JOURNAL OF BACTERIOLOGY, Feb. 1992, p. 1417–1422
0021-9193/92/041417-06$02.00/0

Vol. 174, No. 4

# The Subtilin Gene of *Bacillus subtilis* ATCC 6633 Is Encoded in an Operon That Contains a Homolog of the Hemolysin B Transport Protein

YONG JOON CHUNG, MARK T. STEEN, AND J. NORMAN HANSEN*

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

Received 17 June 1991/Accepted 9 December 1991

**Sequence analysis upstream from the subtilin structural gene (*spaS*) in *Bacillus subtilis* ATCC 6633 revealed several open reading frames, SpaB, SpaC, and SpaD. SpaB, consisting of 599 amino acid residues, shows excellent homology with a variety of membrane translocator proteins, such as HlyB from *Escherichia coli* and some mammalian multidrug resistance proteins. When the *spaB* gene was interrupted by integration of a chloramphenicol acetyltransferase gene, the ability of the cell to produce subtilin, as determined by a halo assay, was lost. The homology of SpaB to translocator proteins, including transmembrane and ATP-binding regions, suggests that SpaB may play a role in subtilin secretion. The SpaB open reading frame overlaps with another open reading frame called SpaC, and the possibility that the SpaB and SpaC proteins become fused by frameshifting is considered. Regions of homology between SpaD (177 residues) and HlyD were also found, suggesting that SpaD may participate with SpaB in translocation of subtilin through the membrane. Although no readily interpretable homologies to SpaC (442 residues) were found, its sequence suggests that it is membrane associated. The absence of ρ-independent transcription terminators between these open reading frames suggests that they are all part of the same operon.**

Subtilin is an antimicrobial peptide that is synthesized by a ribosomal mechanism in *Bacillus subtilis* ATCC 6633 (1). The mature 32-residue peptide contains many unusual amino acids, such as lanthionine and dehydroalanine, that are introduced by posttranslational processes (1). The specific biochemical events as well as the genes and proteins involved in the maturation of the subtilin precursor peptide are not known. The ability to produce subtilin has recently been transported (5) from the natural producer (*B. subtilis* ATCC 6633) to a strain that does not produce subtilin (*B. subtilis* 168). Because this transfer was achieved by competence transformation, it seems likely that the genes involved with subtilin maturation are located near the gene for the subtilin precursor peptide, and they may be in the same operon. The presence of a strong terminator immediately downstream from the subtilin gene (1, 5) appears to define the 3′ end of an operon, so we have studied the sequences upstream from the subtilin gene. In this paper, we present the sequences of several open reading frames (ORFs) upstream from the subtilin gene and show that interruption of one of them (SpaB) destroys the ability of the cell to secrete active subtilin.

**Sequences and organization of ORFs upstream from the subtilin structural gene (*spaS*).** The nucleotide sequence of the region upstream from the *spaS* gene, the deduced protein sequences, and putative ribosome binding sites are shown in Fig. 1. Major restriction sites and ORFs in this sequence are shown in Fig. 2.

**SpaB has extensive sequence homology to the HlyB transport protein from *Escherichia coli*.** The deduced protein sequences for the ORFs in the genes designated *spaB*, *spaC*, and *spaD* were used for homology searches of the GenBank data base with the TFASTA program in the Genetics Computer Group (University of Wisconsin) program package. The best homologies were found with the deduced SpaB protein sequence, which showed extensive homologies to a variety of membrane translocator proteins, such as the HlyB protein, which is involved with export of the hemolysin A protein (hemolysin toxin) in *E. coli* (6). The homologies between SpaB and HlyB are shown in Fig. 3. Included in the regions of homology are five transmembrane helices and an ATP-binding region. The HlyB protein has previously been shown to have homologies to many different membrane translocators, including human and other mammalian proteins involved with multidrug resistance (3). As would be expected, the SpaB protein also shows homologies to these proteins (data not shown).

***spaB* and *spaC* have overlapping reading frames.** The overlap shown for the SpaB and SpaC ORFs (Fig. 2) is unusual in prokaryotes. The overlapping region contains a sequence that could act as a ribosome binding site for *spaC*, but it also has characteristics of a frameshift sequence (see the legend to Fig. 2). Such a frameshift during translation would result in fusion of the SpaC protein to the SpaB protein. This could result in three distinct types of proteins, separate SpaB and SpaC proteins and a SpaB-SpaC fusion protein. We have not yet attempted to determine whether frameshifting occurs.

