JOURNAL OF BACTERIOLOGY, Oct. 1992, p. 6699–6702
0021-9193/92/206699-04$02.00/0

Vol. 174, No. 20

# Determination of the Sequence of *spaE* and Identification of a Promoter in the Subtilin (*spa*) Operon in *Bacillus subtilis*

YONG JOON CHUNG AND J. NORMAN HANSEN*

*Department of Chemistry and Biochemistry, University of Maryland, College Park, Maryland 20742*

Received 18 February 1992/Accepted 10 August 1992

**An 851-residue open reading frame (ORF) called SpaE has been discovered in the subtilin (*spa*) operon. Interruption of this ORF with a chloramphenicol acetyltransferase gene destroys the ability of *Bacillus subtilis* LH45Δc (a derivative of *B. subtilis* 168) to produce subtilin, which is an antimicrobial peptide belonging to the class of ribosomally synthesized peptide antibiotics called lantibiotics. SpaE shows strong homology to NisB, which is in the nisin (*nis*) operon in *Lactococcus lactis* ATCC 11454. Despite the strong sequence homology between SpaE and NisB, the *spaE* and *nisB* genes occupy very different locations in their respective operons, indicating that they have been evolving separately for a long time. Primer extension analysis was employed to identify a promoter upstream from the *spaE* gene, which appears to define the 5′ end of the *spa* operon, which contains four other ORFs (Y. J. Chung, M. T. Steen, and J. N. Hansen, J. Bacteriol. 174:1417–1422, 1992).**

Nisin and subtilin are structurally homologous members of a class of antimicrobial peptides called lantibiotics. Nisin is a 34-residue peptide produced by *Lactococcus lactis* ATCC 11454 (6), and subtilin is a 32-residue peptide produced by *Bacillus subtilis* ATCC 6633 (5). Nisin, subtilin, and the other lantibiotics contain unusual lanthionine and dehydro residues and are formed from ribosomally synthesized peptide precursors that undergo extensive posttranslational modification and export outside the cell. The genes for the prepeptides of subtilin and of nisin have been cloned, and their kinetics of expression have been characterized (1, 2). Because lantibiotics have novel structures, it is probable that the maturation pathway involves genes that encode novel proteins. We have been searching for genes involved with lantibiotic maturation within the operons that contain the genes for the precursor peptides. The nisin structural gene (*nisA*) appears to be cotranscribed with a downstream gene (*nisB*) (11), and the subtilin structural gene (*spaS*) is in the same transcriptional unit as several genes (*spaD*, *spaB*, and *spaC*) that lie upstream from *spaS* (3). The deduced SpaB protein sequence shows extensive homology to a variety of transport proteins, including the hemolysin B and multidrug resistance proteins (3), and may be involved in subtilin export.

The nucleotide sequence upstream from the previously sequenced (3) *spa* genes was determined by cloning restriction fragments into M13mp18 or M13mp19 and sequencing by the dideoxy chain termination method with the Sequenase kit supplied by U.S. Biochemical Corp. (Cleveland, Ohio). DNA was isolated by standard methods (9). Long restriction fragments were cloned into pTZ18R and incrementally shortened with the Erase-a-base kit supplied by Promega Corp. (Madison, Wis.). The shortened plasmids were purified with GeneClean provided by Bio 101 (La Jolla, Calif.). To obtain the DNA containing this upstream region, a synthetic oligonucleotide based on the sequence of the 5′ end of the *spaD* gene was used as a hybridization probe to clone a 10-kb *Pst*I restriction fragment that contained the 5′ end of *spaD* and its upstream flanking region. The bacterial strain used was LH45, which is a chloramphenicol-resistant subtilin-producing strain of *B. subtilis* (8) derived from *B. subtilis* 168 by transformation with DNA from *B. subtilis* ATCC 6633, which is the natural producer of subtilin. Strain LH45Δc is a chloramphenicol-sensitive derivative of LH45 derived by spontaneous loss of the *cat* gene (8). *Pst*I fragments of genomic DNA prepared from *B. subtilis* LH45 were size selected on agarose and cloned into *Escherichia coli*, and clones containing the 10-kb *Pst*I fragment were identified by hybridization with the probe. This fragment was mapped with restriction enzymes, subcloned, and partially sequenced. Some primers were synthesized with a Biosearch model 8500 synthesizer. Approximately 3.5 kb of the sequence is shown in Fig. 1. The sequence reveals an 851-residue open reading frame (ORF). The gene for this ORF is designated *spaE*, and the deduced protein sequence is designated SpaE. The lack of any obvious terminator sequence between the *spaE* gene and its immediately downstream *spaD* gene argues that *spaE* and *spaD* are in the same transcriptional unit.

