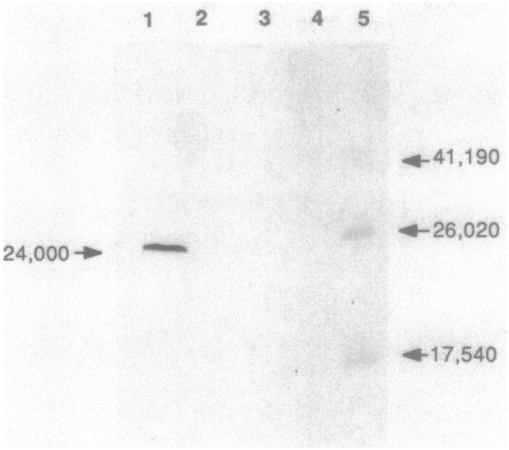

FIG. 5. Western blot (immunoblot) analysis of the *comA* protein. Extracts of several strains were electrophoresed, blotted, and reacted with anti-peptide antibody raised against a hydrophilic peptide from the predicted sequence of the *comA* product. Lane 1, Strain carrying *comA* ORF1 on a multicopy plasmid (pBD440); lane 2, *com*$^+$ (wild type) strain; lane 3, *comA124* strain; lane 4, strain carrying the plasmid vector pBD347 alone; lane 5, prestained size markers. The apparent molecular weights of the prestained standards are indicated, together with the inferred molecular weight of the signal detected in lane 1.

Detection of *comA* protein by immunoblotting. Anti-peptide antibody was raised against the predicted ORF1 product as described in the Materials and Methods and was tested against extracts prepared from the *com*$^+$ strain (BD630), the *comA124* strain, the strain carrying ORF1 on a multicopy plasmid (pBD440), and the strain carrying the plasmid vector (pBD347) alone. All of these strains were isogenic. Figure 5 shows the results of immunoblot analysis. The anti-peptide antibody to ORF1 reacted strongly with a single protein with a molecular weight of approximately 24,000, which corresponded to the predicted molecular weight of the ORF1 product. This signal was detected only in the strain that harbored the multicopy *comA* plasmid, presumably because of the low abundance of the protein. The reaction of the antibody with a protein of the expected size is strong evidence that the predicted *comA* protein is present in *B. subtilis*.

DISCUSSION

The *comA* and *comB* promoters. In this report we presented the nucleotide sequences of two early competence genes of *B. subtilis* that are required for the transcription of several late-expressing *com* genes (18). One of these genes (*comA*) is also required for the expression of the growth stage-regulated molecule surfactin (35). Studies of *lacZ* fusions to *comA* and *comB* have shown that these genes are transcribed in all media tested and throughout growth (1). It is thus not surprising that they both appeared to utilize fairly typical σ^A-type promoter sequences, since the latter form of RNA polymerase holoenzyme transcribes most vegetatively expressed promoters. The suboptimal spacing of the −35 and −10 sequences was consistent with the rather low levels of expression of these genes that have been inferred from β-galactosidase measurements (1). A degree of uncertainty exists with regard to the precise location of the *comA* 5′

mRNA terminus because of the failure of the primer extension experiment. However, the uncertainty is probably not greater than 2 or 3 bases, and no other recognizable promoter sequence was present in the vicinity of the experimentally determined 5′ terminus. An additional general uncertainty was due to the inability of either S1 nuclease or primer extension analysis to rigorously assign a transcriptional start site, since posttranscriptional processing may have played a role. Nevertheless, we provisionally conclude that both *comA* and *comB* are read by the σ^A form of RNA polymerase. *comA* and *comB* are transcribed in the same direction and are located about 3 kb from one another on the *B. subtilis* chromosome (18). Nevertheless, they appear to be transcriptionally independent. As noted previously (18), low-resolution S1 nuclease mapping of both *comA* and *comB* showed that probe-sized DNA was protected, suggesting that readthrough from upstream promoters may also be occurring. Further analysis of transcription should clarify this.

The *comA* and *comB* gene products. The *comB* transcription unit consists of a single ORF. The predicted protein encoded by *comB* bears no significant similarity to any protein in the translated GenBank data base. The *comB* protein would be quite hydrophilic and would possess a preponderance of negative charges in solution, and it would contain 7 cysteine residues.