**Homologies between SpaD and HlyD.** Whereas the HlyB protein has been shown to function in binding and insertion of hemolysin into the membrane, translocation through the membrane requires a second protein called HlyD (6). The Spa ORFs were searched accordingly for homologies to HlyD. Two separate homologies between SpaD and HlyD were found (Fig. 3). One involves a 30-residue region located about 12 residues from the N terminus of SpaD and a region located about 15 residues from the C terminus of HlyD. The second homology was between a 45-residue region in the middle of SpaD and a region about 100 residues from the C

* Corresponding author.

```
CTAAGCCTTATGGAGGCTGCAGAACAGCTTTTCTGTGAAGACAGCAAAGTTGTTGAAATGATGATTAGAATGCACCGGATGAAAGATATTACGATAAGC 99
spaD--> M E A A E Q L F C E D S K V V E M M I E M H E M K D I T I S
AAGGAAATTGCAGGCATGGTTTCGGTTATACAGTTTTTAGAACAGTTCGAGCTAACGTTTGAAGAACAGTTAACTTTTTTAGAGAGAAATTCCTTACAG 198
K E I A G M V S V I Q F L E Q F E L T F E E Q L T F L E R N S L Q
AATGAGTATCGTACTGAATTTAAAAAGGATAGAGAAATGTATATTGAAATATGCAATTCTGACAGAGATTGGGATAATCTCAAGAAAACAAGTGATGGC 297
N E Y R T E F K K D R E M Y I E I C N S D R D W D N L K K T S D G
GGTATGTTATATGAAACTTTGAAAACAAGAAAAATGGCTGCAGCTCATTATGCATTTTTAATCAAAAAGGCATTTGATAACAAAGATGAAGTTTATTCA 396
G M L Y E T L K T R K M A A A H Y A F L I K K A F D N K D E V Y S
CGTATAGGAAGTATCATCCATCTGCATTGCAATCGTTTATTCGGAACCGACAGAGAACTGGAAAATAAAATTCTCACCCTATGCAGACATTCTTTATAT 495
R I G S I I H L H C N R L F G T D R E L E N K I L T L C R H S L Y
GCGCAACGATATCAAAAGATGAATGGTAGTTTAGCATGGAAGTAAAGGAACAACTGAAACTAAAAGAGCTGCTGTTTATCATGAAACAAATGCCTAAGA 594
A Q R Y Q K M N G S L A W K - RBS spaB--> M K Q M P K T
CGTTCAAGTTGATTTTCACCTTAGAAAGATCACTGTTTTTAAAGTTGATTGCATTCAGTATTATTACCGGTATTCTGCCAATTGTTTCACTATATATTT 693
F K L I F T L E R S L F L K L I A F S I I T G I L P I V S L Y I S
CACAAGAACTGATTAATTCCCTTGTGACTATCCGGAAAGATGTTTCAATTGTTATCACCATTTTCTTGACATATCTAGGGGTATCTTTTTTTTCGGAGC 792
Q E L I N S L V T I R K D V S I V I T I F L T Y L G V S F F S E L
TCATTTCGCAGATTTCCGAATTTTATAATGGTAAATTTCAATTAAATATTGGTTATAAACTCAACTATAAAGTCATGAAAAAGAGCAGTAATTTAGCTC 891
I S Q I S E F Y N G K F Q L N I G Y K L N Y K V M K K S S N L A L
TCAAGGATTTTGAAAATCCAGAAATATATGACAAGTTAGAGAGGGTAACGAAAGAGATCAGTTATAAGCCCTATCAGATTATTCAAGCCATCATTACAA 990
K D F E N P E I Y D K L E R V T K E I S Y K P Y Q I I Q A I I T M
TGACCACATCATTTGTGACGCTACTCTCATCAATTGCGTTTTTTAATGTCGTGGAACCCTAAAGTCTCACTTCTGTTGTTAGTTATTCCGGTAATTTCTC 1089
T T S F V T L L S S I A F L M S W N P K V S L L L L V I P V I S L
