JOURNAL OF BACTERIOLOGY, Sept. 1985, p. 824–831
0021-9193/85/090824-08$02.00/0

Vol. 163, No. 3

Nucleotide Sequence and Promoter Region for the Neutral Protease Gene from *Bacillus stearothermophilus*

MASAHIRO TAKAGI, TADAYUKI IMANAKA,* AND SHUICHI AIBA

Department of Fermentation Technology, Faculty of Engineering, Osaka University, Yamada-oka, Suita-shi, Osaka 565, Japan

Received 3 April 1985/Accepted 29 May 1985

The thermostable neutral protease gene *nprT* of *Bacillus stearothermophilus* was sequenced. The DNA sequence revealed only one large open reading frame, composed of 1,644 bases and 548 amino acid residues. A Shine-Dalgarno sequence was found 9 bases upstream from the translation start site (ATG), and the deduced amino acid sequence contained a signal sequence in its amino-terminal region. The sequence of the first 14 amino acids of purified extracellular protease completely matched that deduced from the DNA sequence starting at GTC (Val), 687 bases (229 amino acids) downstream from ATG. This suggests that the protease is translated as a longer polypeptide. The amino acid sequence of the extracellular form of this protease (319 amino acids) was highly homologous to that of the thermostable neutral protease from *Bacillus thermoproteolyticus* but less homologous to the thermolabile neutral protease from *Bacillus subtilis*. A promoter region determined by S1 nuclease mapping (TTTTCC for the −35 region and TATTTT for the −10 region) was different from the conserved promoter sequences recognized by the known σ factors in bacilli. However, it was very homologous to the promoter sequence of the *spo0B* gene from *B. subtilis*. The guanine-plus-cytosine content of the coding region of the *nprT* gene was 58 mol%, while that of the third letter of the codons was much higher (72 mol%).

Microbial proteases have been divided into four groups based on their mechanisms of action. These groups are serine, metal, thiol, and acid proteases.

The extracellular proteases of the genus *Bacillus* are mainly either serine (alkaline) or metal (neutral) enzymes (25). Neutral and alkaline proteases from *Bacillus subtilis* and *Bacillus amyloliquefaciens* have been cloned and sequenced (26, 30, 33, 36, 37). These studies revealed the existence of a pre-"pro" structure. However, the promoter region and the regulation of gene expression in these proteases are not necessarily well characterized. In contrast to these thermolabile proteases, thermostable neutral protease from *Bacillus thermoproteolyticus* (thermolysin) is well characterized enzymatically. The primary and three-dimensional structures and the active site of the thermolysin have been determined (12, 17, 29).

We have already cloned the structural gene of thermostable neutral protease from *Bacillus stearothermophilus* and examined its expression in *B. stearothermophilus* and *B. subtilis* (4). This paper describes both the nucleotide sequence of the cloned neutral protease gene and determination of the promoter region. Furthermore, we compare the amino acid sequence and the guanine-plus-cytosine content of the neutral protease with those same components from other *Bacillus* proteases.

MATERIALS AND METHODS

Bacterial strains, plasmids, phages, and culture conditions. The *Bacillus* strains used were *B. stearothermophilus* CU21 (Str^r) (9) and *B. subtilis* MT-2 (*trpC2 leuC7 hsdR hsdM* Npr^- (4). *Escherichia coli* K-12 JM105 [Δ(*lac-pro*) *thi rpsL endA sbcB15 hspR4*/F′ *traD36 proA*$^+$*B*$^+$*lacI*^{q} *ZΔM15*](19) was used as a host for phages M13mp10 and M13mp11 (19). The plasmids used were pNP22 (Km^r Tc^r *nprT*$^+$ [structural gene of neutral protease from *B. stearothermophilus*]) and pNP28 (Km^r Tc^r *nprT*$^+$) (4). Bacteria were grown in L broth or YPC medium or on L or LC agar (4). The antibiotics used were kanamycin and tetracycline (5 μg of each per ml) for both *B. stearothermophilus* and *B. subtilis*.

Plasmid isolation, restriction enzyme treatment, ligation of DNA, and transformation. Plasmid DNA was prepared by either the rapid alkaline extraction method or CsCl-ethidium bromide equilibrium density gradient centrifugation as described previously (9). Treatment of DNA with restriction enzymes and ligation of DNA with T4 DNA ligase were done as recommended by the manufacturer. Transformation of *B. stearothermophilus* protoplasts with plasmid DNA was done as described previously (9). For transformation of *B. subtilis* with plasmid DNA, competent cells were prepared as described previously (8).

Gel electrophoresis for DNA analysis and isolation. For the analysis of DNA, gel electrophoresis with agarose or polyacrylamide was carried out under standard conditions (15). Recovery of DNA from either low-melting-point agarose or polyacrylamide gels was performed by the standard method (15).

DNA sequencing analysis. DNA sequencing was performed by either the method of Maxam and Gilbert (18) or the dideoxy method with *E. coli* K-12 JM105 and M13 phages (19). Both strands were sequenced.

Detection of protease-producing colonies and assay of protease activity. Protease-producing colonies on plates were detected, and protease activity was assayed for casein hydrolytic activity as described previously (4). We defined 1 U of protease as the quantity required to increase the absorbance at 275 nm by an equivalent of 1 μg of tyrosine per min at 37°C.

* Corresponding author.