Further sequence analysis revealed an intergenic space upstream from the *spaE* gene. This space contained a promoter-like sequence, and upstream from this was an inverted repeat with characteristics of a ρ-independent terminator. This putative terminator is located immediately downstream from an ORF called ORF X since its sequence is incomplete and it is not yet known whether it is involved with subtilin biosynthesis. The presence of this terminator suggests that expression of the *spa* operon is dependent on the promoter sequence that lies between this terminator and the *spaE* gene. Experiments to identify this promoter are described below.

**Interruption of the *spaE* gene prevents production of active subtilin.** The role of the *spaE* gene in subtilin production was tested by integration of a *cat* gene into the SpaE coding region, thus interrupting the continuity of the reading frame. The strategy was similar to that previously used to show that the *spaB* gene is required for subtilin production (3). The *spaE*-containing *Sst*I-*Pst*I restriction fragment (see Fig. 2) was cloned into pTZ18R, and then the *cat* gene was cloned into the *Xba*I site of the insert. This plasmid was linearized by restriction with *Sst*I and *Pst*I, introduced into LH45Δc cells by transformation, and selected on chloramphenicol plates. The transformants should be the result of integration

* Corresponding author.

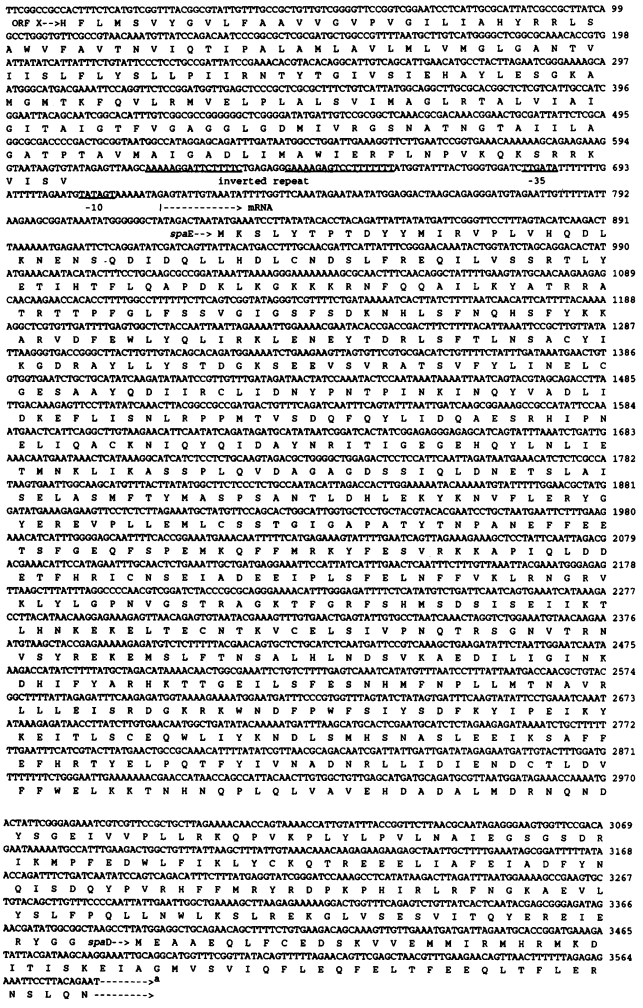
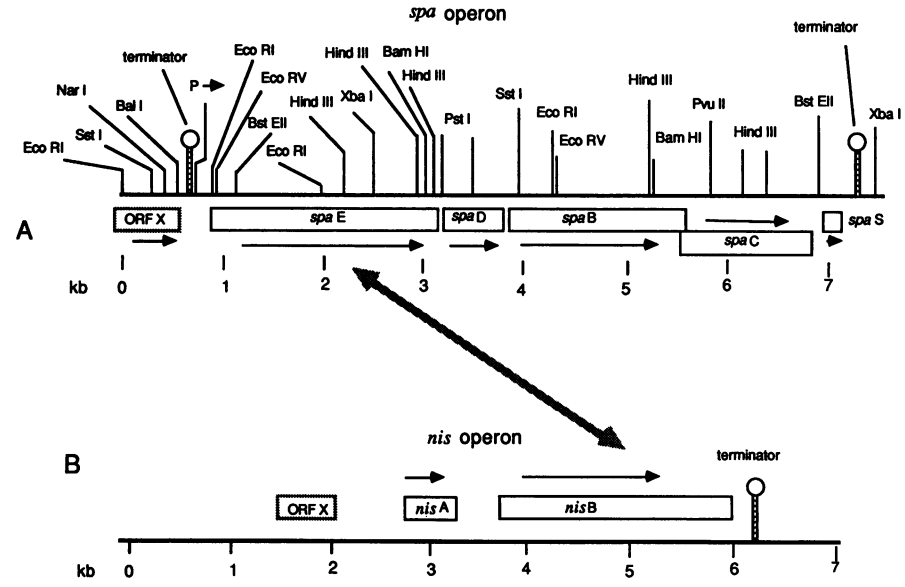