TATTCTATTTCTTGAAAATTGGACAGGAAGAATTCTTCATACACTGGAAACGCGCGGGAAAGGAAAGGAAATCTTGGTATATCAGCTATATACTCACAC 1188
F Y F L K I G Q E E F F I H W K R A G K E R K S W Y I S Y I L T H
ATGATTTTTCTTTTAAAGAATTAAAATTGTATAACCTTAAGGATTATTTATTAAACAAGTACTGGGATATCAAAAAATCATTTATAGAACAAGATACAA 1287
D F S F K E L K L Y N L K D Y L L N K Y W D I K K S F I E Q D T K
AGATATTAAGAAAGAAAACTTTATTGAACCTAATATATGAGATTGCAGTGCAATTAGTCGGAGCTGTCATTATATTTATTGCCATCATGTCTGCTTTTG 1386
I L R K K T L L N L I Y E I A V Q L V G A V I I F I A I M S A F A
CGGGAAAAATAATGGTCGGTAATGTAATGAGTTATATCAGATCGGTTTCGCTGGTGCAAAATCACTCACAATCTATCATGACAAGCATTTATTCGATCT 1485
G K I M V G N V M S Y I R S V S L V Q N H S Q S I M T S I Y S I Y
ATAACAGCAATCTCTATATGAATCAATTGTATGAATTTCTGGAATTAAAAGAAGAGAAAAGTCAAGGTCACAAGAAGCCGATTGTTGAGCCTATTCATT 1584
N S N L Y M N Q L Y E F L E L K E E K S Q G H K K P I V E P I H S
CTGTTGTTTTTCAAAATGTCAGCTTTATCTATCCCAATCAAGGAGAACAGACACTGAAACATATTAACGTTTCATTGCATAAAGGAGAACGTGTAGCCA 1683
V V F Q N V S F I Y P N Q G E Q T L K H I N V S L H K G E R V A I
TTGTTGGACCAAACGGATCAGGGAAAAGTACATTCATTAAGCTCCTTACCGGCTTATATGAAGTTCAGCAGGGCGACATTTTGATTAATGGAATAAATA 1782
V G P N G S G K S T F I K L L T G L Y E V Q Q G D I L I N G I N I
TTAAAGAACTGGATATGGACAGTTATATGAATCAAATTGCAGCGCTATTTCAAGACTTTATGAAATACGAAATGACATTAAAAGAAAATATAGGATTCG 1881
K E L D M D S Y M N Q I A A L F Q D F M K Y E M T L K E N I G F G
GCCAAATTGATAAACTACATCAAACAAACAAAATGCATGAAGTTCTCGATATTGTGAGAGCCGATTTCTTAAAAAGCCACTCTTCCTATCAATTCGATA 1980
Q I D K L H Q T N K M H E V L D I V R A D F L K S H S S Y Q F D T
CACAGCTCGGACTATGGTTTGATGAAGGAAGACAGCTATCCGGTGGACAATGGCAAAAGATTGCTTTGGCAAGGGCTTATTTTAGGGAAGCTTCTTTGT 2079
Q L G L W F D E G R Q L S G G Q W Q K I A L A R A Y F R E A S L Y
ATATTTTGGACGAGCCAAGTTCTGCGCTGGATCCAATTGCAGAAAAAGAAACCTTCGATACTTTTTTTCAGTTTGTCAAAAGATAAAATAGGAATATTTA 2178
I L D E P S S A L D P I A E K E T F D T F F S L S K D K I G I F I
TTTCTCATCGATTGGTTGCTGCTAAACTTGCAGATCGGATCATTGTTATGGATAAAGGAGAGATTGTAGGAATAGGCACACATGAAGAGCTATTAAAAA 2277
S H R L V A A K L A D R I I V M D K G E I V G I G T H E E L L K T
CATGTCCGCTGTATAAAAAAATGGATGAATCCGAGAACTATATGAACCCACTGGAGGAGGAAGGCAGTAAATGGAAAGAGGCACTGTATCAAGGATAGA 2376
C P L Y K K M D E S E N Y M N P L E E E G S K W K E A L Y Q G -
RBS spaC--> M E R G T V S R I E
AGTTGAAATTGTGAAAGAAATGGCTCGTCAAATTTCTAATTACGACAAAGTGCTTGAAATAGTGAATCAAAAAGATAACTTTCGCAGCATTGGCGAAGT 2475
V E I V K E M A R Q I S N Y D K V L E I V N Q K D N F R S I G E V
TCCGCTTATACCTTGGAAATCAACTGCTTTGAGTCATGGTATACCAGGCATTTGTATGCTGTATGGCGAATTGCACGCTCATTTTCCTGAGGAAGGATG 2574
P L I P W K S T A L S H G I P G I C M L Y G E L H A H F P E E G W
```