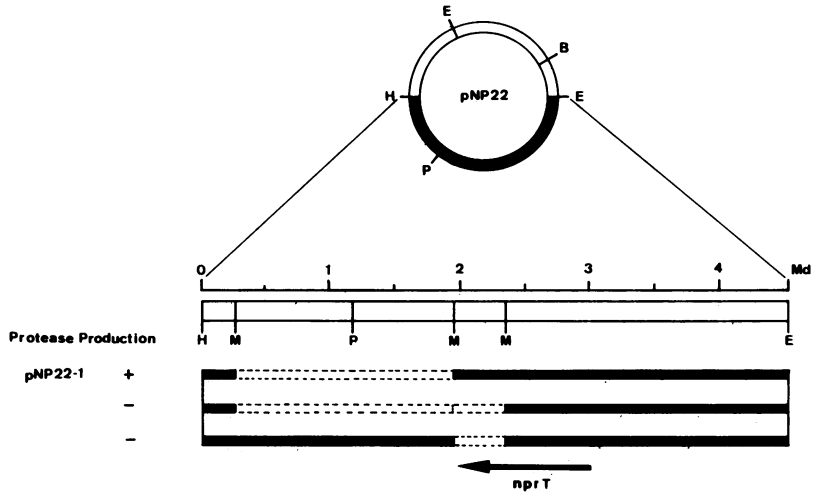


FIG. 1. Structure of plasmid pNP22 and its derivatives. Black bars indicate the DNA from *B. stearothermophilus* CU21. The neutral protease gene is contained in a black bar of the 4.5-MDa *Hin*dIII-*Eco*RI fragment of pNP22 (4). Broken parallel lines indicate the deletion of *Mlu*I fragments. Cleavage sites of *Bgl*II, *Eco*RI, *Hin*dIII, *Mlu*I, and *Pst*I are indicated by B, E, H, M, and P, respectively. The position and direction of the *nprT* gene are indicated by the arrow.

Purification of extracellular protease. Extracellular protease was purified by the method described previously (4). After the column chromatography of DEAE-Sephadex A50 as a final purification step, the enzyme fractions were collected and immediately inactivated to minimize self-digestion by adding trichloroacetic acid to a final concentration of 2%. The protein was precipitated by centrifugation (15,000 × *g*, 30 min) and dissolved in 0.8 ml of 98% formic acid solution. Desalting and further purification were performed by reverse-phase column chromatography with a high-performance liquid chromatography system from Gilson Medical Electronics, Villiers le Bell, France. The peak fractions of absorbance at 280 nm were collected and then freeze-dried. This sample was used for amino acid analysis.

Polyacrylamide gel electrophoresis for protein analysis. Sodium dodecyl sulfate-polyacrylamide gel electrophoresis was performed as described previously (16).

Amino-terminal amino acid sequence. Amino-terminal amino acids were analyzed by Edman degradation, by the method of Hunkapiller et al. (7) with a gas-phase sequencer (model 470A; Applied Biosystem Inc., Foster, Calif.). The phenylthiohydantoin derivatives of protein amino acids were analyzed by liquid chromatography (SP 8100; Spectra Physics, San Jose, Calif.).

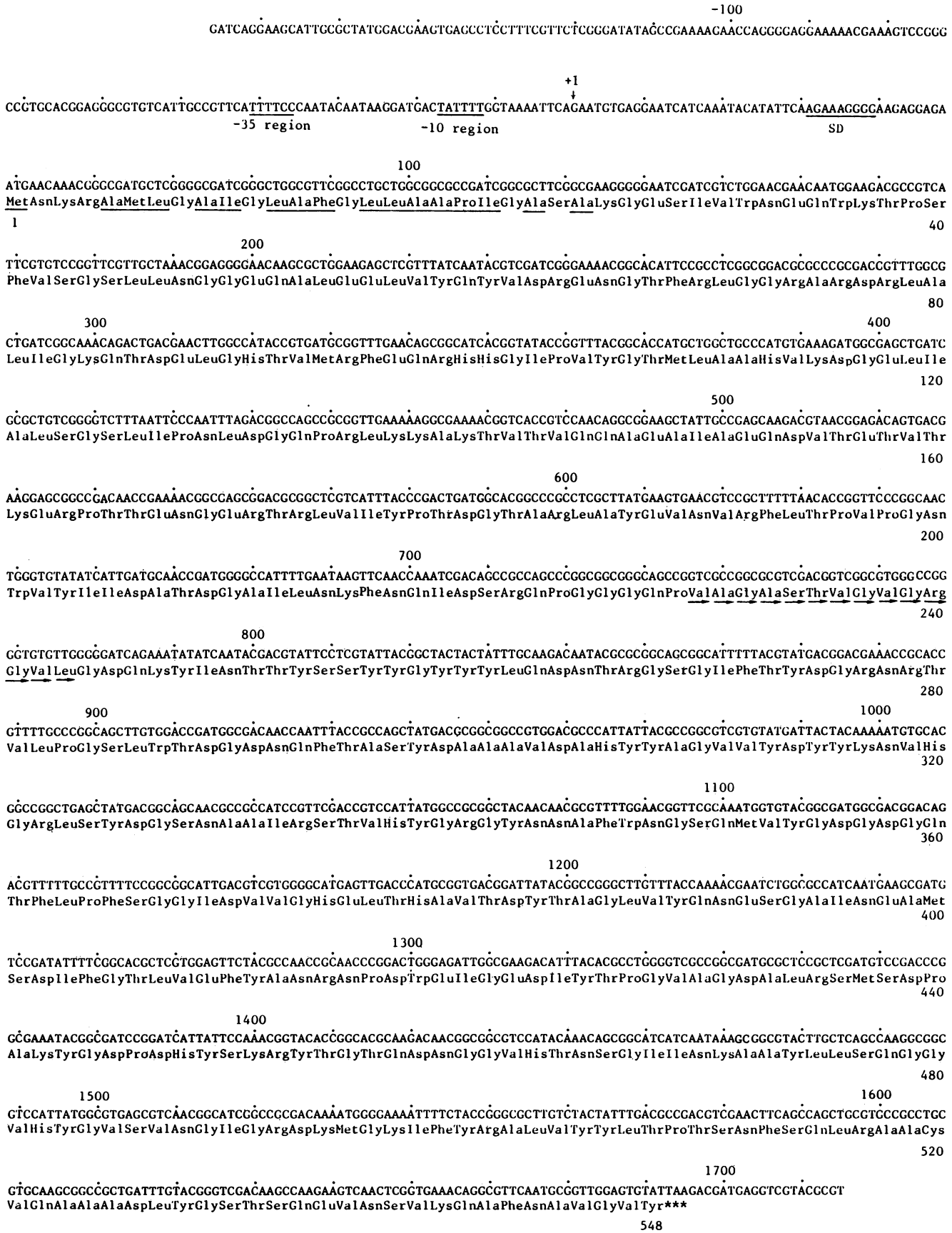
Preparation of RNA fraction. An RNA fraction containing *nprT* mRNA was extracted from *B. stearothermophilus* CU21 cells harboring recombinant plasmid pNP28 (Km^r Tc^r $nprT^+$), by the method of Duvall et al. (2) with slight modifications. The bacteria were grown in L broth containing kanamycin (5 μg/ml) and calcium chloride (2 mM) at 48°C for 18 h. The cells were harvested by centrifugation (8,000 × *g*, 10 min) and then washed three times with cold 0.01 M Tris hydrochloride buffer (pH 8.0). Cell lysis was brought about in the presence of lysozyme (200 μg/ml) at 0°C for 10 min, and the lysate was treated twice with phenol and once with chloroform. Nucleic acids were collected by ethanol precipitation and digested with RNase-free DNase I (30 μg/ml) at 37°C for 30 min. The RNA fraction was then extracted with phenol and with chloroform, followed by precipitation with cold ethanol.