The *comA* locus contains two ORFs, of which only the first one is required for competence. The reaction of anti-peptide antibody with a protein of the predicted molecular weight for the ORF1 product suggests that the protein is synthesized in vivo and that our nucleotide sequence is correct. The ORF1 product, like that of *comB*, is hydrophilic with a formal preponderance of negatively charged amino acids. Comparison of the predicted sequence of ORF1 with the GenBank data base has revealed that this protein is similar to several members of the effector class of the procaryotic two-component regulatory system (38, 44, 59). Figure 6 presents this comparison for a selected group of effectors. The *comA* protein was similar to the various effector-type proteins predominantly in its amino-terminal region. This is characteristic of the effector proteins as a class. Particularly noteworthy were the conserved tandem D residues at positions 8 to 9 (on the *comA* sequence) and the conserved D residue at position 55 (and possibly at position 50). Patches of sequence similarity were located surrounding these residues. These residues are thought to be potential sites of phosphorylation (58).

A more quantitative comparison was also carried out by using the RDF program (26) to test the significance of the matches between the *comA* protein and each of several other effectors. The *comA* protein sequence was first randomized 100 times. Each of the randomized sequences was then compared with one of the other effector sequences, and a comparison score was obtained. The average of these scores (A) and their standard deviation (S) was computed, and a z value was then obtained, as follows: $z = (A - C)/S$, where C is the original comparison score between the *comA* product and the effector protein. The values (numbers of standard deviations) obtained were as follows: *degU* (21, 25), 23; ORF2 of *uvrC* (31), 23; *ntrC* (5), 9.1; *fixJ* (10), 7.4; ORF2 of *agr* (41), 7.1; *uhpA* (16), 6.8; *virG* (59), 6.3; *dye* (11), 4.8; *phoP* (48), 4.3; *ompR* (7), 3.5; *cheB* (52), 2.3; *dctD* (43), 1.7; *spo0F* (55, 62), 0.45; *cheY* (52), 0.4; *spo0A* (15, 24), −0.9.

These values, which ranged from not significant to highly significant, suggest that the *comA* protein is particularly related to the products of the *B. subtilis* gene *degU* (21, 25, 53) and to ORF2 of the *E. coli uvrC* locus (31, 49). Examination of the matches between the *comA* product and these proteins revealed that the similarities extend over the entire length of the polypeptides. This was also true of the comparison between the *comA* and *fixJ* sequences. The particular closeness of these few proteins is not merely a reflection of phylogenetic relatedness. For instance, the similarities between the *comA* product and those of *spo0A* and *spo0F* are weak, although these are all *B. subtilis* proteins. The similarity of the *comA* protein to the *B. subtilis phoB* product was intermediate in strength. The product of *uvrC*-ORF2 was very similar to that of *comA*, although the former is an *E. coli* protein. These relationships may therefore reflect the conservative influence of functional constraints. The role of the *uvrC*-ORF2 protein is unclear. Although it is cotranscribed with *uvrC*, inactivation of this ORF does not produce a recognizable phenotype (31). However, we may reasonably suspect that the *uvrC*-ORF2 polypeptide may play a role in transducing environmental information to the DNA repair system, perhaps under particular conditions or at a particular stage of growth. There is a relationship between DNA repair and competence in *B. subtilis*. A set of genes that is induced when DNA is damaged is also induced in the competent state (27). It is not yet known whether the *comA* product is required for induction of this set of genes, but we suspect that *comA* and *uvrC*-ORF2 may play analogous roles. The *degU* gene product is not only involved in regulating the synthesis of several *B. subtilis* extracellular enzymes during the transition to stationary phase (23) but is also required for the development of competence (53). Perhaps, then, the *comA*, *degU*, and *uvrC*-ORF2 proteins comprise a subclass of effectors that are related functionally, and the similarities of these molecules throughout their lengths may reflect some special features of their modes of action. Also noteworthy was the similarity of the *comA* gene product to that of *malT*, a transcriptional regulator of the *E. coli* maltose regulon (6, 47) (data not shown). This similarity was significant (z value, 9.3 standard deviations) as was the similarity between the *malT* and *degU* proteins (z = value, 10.2) (25). The *comA* and *degU* proteins resembled the C-terminal moiety of the 901-amino-acid *malT* protein over most of their lengths. However, there was no significant resemblance of the *malT* and *uvrC*-ORF2 products (z value, −0.9), suggesting that the relationships among these various regulatory proteins are complex.