FIG. 1. Nucleotide sequence upstream from the subtilin structural gene. The sequence of DNA cloned from a bacteriophage λ library of *B. subtilis* ATCC 6633 DNA (1) was determined by the dideoxy chain termination method and by using the Sequenase II sequencing kit supplied by United States Biochemical Corp., Cleveland, Ohio. The structural gene for the subtilin precursor (*spaS*) has been previously reported (1). Genes are designated *spaB*, *spaC*, and *spaD* for the putative ORFs SpaB, SpaB, and SpaD, respectively. Putative ribosome binding sites (RBS) are indicated. The SpaD ORF is methionine rich in its N-terminal region, making identification of the preferred initiation codon difficult.

```
GGACGACATAGGACATCAATATTTATCGATTTTAGTAAATGAAATAAAAGAAAAAGGGCTGCATACTCCTTCAATGTTTTCTGGAGCAGCTGGCATTGG 2673
 D  D  I  G  H  Q  Y  L  S  I  L  V  N  E  I  K  E  K  G  L  H  T  P  S  M  F  S  G  A  A  G  I  G
ACTGGCAGCAATCTGTTTATCGCAACGTTTCACCTATTACAATGGTTTGATTTCTAGTATTAATGAATATTTGGCTGAAACCGTACCTCAATTGTTAAC 2772
 L  A  A  I  C  L  S  Q  R  F  T  Y  Y  N  G  L  I  S  S  I  N  E  Y  L  A  E  T  V  P  Q  L  L  T
TGAATTTGATCAGCGGCAAGTGTGTATGAGTGACTACGATGTCATAGAAGGAGTTAGCGGAATTGCCAACTATCTGCTGTTATTTCAAGAAGATAAAGC 2871
 E  F  D  Q  R  Q  V  C  M  S  D  Y  D  V  I  E  G  V  S  G  I  A  N  Y  L  L  L  F  Q  E  D  K  A
TATGGGAGATTTGTTAATTGATATTTTGAAATATCTCGTTAGGTTAACTGAAGACATTATCGTGGATGGAGAAAAAGTTCCTGGATGGCATATTCCATC 2970
 M  G  D  L  L  I  D  I  L  K  Y  L  V  R  L  T  E  D  I  I  V  D  G  E  K  V  P  G  W  H  I  P  S
TCAACATCAGTTTACTGATATTGAAAAAAAAAGCTTATCCTTACGGCAATTTTAATATGGGATTAGCACATGGGATACCTGGTCCTATTTGTGTTCTCTC 3069
 Q  H  Q  F  T  D  I  E  K  K  A  Y  P  Y  G  N  F  N  M  G  L  A  H  G  I  P  G  P  I  C  V  L  S
TTCAGCACTCATACAGGGAATTAAAGTAAAGGGACAAGAACGTGCGATTGAAAAAATGGCGAATTTCCTACTGGAATTTTCTGAAAAAGAGCAAGACAG 3168
 S  A  L  I  Q  G  I  K  V  K  G  Q  E  R  A  I  E  K  M  A  N  F  L  L  E  F  S  E  K  E  Q  D  S