Preparation of DNA probe. The 241-base-pair (bp) *Mbo*I fragment, which contains 26 bp of coding region for the *nprT* gene and its 5′ flanking region, was treated with bacterial alkaline phosphatase, and both 5′ ends were labeled in the presence of [γ-^{32}P]ATP with T4 polynucleotide kinase. The labeled DNA fragment was treated with *Hha*I, and the resulting 223-bp fragment was separated by polyacrylamide gel electrophoresis and isolated from the gel. This DNA fragment, which was labeled at one 5′ end (*Mbo*I site), was used as a hybridization probe of S1 nuclease mapping.

S1 nuclease mapping. S1 nuclease mapping was performed by the method of Gilman and Chamberlin (5) with some modifications. The radioactive DNA probe fragment (about 100,000 cpm) and the RNA fraction (about 200 μg) were mixed, precipitated with ethanol, and dissolved in 30 μl of hybridization buffer (80% formamide, 40 mM PIPES [piperazine-*N*,*N*′-bis(2-ethanesulfonic acid)], 1 mM EDTA, 0.4 M NaCl [pH 6.4]). The mixture was incubated at 72°C for 15 min and then cooled immediately down to 59°C. This temperature for hybridization was altered based on the guanine-plus-cytosine content of the DNA probe. The incubation was continued for 18 h. After the addition of 300 μl of ice-cold S1 nuclease buffer (0.28 M NaCl, 0.05 M sodium acetate (pH 4.6), 4.5 mM $ZnSO_4$, 20 μg of carrier single-stranded DNA per ml), the reaction mixture was treated with S1 nuclease (400 U/ml) at 37°C for 30 min. The reaction product was precipitated with ethanol, dissolved in sequencing dye buffer, and electrophoresed in an 8% polyacrylamide sequencing gel.

Enzymes and chemicals. Restriction endonucleases, T4 DNA ligase, polynucleotide kinase, bacterial alkaline phosphatase, an M13 sequencing kit, RNase-free DNase I, S1 nuclease, and low-melting-point agarose were purchased from Takara Shuzo Co., Kyoto, Japan. Kanamycin and tetracycline were from Sigma Chemical Co., St. Louis, Mo. [γ-^{32}P]ATP was purchased from New England Nuclear Corp., Boston, Mass., whereas [α-^{32}P]dCTP was from Amersham International plc., Buckinghamshire, England. All other chemicals used were from Wako Pure Chemical Industries, Osaka, Japan.