FIG. 1. Nucleotide sequence of the *spaE* gene. The organization of the *spaE* gene within the *spa* operon is shown in Fig. 2. The transcription start site of the *spaE* gene was determined by primer extension (Fig. 4) to give an mRNA that begins with the sequence 5′-GAGTATTGTAAATA→. The −10 and −35 regions of the promoter that correspond to this start site are underlined. Upstream from *spaE* is part of an ORF of unidentified function (ORF X). Immediately downstream from ORF X is an inverted repeat (underlined) with a stretch of T residues that is typical of a ρ-independent terminator. The translation start codon for the SpaE ORF is shown as the first methionine downstream from the transcription start site, although the second methionine has a better ribosome binding site. The deduced protein sequence from the SpaE ORF has strong homology to the NisB ORF from the *nis* operon (Fig. 3). The *Eco*RI restriction site in ORF X shown in Fig. 2 is located approximately 10 nucleotides upstream from residue 1 of the sequence. a, the sequence of the rest of the operon, including *spaD*, *spaB*, *spaC* (3), and *spaS* (2), has been published previously.

of the *cat* gene into *spaE* by a double crossover between the flanking regions, resulting in interruption of the *spaE* gene at the *Xba*I site, which is located about three-fourths of the way into the gene. A halo assay (9) showed that these integrants were unable to produce active subtilin (data not shown). By using the same approach, ORF X was interrupted by integration of the *cat* gene into the *Bal*I site shown in Fig. 2. That the correct insertion occurred was established by Southern analysis of chromosomal DNA (data not shown). Interruption of ORF X at the *Bal*I site had no effect on subtilin production.

**Homology between SpaE and NisB.** The sequence data bases were searched for homologies to the deduced SpaE protein sequence. One very strong similarity, to NisB, was revealed (Fig. 3). We conclude that SpaE and NisB are likely to perform analogous roles in the biosynthesis of subtilin and nisin, respectively. It is notable that, despite this sequence homology, their locations in their respective operons are quite different, as illustrated in Fig. 2. This provides additional support for our earlier conclusion (2) that the *spa* and *nis* operons have been evolving separately for a long time. The roles of SpaE and NisB in lantibiotic synthesis are