CTTGTTTTGGAAAGGAATCATAAGCTTTGAGGAGTATCAATACGGATCTCCGCCAAACGCAGTAAATTTTAGCAGAGATGCTTGGTGTTATGGAAGACC 3267
 L  F  W  K  G  I  I  S  F  E  E  Y  Q  Y  G  S  P  P  N  A  V  N  F  S  R  D  A  W  C  Y  G  R  P
TGGTGTATGTCTAGCGCTTGTTAAAGCCGGAAAAGCACTTCAAAATACTGAACTTATAAATATCGGAGTACAGAATTTAAGATACACAATATCTGATAT 3366
 G  V  C  L  A  L  V  K  A  G  K  A  L  Q  N  T  E  L  I  N  I  G  V  Q  N  L  R  Y  T  I  S  D  I
ACGGGGAATTTTTTCACCTACCATATGTCATGGTTACAGCGGTATCGGTCAGATTCTGCTTGCTGTAAATCTACTCACCGGACAGGAGTATTTTAAGGA 3465
 R  G  I  F  S  P  T  I  C  H  G  Y  S  G  I  G  Q  I  L  L  A  V  N  L  L  T  G  Q  E  Y  F  K  E
AGAGCTTCAAGAGATTAAACAAAAGATCATGAGCTACTATGACAAGGATTATATCTTTGGATTCCATAACTATGAATCAATGGAAGGGGACGAAGCAGT 3564
 E  L  Q  E  I  K  Q  K  I  M  S  Y  Y  D  K  D  Y  I  F  G  F  H  N  Y  E  S  M  E  G  D  E  A  V
ACCTTTGCAGTACGTTGGTTTGTTGGATGGAGCTGTAGGTGTAGGCTTAGGGGTATTAAACATGGAATTAGGCTCAAAAACAGATTGGACAAAAGCATT 3663
 P  L  Q  Y  V  G  L  L  D  G  A  V  G  V  G  L  G  V  L  N  M  E  L  G  S  K  T  D  W  T  K  A  L
ATTAATTTAATAAAAAAAGGAAAAAAATGATAAAATCTTGATATTTGTCTGTTACTATTTAGGTATTGAAAGGAGGTCACCAATATGTCAAAGTTCGAT 3762
 L  I  -                                                    RBS spaS --> M  S  K  F  D
GATTTCGATTTGGATGTTGTGAAAGTCTCTAAACAAGACTCAAAAATCACTCCGCAATGGAAAAGTGAATCACTTTGTACACCAGGATGTGTAACTGGT 3861
 D  F  D  L  D  V  V  K  V  S  K  Q  D  S  K  I  T  P  Q  W  K  S  E  S  L  C  T  P  G  C  V  T  G
GCATTGCAAACTTGCTTCCTTCAAACACTAACTTGTAACTGCAAAATCTCTAAATAAGTAAAACCATTAGCATCACCTTGCTCTGACTCCTTGCACTTC 3960
 A  L  Q  T  C  F  L  Q  T  L  T  C  N  C  K  I  S  K  -
TGAGTGTTATACATACTTATTTTCATAGAGTCGGGACAAGAAAATGAAGTAAAAAACGACGGGTGTGAAAGAGTTTATATTCACACCCGTTTTTATATT 4059
CGGCTTTAAGGAGGAACACAATTGTAGAACGGAAGAACGGTTATTTTCGATCATGCGTTTTGAATAACATTCCAATAAAAATTCCAGTCTCTTCCTCAA 4158
ATGCAGACAAAGGATGAAGGACTTAAGGGTACTTACCAGGTTTTATGGTTAAGAATATTTCTAAGAACATCATATTTTTTATTAGGAAATTAATAAATG 4257
AGATTGATCACTCTAGA
```