Analysis of primary-structure homology of protein. Amino


-100
GATCAGGAAGCATTGCGCTATGCACGAAGTGAGCCTCCTTTCGTTCTCGGGATATAGCCGAAAACAACCACGGGAGGAAAAACGAAAGTCCGGG
+1
CCGTGCACGGAGGGCGTGTCATTGCCGTTCATTTTCCCAATACAATAAGGATGACTATTTTGGTAAAATTCAGAATGTGAGGAATCATCAAATACATATTCAAGAAAGGGCAAGACGAGA
-35 region
-10 region
SD
100
ATGAACAAACGGGCGATGCTCGGGGCGATCGGGCTGGCGTTCGGCCTGCTGGCGGCGCCGATCGGCGCTTCGGCGAAGGGGGAATCGATCGTCTGGAACGAACAATGGAAGACGCCGTCA
MetAsnLysArgAlaMetLeuGlyAlaIleGlyLeuAlaPheGlyLeuLeuAlaAlaProIleGlyAlaSerAlaLysGlyGluSerIleValTrpAsnGluGlnTrpLysThrProSer
1
40
200
TTCGTGTCCGGTTCGTTGCTAAACGGAGGGGAACAAGCGCTGGAAGAGCTCGTTTATCAATACGTCGATCGGGAAAACGGCACATTCCGCCTCGGCGGACGCGCCCGCGACCGTTTGGCG
PheValSerGlySerLeuLeuAsnGlyGlyGluGlnAlaLeuGluGluLeuValTyrGlnTyrValAspArgGluAsnGlyThrPheArgLeuGlyGlyArgAlaArgAspArgLeuAla
80
300
400
CTGATCGGCAAACAGACTGACGAACTTGGCCATACCGTGATGCGGTTTGAACAGCGGCATCACGGTATACCGGTTTACGGCACCATGCTGGCTGCCCATGTGAAAGATGGCGAGCTGATC
LeuIleGlyLysGlnThrAspGluLeuGlyHisThrValMetArgPheGluGlnArgHisHisGlyIleProValTyrGlyThrMetLeuAlaAlaHisValLysAspGlyGluLeuIle
120
500
GCGCTGTCGGGGTCTTTAATTCCCAATTTAGACGGCCAGCCGCGGTTGAAAAAGGCGAAAACGGTCACCGTCCAACAGGCGGAAGCTATTGCCGAGCAAGACGTAACGGAGACAGTGACG
AlaLeuSerGlySerLeuIleProAsnLeuAspGlyGlnProArgLeuLysLysAlaLysThrValThrValGlnGlnAlaGluAlaIleAlaGluGlnAspValThrGluThrValThr
160
600
AAGGAGCGGCCGACAACCGAAAACGGCCAGCGGACGCGGCTCGTCATTTACCCGACTGATGGCACGGCCCGCCTCGCTTATGAAGTGAACGTCCGCTTTTTAACACCGGTTCCCGGCAAC
LysGluArgProThrThrGluAsnGlyGluArgThrArgLeuValIleTyrProThrAspGlyThrAlaArgLeuAlaTyrGluValAsnValArgPheLeuThrProValProGlyAsn
200
700
TGGGTGTATATCATTGATGCAACCGATGGGGCCATTTTGAATAAGTTCAACCAAATCGACAGCCGCCAGCCCGGCGGCGGGCAGCCGGTCGCCCGGCGCGTCGACGGTCGGCGTGGGCCGG
TrpValTyrIleIleAspAlaThrAspGlyAlaIleLeuAsnLysPheAsnGlnIleAspSerArgGlnProGlyGlyGlyGlnProValAlaGlyAlaSerThrValGlyValGlyArg
240
800
GGTGTGTTGGGGGATCAGAAATATATCAATACGACGTATTCCTCGTATTACGGCTACTACTATTTGCAAGACAATACGCGCGGCAGCGGCATTTTTACGTATGACGGACGAAACCGCACC
GlyValLeuGlyAspGlnLysTyrIleAsnThrThrTyrSerSerTyrTyrGlyTyrTyrTyrLeuGlnAspAsnThrArgGlySerGlyIlePheThrTyrAspGlyArgAsnArgThr
280
900
1000
GTTTTGCCCGGCAGCTTGTGGACCGATGGCGACAACCAATTTACCGCCAGCTATGACGCGGCGGCCGTGGACGCCCATTATTACGCCGGCGTCGTGTATGATTACTACAAAAATGTGCAC
ValLeuProGlySerLeuTrpThrAspGlyAspAsnGlnPheThrAlaSerTyrAspAlaAlaAlaValAspAlaHisTyrTyrAlaGlyValValTyrAspTyrTyrLysAsnValHis
320
1100
GGCCGGCTGAGCTATGACGGCAGCAACGCCGCCATCCGTTCGACCGTCCATTATGGCCGCGGCTACAACAACGCGTTTTGGAACGGTTCGCAAATGGTGTACGGCGATGGCGACGGACAG
GlyArgLeuSerTyrAspGlySerAsnAlaAlaIleArgSerThrValHisTyrGlyArgGlyTyrAsnAsnAlaPheTrpAsnGlySerGlnMetValTyrGlyAspGlyAspGlyGln
360
1200
ACGTTTTTGCCCGTTTTCCGGCGGCATTGACGTCGTGGGGCATGAGTTGACCCATGCGGTGACGGATTATACGGCCGGGCTTGTTTACCAAAACGAATCTGGCGCCATCAATGAAGCGATG
ThrPheLeuProPheSerGlyGlyIleAspValValGlyHisGluLeuThrHisAlaValThrAspTyrThrAlaGlyLeuValTyrGlnAsnGluSerGlyAlaIleAsnGluAlaMet
400
1300
TCCGATATTTTCGGCACGCTCGTGGAGTTCTACGCCAACCGCAACCCGGACTGGGAGATTGGCGAAGACATTTACACGCCTGGGGTCGCCGGCGATGCGCTCCGCTCGATGTCCGACCCG
SerAspIlePheGlyThrLeuValGluPheTyrAlaAsnArgAsnProAspTrpGluIleGlyGluAspIleTyrThrProGlyValAlaGlyAspAlaLeuArgSerMetSerAspPro
440
1400
GCGAAATACGGCGATCCGGATCATTATTCCAAACGGTACACCGGCACGCAAGACAACGGCGGCGTCCATACAAACAGCGGCATCATCAATAAAGCGGCGTACTTGCTCAGCCAAGGCGGC
AlaLysTyrGlyAspProAspHisTyrSerLysArgTyrThrGlyThrGlnAspAsnGlyGlyValHisThrAsnSerGlyIleIleAsnLysAlaAlaTyrLeuLeuSerGlnGlyGly
480
1500
1600
GTCCATTATGGCGTGAGCGTCAACGGCATCGGCCGCGACAAAATGGGCAAAATTTTCTACCGGGCGCTTGTCTACTATTTGACGCCGACGTCGAACTTCAGCCAGCTGCGTCCCGCCTGC
ValHisTyrGlyValSerValAsnGlyIleGlyArgAspLysMetGlyLysIlePheTyrArgAlaLeuValTyrTyrLeuThrProThrSerAsnPheSerGlnLeuArgAlaAlaCys
520
1700
GTGCAAGCGGCCGCTGATTTGTACGGGTCGACAAGCCAAGAAGTCAACTCGGTGAAACAGGCGTTCAATGCGGTTGGAGTGTATTAAGACGATGAGGTCGTACGCGT
ValGlnAlaAlaAlaAspLeuTyrGlySerThrSerGlnGluValAsnSerValLysGlnAlaPheAsnAlaValGlyValTyr***
548

acid sequence homology was checked with an NEC PC-8001 computer (Nippon Electric Co., Tokyo, Japan) by the method of Novotny (24).

RESULTS

Location of *nprT*. We have shown that the structural gene for the thermostable neutral protease from *B. stearothermophilus* is cloned in a 4.5-megadalton (MDa) *Hin*dIII-*Eco*RI fragment of plasmid pNP22 (4). Based on the restriction enzyme analysis of pNP22, three *Mlu*I cleavage sites were found only within the 4.5-MDa fragment (Fig. 1). Using *Mlu*I sites, we attempted to construct the three types of deletion plasmids. pNP22 was digested with *Mlu*I, ligated, and used to transform *B. subtilis* MT-2 (Npr^-). Plasmids from transformants which exhibited either Km^r Tc^r Npr^+ or Km^r Tc^r Npr^- were extracted, and a correlation between the deletion and protease production was examined (Fig. 1). One *Mlu*I fragment (0.4 MDa) was indispensable, whereas the other fragment (1.8 MDa) could be omitted for protease production. These results showed that the *nprT* gene is located around the 0.4-MDa *Mlu*I fragment. A deletion plasmid, lacking only the 1.8-MDa *Mlu*I fragment, was designated as pNP22-1 (Km^r Tc^r $nprT^+$), and the plasmid was used for nucleotide sequence determination.