In the case of the regulation of nitrogen metabolism in *E. coli*, it is known that the effector (*ntrC*) becomes a transcriptional activator when it is phosphorylated (36, 58). The C-terminal region of *ntrC* is directly involved in modulating transcription (33). Since *comA* is required for the transcription of several late competence genes, it was reasonable to search for evidence that its protein product may interact with DNA. Figure 7 presents a comparison that suggests the presence of a helix-turn-helix-type DNA-binding motif in the C-terminal region of the predicted *comA* sequence. Similar sequences were evident at corresponding locations in the C-terminal regions of the *uhpA*, *uvrC*-ORF2, and *fixJ* proteins (Fig. 6, positions 171 to 185 on the *comA* sequence).

Possible biological roles for the *comA* and *comB* products. The analysis presented above strongly suggests that the *comA* gene product is a transcriptional effector that responds to some signal that is essential for the initiation of competence development. Nakano and Zuber (35) have shown that *comA* (called *srfB* in their report) is also required for expression of a surfactin molecule in *B. subtilis* during the stationary phase of growth, demonstrating that this effector

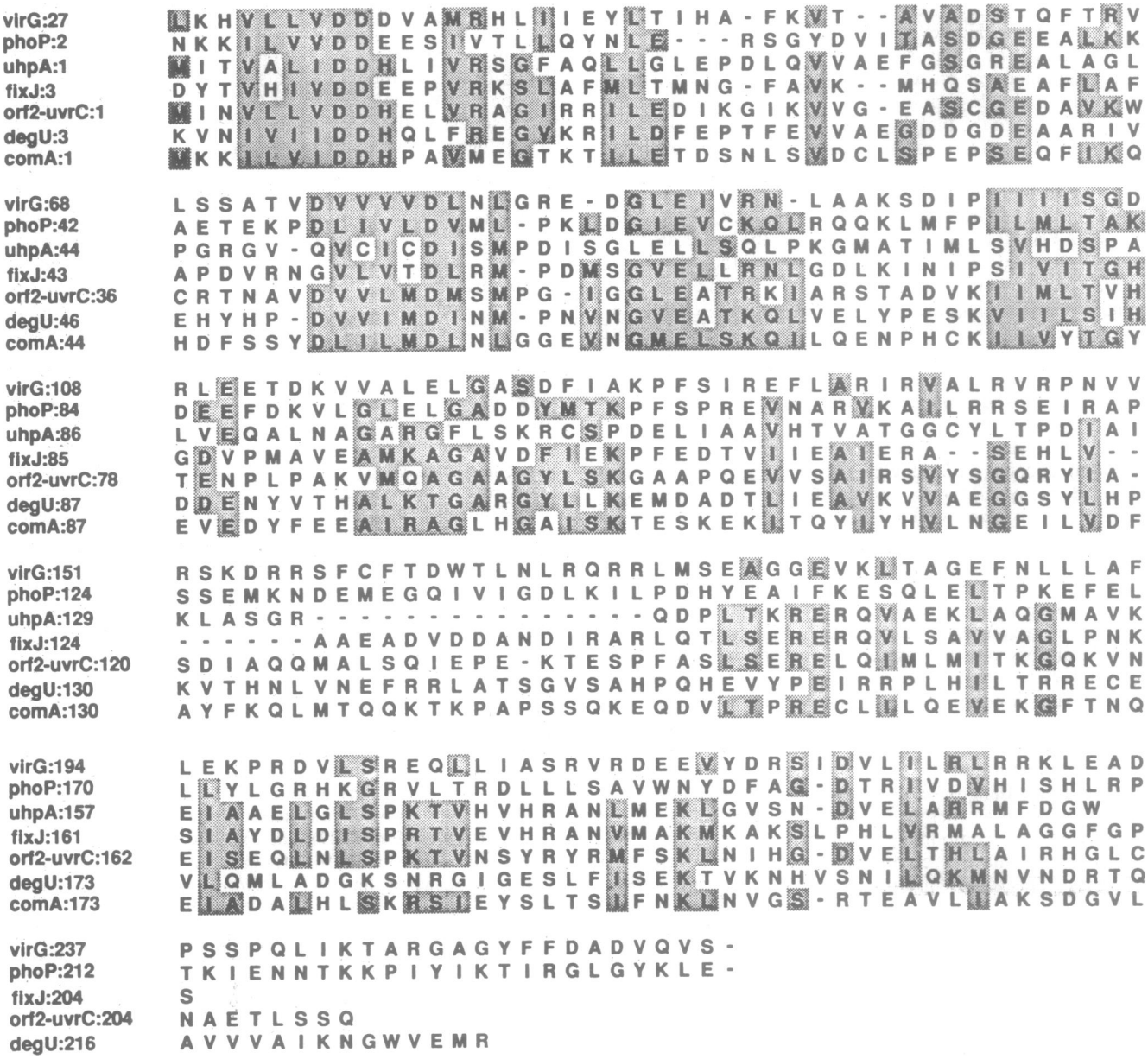

FIG. 6. Comparison of the predicted amino acid sequences of the *comA* protein with those of other effector proteins. The FASTP program (26) was used to align the sequences. In a few places (notably, in the case of *uhpA*), the alignment was improved by manual insertion of dashes. Amino acids that were regarded as functionally related are shaded. Functionally related residues are classified as follows: M, L, I, and V; H, K, and R; D and E; Q and N; A, G, S, and T; F, W, and Y; P; and C. The sources of the protein sequences are given in the text.

is responsible for the regulation of a disparate set of genes. In addition, competence development requires *spo0A* and *degU*, whose products have also been found to be members of the effector protein class (15, 21, 24, 25, 53) and to be required for the expression of several disparate late growth functions. Why are at least three effectors required? A reasonable speculation is that complexity at this level enables the cell to assess several independent environmental or intracellular conditions, a particular combination of which must be satisfied before competence is initiated. It is interesting that a mutation in any one of these three effector genes confers a leaky competence phenotype, typified by a several hundred- to several thousand-fold deficiency. In contrast, mutations in the late competence genes confer absolute deficiency (at least 10^6-fold) (19). This is consistent with the proposed regulatory roles of these effectors and suggests that in the absence of any one of them, the others are still capable of carrying out a residual level of transcriptional activation.