FIG. 1—*Continued.*

terminus of HlyD. Although the regions of homology are not large in comparison to the homologies between SpaB and HlyB, they are in nearly perfect alignment. Moreover, the SpaD sequence is so short (177 residues) that there is relatively little opportunity for random-chance homologies, and the fact that the homologies encompass nearly one-half of the SpaD ORF increases confidence in their significance. It is intriguing that one homology is between the N terminus of SpaD and the C terminus of HlyD, which indicates a reordering of the domains. It has been shown that the cytoplasmic hemolysin protein contains its export recognition signal at its C terminus and that the hemolysin protein is transported without modification (6). In contrast, the subtilin precursor contains an N-terminal leader that is cleaved during the maturation process (1). This leader does not contain a typical export signal (1), suggesting export by an alternate mechanism. If the SpaB protein participates in the translocation of subtilin, one would also expect to find a protein that performs the role that corresponds to that of HlyD, namely, the translocation that follows binding and insertion of the precursor peptide into the membrane. If SpaD performs this role, the fact that SpaS has an N-terminal leader that undergoes cleavage, whereas hemolysin is translocated without cleavage, suggests that its mode of interaction with the translocated peptide may be quite dif-

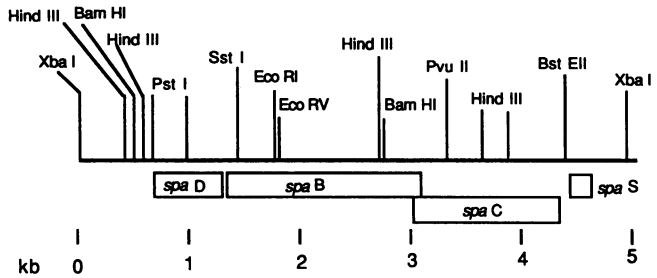
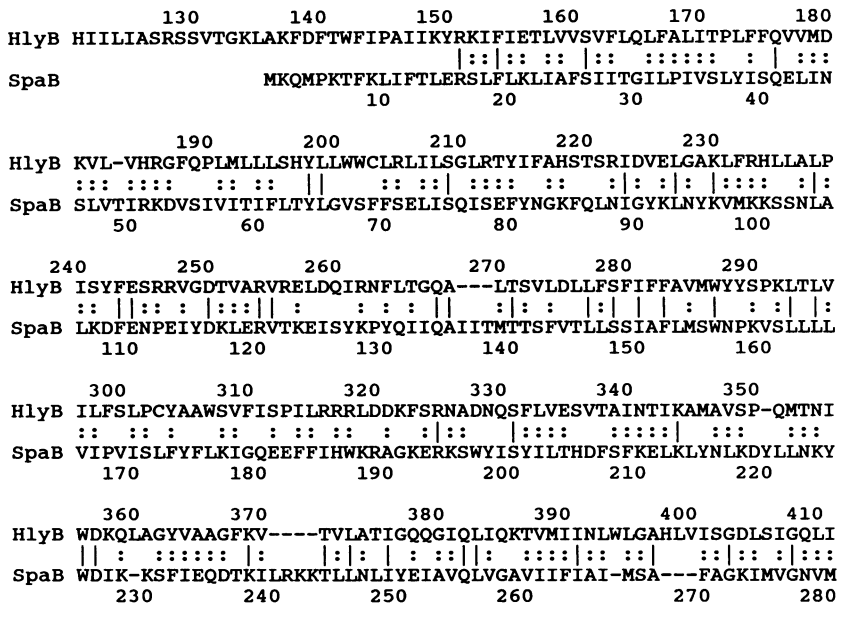
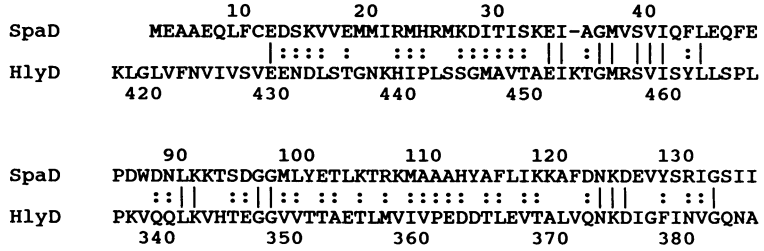



FIG. 2. Organization of ORFs within the subtilin (*spa*) operon. The restriction sites and ORFs in the sequence (Fig. 1) within a 5-kb natural *Xba*I fragment that contains the subtilin gene are indicated. The deduced protein sequence of *spaB* contains five transmembrane helices involving amino acids 17 to 39, 56 to 74, 135 to 154, 160 to 176, and 248 to 280 (the sequence of the deduced SpaB protein is shown in Fig. 3). SpaB is clearly a membrane-associated protein. There is also an ATP-binding domain at amino acid positions 355 to 545. SpaB also shows strong homology to HlyB in the hemolysin operon, and SpaD shows homologies to HlyD (Fig. 3). The overlapping region between the ORF of *spaB* and the ORF of *spaC* contains a sequence (TGGAGGAGGAAGGCAGTAAATG GAAAGA) that is typical of a prokaryotic frameshift sequence (4, 9, 10). It contains both a potential ribosome binding site (GGAGGA) and other suitable components of a frameshift region (AGGCAG TAAATG). The SpaB and SpaC proteins could accordingly be expressed as separate translation products or as a SpaB-SpaC fusion protein.

Comparison of HlyB with SpaB




Comparison of SpaD with HlyD



ferent from that of HlyD. One would accordingly expect that SpaD and HlyD, even if they perform homologous functions, might possess limited sequence homology. Further experiments to confirm the role of SpaD are in progress.