Nucleotide sequence of *nprT*. The nucleotide sequence of the *nprT* gene and its flanking regions are shown in Fig. 2. Starting from the codon ATG at nucleotide +49 and terminating in the TAA nonsense codon at nucleotide +1693, the single open reading frame was composed of 1,644 nucleotides (548 amino acid residues). Numerals above the nucleotide sequences in Fig. 2 are counted from the transcription start site as +1. Determination of the transcription start site is explained below. At 9 bases upstream from the ATG codon, there is a 9-base sequence, AGAAAGGGG (+31 to +39), which is complementary to the 3′ terminus of the 16S rRNA from *B. stearothermophilus* and *B. subtilis* (13); hence, it is the most probable ribosome-binding site (Shine-Dalgarno sequence) of the *nprT* gene. The free energy of formation of the most stable pairing was −24.0 Kcal (ca. −100 kJ)/mol (28).

Amino-terminal amino acids deduced from the nucleotide sequence exhibited a typical signal sequence with basic amino acids (Lys, Arg), followed by a hydrophobic core. Although the cleavage site of the signal sequence is still uncertain, hydrophobic residues within 25 amino-terminal amino acids are underlined in Fig. 2. The existence of the ribosome-binding site and the signal sequence supports the hypothesis that the open reading frame codes for the *nprT* gene. However, the molecular weight of the protease calculated from the amino acid sequence (59,514) was much higher than that of the extracellular form of the protease determined by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (36,000) (4).

Amino-terminal amino acid sequence of the extracellular neutral protease. The extracellular form of the protease was purified from the culture broth of *B. stearothermophilus* CU21(pNP28), and the amino-terminal amino acid sequence was determined through 14 cycles of Edman degradation by the sequencer. The first 14 amino acids were Val-Ala-Gly-Ala-Ser-Thr-Val-Gly-Val-Gly-Arg-Gly-Val-Leu. The sequence of the 14 amino acids completely matched that deduced from the nucleotide sequence starting at GTC (+736 to +738, Val). The matched amino acids are indicated with arrows in Fig. 2. Amino acid compositions of the extracellular protease agreed with those assessed from the nucleotide sequence (data not shown).

These results justify the hypothesis that the open reading frame codes for the neutral protease gene. The molecular weight of the extracellular protease, calculated as 34,579 from the amino acid sequence (319 amino acid residues), was within a permissible range (36,000) from the direct assessment mentioned above. A long amino acid sequence (229 amino acid residues), including the signal sequence, precedes the amino terminus of the extracellular enzyme. This fact suggests the existence of a pro sequence as is inferred in other protease genes of *B. subtilis* and *B. amyloliquefaciens* (26, 30, 33, 36, 37).

Determination of the transcription start site of the *nprT* gene. Although the promoter sequence was searched for in the 5′ flanking region (about 1.0 kilobase pairs) of the *nprT* gene, no consensus sequences for σ factors (σ^{55}, σ^{37}, σ^{32}, σ^{29}, and σ^{28}) (10) were found. To determine the promoter region, a transcriptional study was done by S1 nuclease mapping. The RNA fraction containing mRNA for the *nprT* gene was extracted and hybridized with the *Hha*I-*Mbo*I 223-bp (−149 to +74) fragment, whose *Mbo*I 5′ terminus was labeled with ^{32}P, as the DNA probe. The heteroduplex mixture was treated with S1 nuclease and loaded on a sequencing gel. DNA probes with and without base-specific chemical degradation (A+G and T+C) were loaded simultaneously (Fig. 3).

Two transcription start sites in Fig. 3 were found; since the minor one (thin arrow) was too weak to be worthy of further analysis, the major start site (thick arrow) was taken.

As a result, the transcription start site was fixed as guanine (+1), which was 49 bases upstream from the translation start site. Consequently, a rationale would suggest a TTTTCC sequence (−41 to −36) and a TATTTT sequence (−17 to −12), being 18 bp apart as are the −35 and −10 regions, respectively. Indeed, these sequences are highly homologous to the promoter sequence of the *spo0B* gene of *B. subtilis* (TTTTCT for the −35 region and TATAAT for the −10 region) (1) (Fig. 2). However, in vitro RNA polymerase studies are needed to show what kind of σ factor recognizes the promoter sequence.

DISCUSSION

We determined the complete nucleotide sequence of the gene for the thermostable neutral protease from *B. stearothermophilus*. The open reading frame, composed of 1,644 bases and 548 amino acid residues, was found. The amino-terminal amino acid sequence of the purified extracel-

FIG. 2. Nucleic and amino acid sequences of the neutral protease gene. The nucleotide sequence is presented from the *Mbo*I site (nucleotide −166) to the *Mlu*I site (nucleotide +1715). The nucleotide sequence is counted from the first base of the transcription start site (G) as +1. The amino acid sequence is shown beneath the nucleotide sequence. The amino-terminal amino acid sequence of extracellular protease, determined by Edman degradation, is specified below the amino acid sequence by arrows. The first amino acid of translation (Met) is counted as 1. Hydrophobic amino acids in the putative signal peptide (amino acids 1 to 25) are underlined. A probable Shine-Dalgarno sequence (nucleotide +31 to +39) and a promoter region (−35 and −10 regions) are shown by solid lines below the nucleotide sequence. Asterisks indicate a stop codon.