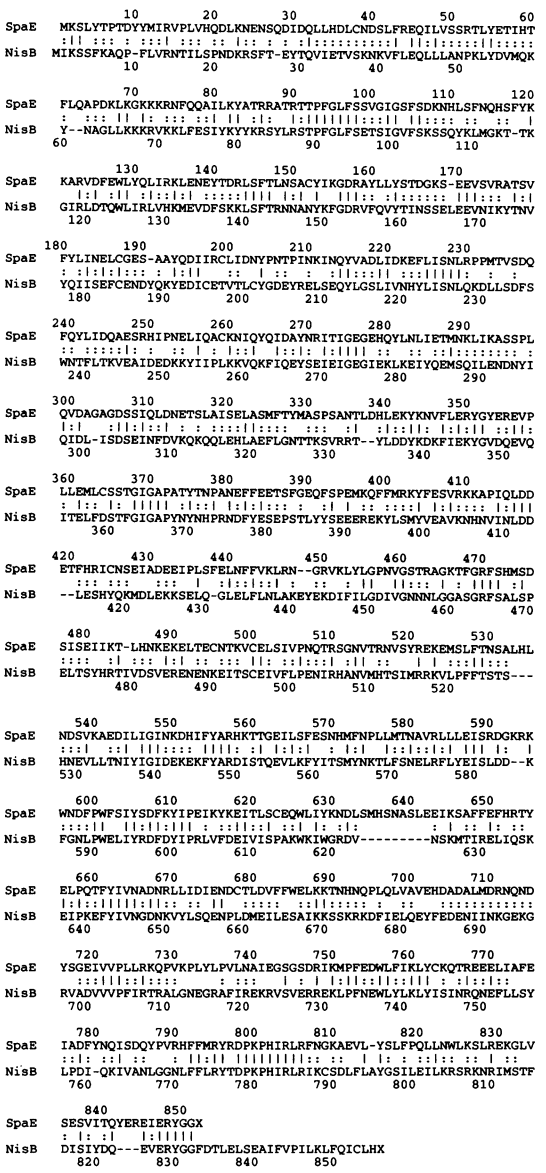


FIG. 3. Sequence homologies between SpaE and NisB. The deduced protein sequence encoded by *spaE* is compared with the corresponding sequence from the previously described ORF that is downstream from the nisin structural gene (*nisA*). The gene for this downstream ORF and the deduced protein sequence are referred to as *nisB* and NisB, respectively. The homology comparison was carried out with programs FASTA and GAP (4). Residues connected by lines are identical, and those connected by dots are similar. Dashes in the sequences represent gaps introduced to improve sequence homology. There are 29% identity and 53% similarity between the two proteins.

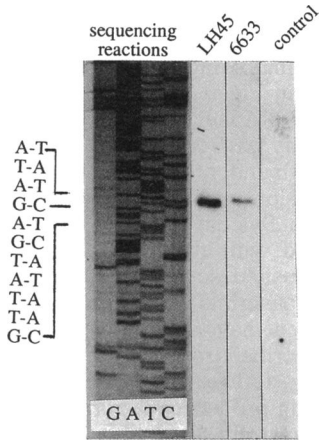



FIG. 4. Identification of promoter by primer extension. A 19-mer primer (5′-GATGTACTAAAGGAACCCG-3′) that is complementary to a sequence within the N-terminal region of the *spaE* mRNA was used in a primer extension reaction as described in Materials and Methods. The same primer was used in dideoxy sequencing reactions of the double-stranded *Eco*RI fragment that contained this region (Fig. 2). The sequencing reaction products and the primer extension reaction products were electrophoresed simultaneously on a denaturing sequencing gel and autoradiographed. The mRNA used as a template was isolated from strains ATCC 6633, the natural subtilin producer, and LH45, which is a subtilin-producing strain of *B. subtilis* 168 as described by Liu and Hansen (8). The mRNA was isolated from cells that had been incubated for 22 h, which is a time that these cells are known to produce subtilin mRNA (1). The autoradiogram shows the results for sequencing reaction products, the primer extension products from LH45 and ATCC 6633, and products of a control reaction which contained all reaction components except RNA. In the pair of sequences to the left of the autoradiogram, the right-hand sequence was read from the bracketed region of the gel and the left-hand sequence is the reverse complement that corresponds to the mRNA sequence shown in Fig. 1. The single sharp band that corresponds to a G residue in the reverse complement indicates that this residue is a transcription initiation site. The corresponding promoter sequence, including its −10 and −35 regions, is shown in Fig. 1.

unclear. The absence of sequence similarity to other (non-lantibiotic) proteins requires that functional tests be performed to elucidate the roles of SpaE and NisB. Neither protein has a recognizable export signal region or transmembrane helices, so they appear to be restricted to the cytoplasmic side of the membrane. Results presented above showed that interruption of the *spaE* gene destroys the ability of the cell to produce subtilin, although the experiment could not distinguish between the effect being the result of destroying the function of the SpaE protein and being the result of interference with transcription of other downstream *spa* genes.

**Identification of a promoter upstream from the *spaE* gene.** The presence of a ρ-independent terminator sequence in the

FIG. 2. (A) Organization of genes within the *spa* operon. Major restriction sites are shown. The promoter and its orientation are shown (P→). ORF X is an incomplete reading frame of unknown function. The sequence of *spaE* is given in Fig. 1. The sequences for *spaD*, *spaB*, *spaC*, and *spaS* have been published previously (2, 3). Horizontal arrows show the direction of transcription. (B) Organization of genes within the *nis* operon. The *nisA* gene encodes the primary structure of the nisin precursor peptide (1), and *nisB* is an ORF that is transcribed by read-through from *nisA* (11). ORF X is a partially sequenced ORF that has homology to a transposase (1). The terminator structure has characteristics of a ρ-independent terminator that probably defines the 3′ end of the *nis* operon (11). The double-headed shaded arrow indicates homologous *spaE* and *nisB* genes, shown in Fig. 3. The promoter of the *nis* operon (not shown) lies several kilobases upstream (11).