What might be the role of the *comB* protein in these processes? The phenotype of *comB138* is indistinguishable from that of *comA124*. Both mutants are deficient in competence to the same extent (several hundred-fold) and both fail to express late competence genes (18). A double *comA comB* mutant strain is no more competence deficient than is either single mutant strain (unpublished data). These facts suggest that *comB* may play a role in the *comA* signal transduction pathway, perhaps modulating signal recognition by a sensor, transduction of the signal to the *comA* protein, activity of the *comA* protein as a transcriptional factor, etc. One cautionary

```
lambda-cI     QESVADKMGMGQSGVGALFN
ampR          FTHAAIELNVTHSAISQHVK
lysR          LTEAAHLLHTSQPTVSRELA
ilvY          FGRSARAMHVSPSTLSRQIQ
metR          LAAAAAVLHQTQSALSHQFS
tfdO          VGAAARRLHISQPPVTRQIH
comA          NQEIADALHLSKRSIEYSLT
              ****** **********  *
consensus     ppphGp-hGhpppplpph-p
                 A   H      V
```

FIG. 7. Possible helix-turn-helix motif in the predicted *comA* protein. Amino acid residues 170 to 189 from the C-terminal region of the *comA* protein were compared with a group of probable helix-turn-helix DNA-binding sequences that has been compiled recently (20). The sequences selected for this comparison differed from most of the previously recognized helix-turn-helix sequences (39), most notably in the presence of an H residue in place of a conserved G residue (present in λ *cI*). Also shown is an inferred helix-turn-helix consensus sequence taken from Siegele et al. (51), in which p and h represent polar and hydrophobic residues, respectively. The asterisks denote the positions in which the *comA* protein sequence conforms to this consensus sequence.

point must be made concerning the *comB138* phenotype. The insertion point of Tn*917lac* in this mutant was within 1 to 2 bases of the transcriptional start site and was not within the coding region. It is possible that a low level of *comB* transcription occurs from a weak promoter within Tn*917lac* and that the real phenotype of a *comB* null mutation would be different from that of *comB138*.