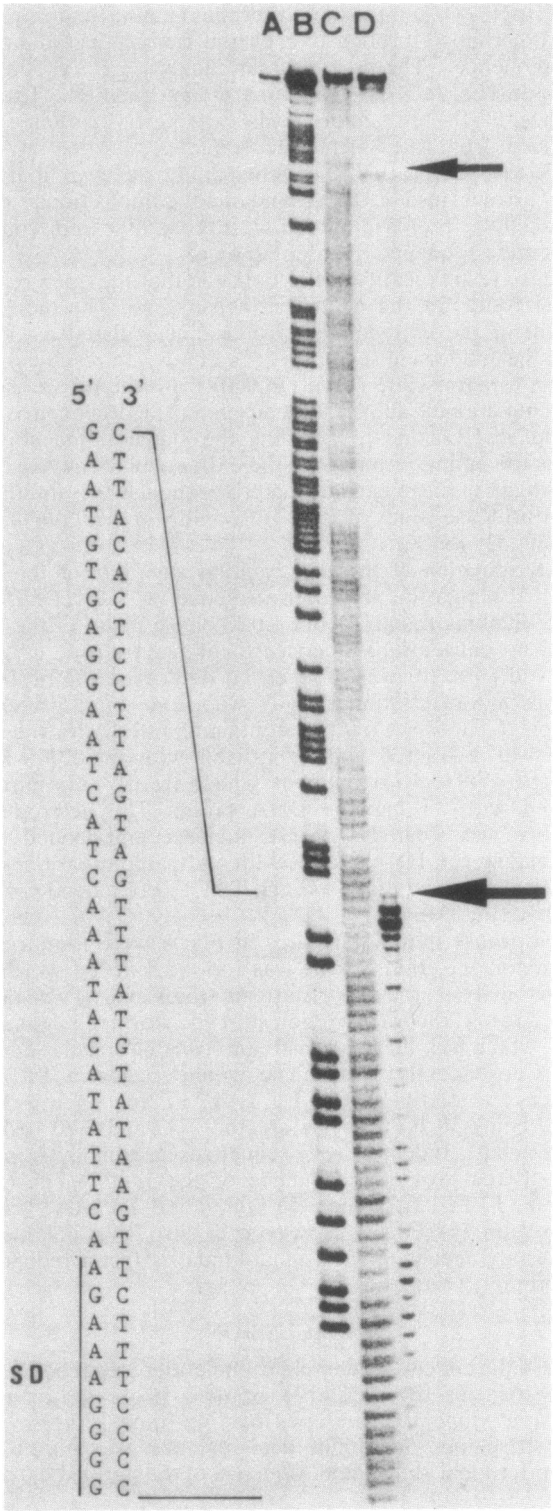


FIG. 3. Mapping of the 5′ end of the *nprT* transcript. After hybridization with an RNA fraction from *B. stearothermophilus* CU21(pNP28) and treatment with S1 nuclease, the DNA probe was analyzed on an 8% sequencing gel. Part of the DNA probe was loaded with and without the base-specific chemical cleavages (purines [A+G] or pyrimidines [T+C]) by the method of Maxam and Gilbert (18). The reaction time for T+C was prolonged to detect all ladders of sequence. Lanes: A, DNA probe; B, A+G; C, T+C; D, treated with S1 nuclease.

lular protease showed an existence of the long polypeptide (229 amino acid residues) preceding the amino terminus of the extracellular protease. This amino acid sequence in the amino-terminal region is much longer than the signal sequence of other secreted proteins. Indeed, the ordinary length of signal sequences is about 30 amino acid residues, and the mature form of protein follows just after the signal sequence (32). This sort of pre-pro structure is found for other proteases from *Bacillus* spp. (26, 30, 33, 36, 37), although its biological significance is obscure.

The amino acid sequence of the extracellular form of this protease was compared with that of another thermostable neutral protease from *B. thermoproteolyticus* (thermolysin), since only the amino acid sequence of extracellular thermolysin has been reported (29). The homology of the two amino acid sequences was very high (85%). We also compared the amino acid sequences deduced from the whole coding regions between the protease from *B. stearothermophilus* and the thermolabile neutral protease from *B. subtilis* (37) (Fig. 4). Homologous regions could be found mostly in the region of the extracellular form, and the homology of the extracellular region was about 30%. In contrast, the amino acid sequence homology between thermolabile neutral proteases from *B. subtilis* and those from *B. amyloliquefaciens* was very high (82%) for all of the coding regions (30, 37).

Although the homology of amino acid sequences varied among neutral proteases from the four different sources (*B. stearothermophilus*, *B. thermoproteolyticus*, *B. subtilis*, and *B. amyloliquefaciens*), there existed a region in which an amino acid sequence as long as 17 amino acid residues was

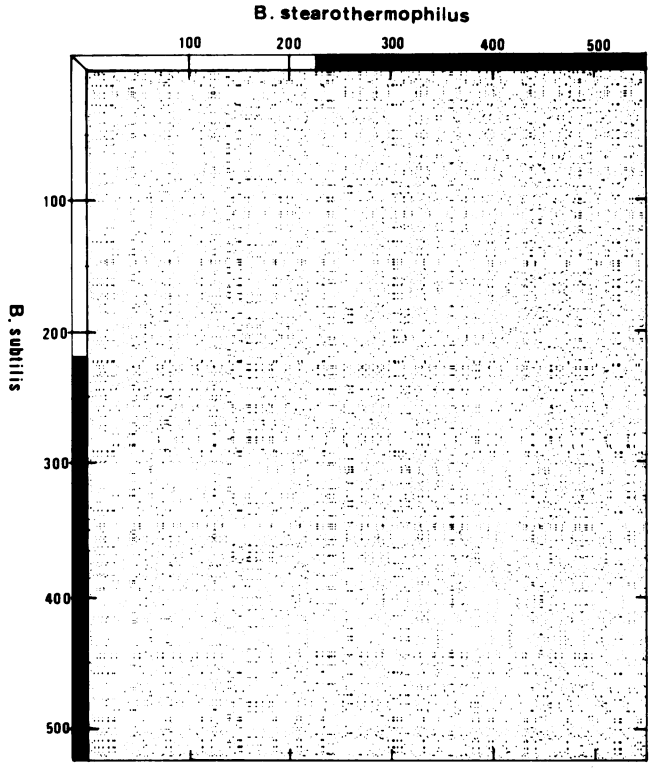


FIG. 4. Computer search for a homologous domain of the amino acid sequence. *B. stearothermophilus* versus *B. subtilis* (37) for the entire coding region. Open and solid bars in both axes correspond to pre-pro and extracellular regions of protease, respectively.