intergenic space upstream from *spaE* suggests that expression of the *spa* genes requires an active promoter between this putative terminator and the *spa* operon. A promoter-like sequence shown in Fig. 1 is a candidate. Primer extension analysis of the mRNA isolated from subtilin-producing cells was used to identify *spaE* transcripts. The cells were isolated after 22 h of growth under the conditions used previously (1), and primer extension experiments were performed with the Primer Extension System kit from Promega by using the protocol provided with the kit. Total cellular RNA was isolated as previously described (12). Figure 4 shows that the only detectable transcripts using a primer sequence from just inside the *spaE* gene have 5′ termini that are consistent with having been transcribed from this promoter. This 5′ terminus is identified in Fig. 1. The same result was obtained for both the natural subtilin producer, *B. subtilis* ATCC 6633, and a subtilin-producing strain that had been constructed (8) from *B. subtilis* 168 by transformation with a restriction fragment of the ATCC 6633 chromosome that contained the *spa* genes. The sequence of this promoter is similar to the vegetative $\sigma^A$ promoter of *B. subtilis* (10), except that the −10 and −35 regions are separated by 20 nucleotides instead of 17. This promoter is temporally regulated and is preferentially expressed during late growth stages (unpublished observations).

The results presented here complete the sequence of a 7-kb operon that contains five ORFs that are required for subtilin biosynthesis, although the number of proteins that are actually produced has not been determined. For example, it is possible that frameshifting may fuse SpaB and SpaC (3). Also, another laboratory has recently reported the sequence of the 3′ end of *spaE* (which is *spaB* by their nomenclature) and concluded, because of a discrepancy of a single nucleotide, that the SpaE and SpaD ORFs are fused (7). An argument against this fusion is that it would put a long C-terminal extension onto SpaE. The homologous NisB protein has no counterpart to this C-terminal extension. This discrepancy cannot be resolved with certainty until the expressed proteins are actually isolated and characterized. In any event, it is possible that the five ORFs in the *spa* operon could represent as few as three expressed proteins. One can legitimately wonder whether three proteins, or even five, are enough to carry out all the identifiable steps of the biosynthetic pathway of subtilin as well as provide immunity. If these are not sufficient, the rest of the genes must be in one or more additional operons. Experiments to determine whether the *spa* operon contains all of the genes required for subtilin biosynthesis are in progress.

**Nucleotide sequence accession number.** Nucleotide sequence accession no. M99263 has been assigned by the GenBank/EMBL Data Bank to the sequence in this article.

This work was supported by NIH grant AI24454, the National Dairy Promotion and Research Board, and Applied Microbiology, Inc., New York, N.Y.

We thank Tomas Kempe of the University of Maryland Protein-Nucleic Acid Laboratory for synthesis of oligonucleotides.

## REFERENCES

1. **Banerjee, S., and J. N. Hansen.** 1988. Structure and expression of a gene encoding the precursor of subtilin, a small protein antibiotic. J. Biol. Chem. **263:**9508–9514.
2. **Buchman, G. W., S. Banerjee, and J. N. Hansen.** 1988. Structure, expression, and evolution of a gene encoding the precursor of nisin, a small protein antibiotic. J. Biol. Chem. **263:**16260–16266.
3. **Chung, Y. J., M. T. Steen, and J. N. Hansen.** 1992. The subtilin gene of *Bacillus subtilis* ATCC 6633 is encoded in an operon that contains a homolog of the hemolysin B transport protein. J. Bacteriol. **174:**1417–1422.
4. **Devereux, J., P. Haeberli, and O. Smithies.** 1984. A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. **12:**387–395.
5. **Gross, E., H. H. Kiltz, and E. Nebelin.** 1973. Subtilin. Part 6: the structure of subtilin. Hoppe-Seyler's. Z. Physiol. Chem. **354:** 810–812.
6. **Gross, E., and J. L. Morell.** 1971. The structure of nisin. J. Am. Chem. Soc. **93:**4634–4635.
7. **Klein, C., C. Kaletta, N. Schnell, and K.-D. Entian.** 1992. Analysis of genes involved in biosynthesis of the lantibiotic subtilin. Appl. Environ. Microbiol. **58:**132–142.
8. **Liu, W., and J. N. Hansen.** 1991. Conversion of *Bacillus subtilis* 168 to a subtilin producer by competence transformation. J. Bacteriol. **173:**7387–7390.
9. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
10. **Moran, C. P., Jr.** 1989. Sigma factors and the regulation of transcription, p. 167–184. *In* I. Smith, R. A. Slepecky, and P. Setlow (ed.), Regulation of procaryotic development. American Society for Microbiology, Washington, D.C.
11. **Steen, M. T., Y. J. Chung, and J. N. Hansen.** 1991. Characterization of the nisin gene as part of a polycistronic operon in the chromosome of *Lactococcus lactis* 11454. Appl. Environ. Microbiol. **57:**1181–1188.
12. **Ulmanen, I., K. Lundström, 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.