The partial inhibition of sporulation in a strain carrying multiple copies of *comA* is subject to several interpretations. It is possible that *comA* transcriptional or translational control regions are titrating a limiting factor that is required for sporulation. It is also possible that the overproduction of *comA* protein (Fig. 5) is responsible for the inhibition. One intriguing possibility is that the *comA* protein competes with an essential sporulation effector molecule for the attentions of a common sensor or a sensor with less than absolute specificity (so-called cross-talk [37]).

For most of the known effectors a sensor (also called a modulator) has been identified which, in at least some cases, is autophosphorylated and transfers a phosphate group to its cognate effector (22, 28, 58, 60). This results in activation of the effector function and the transduction of information to a target system. Whether such a sensor-modulator exists for *comA* is a matter that deserves attention.

ACKNOWLEDGMENTS

We acknowledge valuable discussions with J. Hahn, M. Albano, L. Dempsey, I. Smith, E. Dubnau, F. Breidt, and A. Ninfa. We also thank M. M. Nakano and P. Zuber for communicating results prior to publication. We are especially appreciative of the help and advice provided by J. Oppenheim in preparing the anti-*comA* antiserum.

This work was supported by Public Health Service grant AI10311 from the National Institutes of Health. Computer facilities at the Public Health Research Institute were supported by Public Health Service grant RRNA-02990 from the National Institutes of Health and grant DBM-8502189 from the National Science Foundation. N.G. is a fellow of the French Centre National de la Recherche Scientifique.

LITERATURE CITED

1. **Albano, M., J. Hahn, and D. Dubnau.** 1987. Expression of competence genes in *Bacillus subtilis*. J. Bacteriol. **169**:3110–3117.
2. **Berk, A. J., and P. A. Sharp.** 1977. Sizing and mapping of early adeno-virus mRNAs by gel electrophoresis of S1 endonuclease-digested hybrids. Cell **12**:721–732.
3. **Birnboim, H. C., and J. Doly.** 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res. **7**:1513–1523.
4. **Blake, M., K. Johnston, G. Russell-Jones, and E. Gotschlich.** 1984. A rapid, sensitive method for detection of alkaline phosphatase-conjugated anti-antibody on Western blots. Anal. Biochem. **136**:175–179.
5. **Buikema, W. J., W. W. Szeto, P. V. Lemley, W. H. Orme-Johnson, and F. M. Ausubel.** 1985. Nitrogen fixation specific regulatory genes of *Klebsiella pneumoniae* and *Rhizobium meliloti* share homology with the general nitrogen regulatory gene *ntrC* of *K. pneumoniae*. **13**:4539–4555.
6. **Cole, S. T., and O. Raibaud.** 1986. The nucleotide sequence of the *malT* gene encoding the positive regulator of the *Escherichia coli* maltose regulon. Gene **42**:201–208.