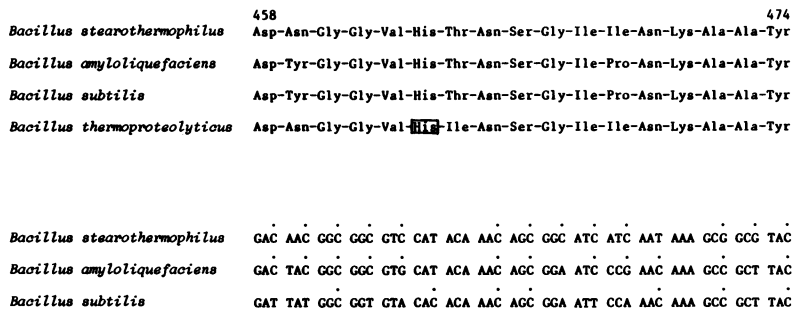


FIG. 5. (A) Amino acid sequence of the highly homologous region among neutral proteases from *B. stearothermophilus*, *B. amyloliquefaciens* (30), *B. subtilis* (37), and *B. thermoproteolyticus* (29). Numerals above the amino acids are counted from the translation start site (Met) as 1 with regard to the neutral protease from *B. stearothermophilus*. The active site of the neutral protease from *B. thermoproteolyticus* (His) (12) is surrounded by a box. (B) The nucleotide sequence of the conserved amino acid region for *B. stearothermophilus*, *B. amyloliquefaciens* (30), and *B. subtilis* (37). The G and C bases in the third letters of the codons are dotted above the nucleotides.

highly conserved (Fig. 5A). Within these conserved amino acid sequences, a histidine residue of thermolysin has been reported as an active site of proteolytic activity (12). These facts suggest that the conserved amino acid sequences are of the same three-dimensional structure and that the histidine residue functions as an active site in these neutral proteases despite the difference in their source and in physicochemical properties such as thermostability.

The promoter of the *nprT* gene was determined as TTTTCC for the −35 region and TATTTT for the −10 region. The promoter here was compared with various *Bacillus* promoters (Table 1). Although the −10 region of the *nprT* promoter was homologous to that recognized by RNA polymerase holoenzyme $E\sigma^{55}$, the −35 regions were not. No homologous sequences could be found among any other promoters in *Bacillus* spp. except for the *spo0B* gene in *B. subtilis* (1).

The guanine-plus-cytosine (G+C) content is reported to be important for the stability of DNA structure (27). Actually, the third letter of the codon is less specific than the first two for a given amino acid. The G+C content of the coding region for the *nprT* gene was 58 mol%, while that of the third letter of the codons was higher (72 mol%). These values were compared with those for the genes of mesophiles (*B. subtilis* and *B. amyloliquefaciens*) and thermophiles (*B. stearothermophilus* and *Thermus thermophilus*) (Table 2). The G+C content of the genes from the moderate thermophile *B. stearothermophilus* was higher than that of mesophilic *Bacillus* genes. This tendency was manifest in the case of an extreme thermophile, *T. thermophilus* (11). These data support an argument that the higher the G+C content becomes, especially in the third letter of codons, the higher the optimum growth temperature of each bacterium tends to be.

Codon usage for the conserved amino acid sequence containing the active site for neutral protease was examined (Fig. 5B). Out of 17 positions, G or C was the third letter of the codon in 13 (76.5 mol%) of the positions in *B. stearothermophilus*, 12 (70.6 mol%) of those in *B. amyloliquefaciens*, and 7 (41.1 mol%) of those in *B. subtilis*. The percentage for *B. amyloliquefaciens* was closer to that for *B. stearothermophilus* than to that for *B. subtilis*. As was referred to previously, the amino acid sequence of neutral protease for *B. amyloliquefaciens* was homologous to that of *B. subtilis* rather than *B. stearothermophilus*. On the contrary, *B. amyloliquefaciens* is reported to produce thermostable α-amylase, whose amino acid sequence is fairly homologous to that of *B. stearothermophilus* (23). Taking these observations into account, it might be said that *B. amyloliquefaciens* is partly featured by both characters of *B. stearothermophilus* (thermophile) and *B. subtilis* (mesophile).

TABLE 1. Promoter sequences in *Bacillus* spp.

Gene(s) (product)	−35 Region	−10 Region	σ Factors	Reference
B. subtilis				
Vegetative genes	TTGACA	TATAAT	σ^{55}	21
spoVG and *spoVC* (subtilisin)	AGG-TT	GG-ATTG-T	σ^{37}	20, 22, 31, 36
spoVG and *spoVC*	AAATC	TA-TG-TT-TA	σ^{32}	10
spoVG and *spoVC*	TT-AAA	CATATT	σ^{29}	10
Genomic genes	CTAAA	CCGATAT	σ^{28}	6
spo0B	TTTTCT	TATAAT	Unknown	1
Phage SPO1	T-AGGAGA--A	TTT-TTT	σ^{gp28}	14
Phage SPO1	CGTTAGA	GATATT	$\sigma^{gp33\text{-}34}$	10
B. stearothermophilus nprT	TTTTCC	TATTTT	Unknown	This work
B. thuringiensis (crystal protein)	AGTT-CA	ATAA$^{G}_{A}$A	Unknown	35
B. megaterium (protein C)	CTAGTAACAA	CGCAAACAT	Unknown	3

TABLE 2. Guanine-plus-cytosine content of the genes from thermophilic and mesophilic bacteria

Strain and gene product	G+C content (mol%)	G+C content of third letter (mol%)	Reference
B. stearothermophilus			
Neutral protease	58	72	This work
B. subtilis			
Neutral protease	44	42	37
Subtilisin	46	41	26
B. amyloliquefaciens			
Neutral protease	46	49	30
Alkaline protease	49	46	30, 33
B. stearothermophilus			
α-Amylase	50	57	23
Tyrosyl-tRNA synthetase	54	70	34
T. thermophilus			
Isopropylmalate dehydrogenase	70	89	11

ACKNOWLEDGMENTS

We are grateful to F. Sakiyama, Institute for Protein Research, Osaka University, for the determination of the amino-terminal amino acid sequence of neutral protease. We also thank M. Shibazaki for technical assistance.