**Interruption of the *spaB* gene interferes with subtilin production, as determined by a halo assay.** The location of the *spaB* gene a short distance upstream from the subtilin structural gene (*spaS*) raises the possibilities that both genes are in the same operon and that the SpaB protein is involved in subtilin biosynthesis. One way to establish involvement of the *spaB* gene is to inactivate it and observe the effect on subtilin production. Inactivation of the *spaB* gene was achieved by integrating a *cat* gene into the structural region of the *spaB* gene. The strategy to carry out the integration is described in the legend to Fig. 4. The effect of the integration of the *cat* gene on subtilin production was tested by a halo assay. Strain CH5, in which the *spaB* gene has been interrupted, no longer forms the halo that is observed around the parent strain (LH45Δc) (data not shown). We conclude that the loss of halo formation is a consequence of interrupting the *spaB* gene.

There are several ways to explain the loss of halo formation. One possibility is that the SpaB protein is required for some stage of subtilin maturation or secretion and interruption of the *spaB* gene resulted in production of a defective SpaB protein. Another possibility is that there was a polar effect on the transcription of one or more genes that are downstream from the *spaB* gene, including the *spaS* gene, as would be expected if downstream genes were part of the same operon. A third possibility is that the region is somehow involved with regulation of subtilin biosynthesis. Although the exact mechanism of interference with subtilin production is uncertain, the result does establish that the region encompassing the *spaB* gene is involved with subtilin biosynthesis. We note that all the ORFs in this sequence are quite close together, some even overlapping, and that there are no obvious ρ-independent terminator sequences (containing a GC-rich stem-loop followed by a stretch of T residues) between any of them. The absence of such terminators is further evidence in favor of these ORFs all being part of the same operon.

The posttranslational events in the conversion of the subtilin precursor peptide to mature subtilin include dehydration of serines and threonines to form dehydroalanines and dehydrobutyrines and addition of cysteine sulfhydryl groups to some of the dehydro residues to give lanthionine and β-methyllanthionine (1). The leader region of the precursor is cleaved, and mature subtilin is secreted. The sequential order of these events is not known. The leader region is unusual and lacks the features that are characteristic of a typical prokaryotic export signal (1, 2). It is therefore probable that the mechanism of subtilin secretion uses a secretion pathway that is distinct from the pathway

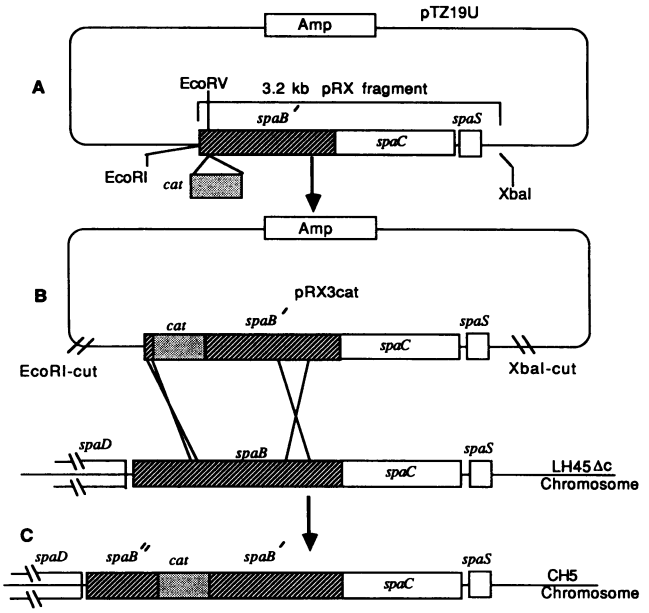



FIG. 4. Strategy for interruption of the *spaB* gene by integration of a *cat* gene. The *cat* gene was derived from plasmid pC194 by using the polymerase chain reaction. (A) The *cat* gene was cloned into the *Eco*RV site located at codon 227 of SpaB, which is about one-third of the way into the gene. The spaB″ region is the portion of the gene that extends upstream from the *Eco*RV site, and spaB′ is the portion that extends downstream from the *Eco*RV site. The *Eco*RV site is within a 3.2-kb natural *Eco*RI-*Xba*I fragment cloned in pTZ19U from the chromosome of *B. subtilis* ATCC 6633 (1). (B) This *cat* gene derivative was linearized by cutting with *Eco*RI and *Xba*I and transformed into competent LH45Δc cells, and transformants were selected on chloramphenicol. LH45Δc (kindly provided by Wei Liu) had been derived from subtilin-producing LH45 (5) by spontaneous deletion of the *cat* gene and contains approximately 40 kb of the ATCC 6633 chromosome, including the region containing the *Eco*RI-*Xba*I fragment (5). LH45Δc can become chloramphenicol resistant by the double-crossover shown in panel B, which results in replacement of the normal *spaB* gene with the one that has been interrupted by the *cat* gene, resulting in strain CH5, which contains the interrupted *spaB* gene in its chromosome, shown in panel C.