7. **Comeau, D. E., K. Ikenaka, K. Tsung, and M. Inouye.** 1985. Primary characterization of the protein products of the *E. coli ompB* locus: structure and regulation of synthesis of the OmpR and EnvZ proteins. J. Bacteriol. **164**:578–584.
8. **Contente, S., and D. Dubnau.** 1979. Marker rescue transformation by linear plasmid DNA in *Bacillus subtilis*. Plasmid **2**: 555–571.
9. **Dale, R. M. K., B. A. McClure, and J. P. Houchins.** 1985. A rapid single-stranded cloning strategy for producing a sequential series of overlapping clones for use in DNA sequencing: application to sequencing the corn mitochondrial 18s rDNA. Plasmid **13**:31–40.
10. **David, M., M. L. Daveran, J. Batut, A. Dedieu, O. Domergue, J. Ghai, C. Hertig, P. Boistard, and D. Kahn.** 1988. Cascade regulation of *nif* gene expression in *Rhizobium meliloti*. Cell **54**:671–683.
11. **Drury, L. S., and R. S. Buxton.** 1985. DNA sequence analysis of the *dye* gene of *Escherichia coli* reveals amino acid sequence homology between the Dye and OmpR proteins. J. Biol. Chem. **260**:4236–4242.
12. **Dubnau, D., R. Davidoff-Abelson, and I. Smith.** 1969. Transformation and transduction in *Bacillus subtilis*: evidence for separate modes of recombinant formation. J. Mol. Biol. **45**:155–179.
13. **Engelman, D. M., T. A. Steitz, and A. Goldman.** 1986. Identifying nonpolar transbilayer helices in amino acid sequences of membrane proteins. Annu. Rev. Biophys. Biophys. Chem. **15**:321–353.
14. **Favaloro, J., R. Treisman, and R. Kamen.** 1980. Transcription maps of polyoma virus-specific RNA analysis by two-dimensional nuclease S1 gel mapping. Methods Enzymol. **65**:718–749.
15. **Ferrari, F. A., K. Trach, D. LeCoq, J. Spence, E. . Ferrari, and J. A. Hoch.** 1985. Characterization of the *spo0A* locus and its deduced product. Proc. Natl. Acad. Sci. USA **82**:2647–2651.
16. **Friedrich, M. J., and R. J. Kadner.** 1987. Nucleotide sequence of the *uhp* region of *Escherichia coli*. J. Bacteriol. **169**:3556–3563.
17. **Guerry, P., D. J. Leblanc, and S. Falkow.** 1973. General method for the isolation of plasmid deoxyribonucleic acid. J. Bacteriol. **116**:1064–1066.
18. **Guillen, N., Y. Weinrauch, and D. Dubnau.** 1989. Cloning and characterization of the regulatory *Bacillus subtilis* competence genes *comA* and *comB*. J. Bacteriol. **171**:5354–5361.
19. **Hahn, J., M. Albano, and D. Dubnau.** 1987. Isolation and characterization of competence mutants in *Bacillus subtilis*. J. Bacteriol. **169**:3104–3109.
20. **Henikoff, S., G. Haughn, J. M. Calvo, and J. C. Wallace.** 1988. A large family of bacterial activator proteins. Proc. Natl. Acad. Sci. USA **85**:6602–6606.
21. **Henner, D. J., M. Yang, and E. Ferrari.** 1988. Localization of *Bacillus subtilis sacU(Hy)* mutations to two linked genes with similarities to the conserved procaryotic family of two-component signaling systems. J. Bacteriol. **170**:5102–5109.
22. **Hess, J. F., K. Oosawa, N. Kaplan, and M. I. Simon.** 1988.

Phosphorylation of three proteins in the signaling pathway of bacterial chemotaxis. Cell **53**:79–87.
23. **Klier, A. F., and G. Rapoport.** 1988. Genetics and regulation of carbohydrate catabolism in *Bacillus*. Annu. Rev. Microbiol. **42**:65–95.
24. **Kudoh, J., T. Ikeuchi, and K. Kurahashi.** 1985. Nucleotide sequences of the sporulation gene *spo0A* and its mutant genes of *Bacillus subtilis*. Proc. Natl. Acad. Sci. USA **82**:2665–2668.
25. **Kunst, F., M. Debarbouille, T. Msadek, M. Young, C. Mauel, D. Karamata, A. Klier, G. Rapoport, and R. Dedonder.** 1988. Deduced polypeptides encoded by the *Bacillus subtilis sacU* locus share homology with two-component sensor-regulator systems. J. Bacteriol. **170**:5093–5101.
26. **Lipman, D. J., and W. R. Pearson.** 1985. Rapid and sensitive protein similarity searches. Science **227**:1435–1441.
27. **Love, P. E., M. J. Lyle, and R. E. Yasbin.** 1985. DNA-damage-inducible (*din*) loci are transcriptionally activated in competent *Bacillus subtilis*. Proc. Natl. Acad. Sci. USA **82**: 6201–6205.
28. **Magasanik, B.** 1988. Reversible phosphorylation of an enhancer binding protein regulates the transcription of bacterial nitrogen utilization genes. TIBS **13**:475–479.
29. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
30. **Monod, M., C. Denoya, and D. Dubnau.** 1986. Sequence and properties of pIM13, a macrolide-lincosamide-streptogramin B resistance plasmid from *Bacillus subtilis*. J. Bacteriol. **167**: 138–147.
31. **Moolenaar, G. F., C. A. Sluis, C. Backendorf, and P. Putte.** 1987. Regulation of the *Escherichia coli* excision repair gene *uvrC*. Overlap between the *uvrC* structural gene and the region coding for a 24 kD protein. Nucleic Acids Res. **15**:4273–4289.
32. **Moran, C. P., Jr., N. Lang, S. F. J. LeGrice, G. Lee, M. Stephens, A. L. Sonenshein, J. Pero, and R. Losick.** 1982. Nucleotide sequences that signal the initiation of transcription and translation in *Bacillus subtilis*. Mol. Gen. Genet. **186**: 339–346.
33. **Morett, E., W. Cannon, and M. Buck.** 1988. The DNA-binding domain of the transcriptional activator protein NifA resides in its carboxy terminus, recognizes the upstream activator sequences of *nif* promoters and can be separated from the positive control function of NifA. Nucleic Acids Res. **16**:11469–11488.
34. **Murray, C. L., and J. C. Rabinowitz.** 1982. Nucleotide sequences of transcription and translation initiation regions in Bacillus phage ϕ29 early genes. J. Biol. Chem. **257**:1053–1062.
35. **Nakano, M. M., and P. Zuber.** 1989. Cloning and characterization of *srfB*; a regulatory gene involved in surfactin production and competence in *Bacillus subtilis*. J. Bacteriol. **171**:5347–5353.
36. **Ninfa, A. J., and B. Magasanik.** 1986. Covalent modification of the *glnG* product, Nr_I, by the *glnL* product, NR_{II}, regulates the transcription of the *glnALG* operon in *Escherichia coli*. Proc. Natl. Acad. Sci. USA **83**:5909–5913.
37. **Ninfa, A. J., E. G. Ninfa, A. N. Lupas, A. Stock, B. Magasanik, and J. Stock.** 1988. Crosstalk between bacterial chemotaxis signal transduction proteins and regulators of transcription of the Ntr regulon: evidence that nitrogen assimilation and chemotaxis are controlled by a common phosphotransfer mechanism. Proc. Natl. Acad. Sci. USA **85**:5492–5496.
38. **Nixon, B. T., C. W. Ronson, and F. M. Ausubel.** 1986. Two component regulatory systems responsive to environmental stimuli share strongly conserved domains with the nitrogen assimilation regulatory genes *ntrB* and *ntrC*. Proc. Natl. Acad. Sci. USA **83**:7850–7854.
39. **Pabo, C. O., and R. T. Sauer.** 1984. Protein-DNA recognition. Annu. Rev. Biochem. **53**:293–321.
40. **Pearson, W. R., and D. J. Lipman.** 1988. Improved tools for biological sequence comparison. Proc. Natl. Acad. Sci. USA **85**:2444–2448.
41. **Peng, H. L., R. P. Novick, B. Kreisworth, J. Kornblum, and P. Schlievert.** 1988. Cloning, characterization and sequencing of an accessory gene regulator (*agr*) in *Staphylococcus aureus*. J. Bacteriol. **170**:4365–4372.
42. **Perkins, J. B., and P. Youngman.** 1984. A physical and functional analysis of Tn917, a *Streptococcus* transposon in the Tn3 family that functions in *Bacillus*. Plasmid **12**:119–138.