mediated by the Sec proteins (7). The SpaB protein, with its homology to other known membrane translocator proteins, is the first indication that the subtilin gene complex contains genes involved with export. The discovery of regions of homology between SpaD and HlyD further supports this idea. Experiments to further characterize these proteins and to search for additional genes for maturation and export of subtilin are in progress.

**Nucleotide sequence accession number.** Nucleotide se-

FIG. 3. Homology between the deduced protein sequences of ORFs in the subtilin operon and those of ORFs in the hemolysin operon. A region of homology between the sequence of HlyB, which is a transport protein encoded in the hemolysin (*hly*) operon of *E. coli* (3, 6) that mediates the export of the hemolysin protein, and the deduced sequence of SpaB in the subtilin (*spa*) operon (cf. Fig. 1) is shown. The numbers refer to amino acid residues in the respective proteins. The SpaB protein is 599 residues long. There is 23% absolute homology (indicated by vertical lines) within the 564-amino-acid region of overlap shown, in addition to many conserved amino acid homologies (indicated by colons). Homologies include both cytoplasmic and membrane-associated domains as well as the ATP-binding domain (cf. the legend to Fig. 2). The HlyB protein has an N-terminal end that is absent from the SpaB protein, but SpaB does contain a homology to the complete C-terminal end of the HlyB protein. Homologies between SpaD and HlyD were also found. HlyD contains 479 amino acids, and the C-terminal region of HlyD contains a homology to the N terminus of SpaD, which contains 177 amino acids. A second homology between the middle of SpaD and a region about 100 residues from the C terminus of HlyD is shown. No clearly identifying homologies to the SpaC protein were found in the protein data bases, although the sequence of SpaC suggests that it is membrane associated. The nucleotide sequences from which HlyB and HlyD were derived were obtained from sequence Gb_ba:ECOHLY.

quence accession number M83944 has been assigned by the GenBank/EMBL Data Bank to the sequence in this article. The sequence Gb_ba:ECOHLY has been assigned the nucleotide sequence accession number M10133 by GenBank/EMBL.

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

## 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. **Holland, I. B., B. Kenny, and M. Blight.** 1990. Haemolysin secretion from E. coli. Biochimie **72**:131–141.
4. **Jacks, T., H. D. Madhani, F. R. Masiarz, and H. E. Varmus.** 1988. Signals for ribosomal frameshifting in the Rous sarcoma virus gag-pol region. Cell **55**:447–458.
5. **Liu, W., and J. N. Hansen.** 1991. Conversion of *Bacillus subtilis* 168 to a subtilin producer by competence transformation. J. Bacteriol. **173**:7387–7390.
6. **Oropeza-Wekerle, R. L., W. Speth, B. Imhof, I. Gentschev, and W. Goebel.** 1990. Translocation and compartmentalization of *Escherichia coli* hemolysin (HlyA). J. Bacteriol. **172**:3711–3717.
7. **Randall, L. L., and S. J. S. Hardy.** 1989. Unity in function in the absence of consensus in sequence: role of leader peptides in export. Science **243**:1156–1159.
8. **Steen, M., 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* ATCC 11454. Appl. Environ. Microbiol. **57**:1181–1188.
9. **Weiss, R. B., D. M. Dunn, J. F. Atkins, and R. F. Gestland.** 1987. Slippery runs, shifty stops, backward steps, and forward hops: −2, −1, +1, +2, +5, and +6 ribosomal frameshifting, p. 687–693. *In* Cold Spring Harbor Symposia on Quantitative Biology: Evolution of Catalytic Function. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
10. **Weiss, R. B., D. M. Dunn, A. E. Dahlberg, J. F. Atkins, and R. F. Gestland.** 1988. Reading frame switch caused by base-pair formation between the 3′ end of 16S rRNA and the mRNA during elongation of protein synthesis in Escherichia coli. EMBO J. **7**:1503–1507.