43. **Ronson, C. W., P. M. Astwood, B. T. Nixon, and F. M. Ausubel.** 1987. Deduced products of C4-dicarboxylate transport regulatory genes of *Rhizobium leguminosarum* are homologous to nitrogen regulatory gene products. Nucleic Acids Res. **15**: 7921–7934.
44. **Ronson, C. W., B. T. Nixon, and F. Ausubel.** 1987. Conserved domains in bacterial regulatory proteins that respond to environmental signals. Cell **49**:579–581.
45. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
46. **Schaeffer, P., J. Millet, and J. Aubert.** 1965. Catabolic repression of bacterial sporulation. Proc. Natl. Acad. Sci. USA **54**:704–711.
47. **Schwartz, M.** 1987. The maltose regulon, p. 1482–1502. *In* F. C. Neihardt (ed.), *Escherichia coli* and *Salmonella typhimurium*. Cellular and molecular biology. American Society for Microbiology, Washington, D.C.
48. **Seki, T., H. Yoshikawa, H. Takahashi, and H. Saito.** 1987. Cloning and nucleotide sequence of *phoP*, the regulatory gene for alkaline phosphatase and phosphodiesterase in *Bacillus subtilis*. J. Bacteriol. **169**:2913–2916.
49. **Sharma, S., T. F. Stark, W. G. Beattie, and R. E. Moses.** 1986. Multiple control elements for the *uvrC* gene unit of *Escherichia coli*. Nucleic Acids Res. **14**:2301–2318.
50. **Shine, J., and L. Dalgarno.** 1974. The 3′-terminal sequence of *Escherichia coli* 16S ribosomal RNA: complementarity to nonsense triplets and ribosome binding sites. Proc. Natl. Acad. Sci. USA **71**:1342–1346.
51. **Siegele, D. A., J. C. Hu, and C. A. Gross.** 1988. Mutations in *rpoD*, the gene encoding the σ^{70} subunit of *Escherichia coli* RNA polymerase, that increase expression of the *lac* operon in the absence of CAP-cAMP. J. Mol. Biol. **203**:29–37.
52. **Stock, A., D. E. Koshland, and J. Stock.** 1985. Homologies between the *Salmonella typhimurium* CheY protein and proteins involved in the regulation of chemotaxis, membrane protein synthesis and sporulation. Proc. Natl. Acad. Sci. USA **82**:7989–7993.
53. **Tanaka, T., and M. Kawata.** 1988. Cloning and characterization of *Bacillus subtilis iep*, which has positive and negative effects on production of extracellular proteases. J. Bacteriol. **170**: 3593–3600.
54. **Towbin, H., T. Staehelin, and J. Gordon.** 1979. Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. Proc. Natl. Acad. Sci. USA **76**:4350–4354.
55. **Trach, K. A., J. W. Chapman, P. J. Piggot, and J. R. Hoch.** 1985. Deduced product of the stage 0 sporulation gene *spo0F* shares homology with the *spo0A*, *ompR*, and *sfrA* proteins. Proc. Natl. Acad. Sci. USA **82**:7260–7264.
56. **Tu, C. P., and S. N. Cohen.** 1980. Translocation specificity of the *Tn3* element: characterization of sites of multiple insertions. Cell **19**:151–160.
57. **Ulmanen, I., K. Lundstrom, P. Lehtovaara, M. Sarvas, M. Ruohonen, and I. Palva.** 1985. Transcription and translation of foreign genes in *Bacillus subtilis* by the aid of a secretion vector. J. Bacteriol. **162**:176–182.
58. **Weiss, V., and B. Magasanik.** 1988. Phosphorylation of nitrogen regulator I (NR_I) of *Escherichia coli*. Proc. Natl. Acad. Sci. USA **85**:8919–8923.
59. **Winans, S. C., P. R. Ebert, S. E. Stachel, M. P. Gordon, and E. W. Nester.** 1986. A gene essential for *Agrobacterium tumefaciens* virulence is homologous to a family of positive regulatory loci. Proc. Natl. Acad. Sci. USA **83**:8278–8282.
60. **Wylie, D., A. Stock, C. Y. Wong, and J. Stock.** 1988. Sensory transduction in bacterial chemotaxis involves phosphotransfer between Che proteins. Biochem. Biophys. Res. Commun. **151**: 891–896.
61. **Yoshikawa, H., H. Sone, T. Seki, F. Kawamura, and H. Saito.**

1986. Revised assignment for the *Bacillus subtilis spo0F* gene and its homology with *spo0A* and two *Escherichia coli* genes. Nucleic Acids Res. **14**:1063–1072.

62. **Youngman, P., J. B. Perkins, and R. Losick.** 1984. A novel method for the rapid cloning in *Escherichia coli* of *Bacillus subtilis* chromosomal DNA adjacent to Tn917 insertions. Mol. Gen. Genet. **195**:424–433.
63. **Youngman, P. J., J. P. Perkins, and R. Losick.** 1983. Genetic transposition and insertional mutagenesis in *Bacillus subtilis* with *Streptococcus faecalis* transposon Tn917. Proc. Natl. Acad. Sci. USA **80**:2305–2309.
64. **Youngman, P., P. Zuber, J. B. Perkins, K. Sandman, M. Igo, and R. Losick.** 1985. New ways to study developmental genes in spore-forming bacteria. Science **228**:285–291.