LITERATURE CITED

1. **Bouvier, J., P. Stragier, C. Bonamy, and J. Szulmajster.** 1984. Nucleotide sequence of the *spo0B* gene of *Bacillus subtilis* and regulation of its expression. Proc. Natl. Acad. Sci. U.S.A. **81**:7012–7016.
2. **Duvall, E. J., D. M. Williams, S. Mongkolsuk, and P. S. Lovett.** 1984. Regulatory regions that control expression of two chloramphenicol-inducible *cat* genes cloned in *Bacillus subtilis*. J. Bacteriol. **158**:784–790.
3. **Fliss, E. R., and P. Setlow.** 1984. Complete nucleotide sequence and start sites for transcription and translation of the *Bacillus megaterium* protein C gene. J. Bacteriol. **158**:809–813.
4. **Fujii, M., M. Takagi, T. Imanaka, and S. Aiba.** 1983. Molecular cloning of a thermostable neutral protease gene from *Bacillus stearothermophilus* in a vector plasmid and its expression in *Bacillus stearothermophilus* and *Bacillus subtilis*. J. Bacteriol. **154**:831–837.
5. **Gilman, M. Z., and M. J. Chamberlin.** 1983. Developmental and genetic regulation of *Bacillus subtilis* genes transcribed by σ^{28}-RNA polymerase. Cell **35**:285–293.
6. **Gilman, M. Z., J. L. Wiggs, and M. J. Chamberlin.** 1981. Nucleotide sequences of two *Bacillus subtilis* promoters used by *Bacillus subtilis* sigma-28 RNA polymerase. Nucleic Acids Res. **9**:5991–6000.
7. **Hunkapiller, M. W., R. M. Hewick, W. J. Dreyer, and L. E. Hood.** 1983. High-sensitivity sequencing with a gas-phase sequenator. Methods Enzymol. **91**:399–413.
8. **Imanaka, T., M. Fujii, and S. Aiba.** 1981. Isolation and characterization of antibiotic resistance plasmids from thermophilic bacilli and construction of deletion plasmids. J. Bacteriol. **146**:1091–1097.
9. **Imanaka, T., M. Fujii, I. Aramori, and S. Aiba.** 1982. Transformation of *Bacillus stearothermophilus* with plasmid DNA and characterization of shuttle vector plasmids between *Bacillus stearothermophilus* and *Bacillus subtilis*. J. Bacteriol. **149**:824–830.
10. **Johnson, W. C., C. P. Moran, Jr., and R. Losick.** 1983. Two RNA polymerase sigma factors from *Bacillus subtilis* discriminate between overlapping promoters for a developmentally regulated gene. Nature (London) **302**:800–804.
11. **Kagawa, Y., H. Nojima, N. Nukiwa, M. Ishizuka, T. Nakajima, T. Yasuhara, T. Tanaka, and T. Oshima.** 1984. High guanine plus cytosine content in the third letter of codons of an extreme thermophile. J. Biol. Chem. **259**:2956–2960.
12. **Kester, W. R., and B. W. Matthews.** 1977. Comparison of the structures of carboxypeptidase A and thermolysin. J. Biol. Chem. **252**:7704–7710.
13. **Kozak, M.** 1983. Comparison of initiation of protein synthesis in procaryotes, eucaryotes, and organelles. Microbiol. Rev. **47**:1–45.
14. **Lee, G., and J. Pero.** 1981. Conserved nucleotide sequences in temporally controlled bacteriophage promoters. J. Mol. Biol. **152**:247–265.
15. **Maniatis, T., E. F. Fritsch, and J. Sambrook, ed.** 1982. Gel electrophoresis, p. 149–185. *In* Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
16. **Matsumura, M., Y. Katakura, T. Imanaka, and S. Aiba.** 1984. Enzymatic and nucleotide sequence studies of a kanamycin-inactivating enzyme encoded by a plasmid from thermophilic bacilli in comparison with that encoded by plasmid pUB110. J. Bacteriol. **160**:413–420.
17. **Matthews, B. W., J. N. Jansonius, P. M. Colman, B. P. Schoenborn, and D. Dupourque.** 1972. Three-dimensional structure of thermolysin. Nature (London) New Biol. **238**:37–41.
18. **Maxam, A. M., and W. Gilbert.** 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. **65**:499–560.
19. **Messing, J.** 1983. New M13 vectors for cloning. Methods Enzymol. **101**:20–78.
20. **Moran, C. P., Jr., W. C. Johnson, and R. Losick.** 1982. Close contacts between σ^{37}-RNA polymerase and a *Bacillus subtilis* chromosomal promoter. J. Mol. Biol. **162**:709–713.
21. **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.
22. **Moran, C. P., Jr., N. Lang, and R. Losick.** 1981. Nucleotide sequence of a *Bacillus subtilis* promoter recognized by *Bacillus subtilis* RNA polymerase containing σ^{37}. Nucleic Acids Res. **9**:5979–5990.
23. **Nakajima, R., T. Imanaka, and S. Aiba.** 1985. Nucleotide sequence of the *Bacillus stearothermophilus* α-amylase gene. J. Bacteriol. **163**:404–409.
24. **Novotny, J.** 1982. Matrix program to analyze primary structure homology. Nucleic Acids Res. **10**:127–131.
25. **Priest, F. G.** 1977. Extracellular enzyme synthesis in the genus *Bacillus*. Bacteriol. Rev. **41**:711–753.
26. **Stahl, M. L., and E. Ferrari.** 1984. Replacement of the *Bacillus subtilis* subtilisin structural gene with an in vitro-derived deletion mutation. J. Bacteriol. **158**:411–418.
27. **Stenesh, J.** 1976. Information transfer in thermophilic bacteria, p. 85–101. *In* M. R. Heinrich (ed.), Extreme environments—mechanisms of microbial adaptation. Academic Press, Inc., New York.
28. **Tinoco, I., Jr., P. N. Borer, B. Dengler, M. D. Levine, O. C. Uhlenbeck, D. M. Crothers, and J. Gralla.** 1973. Improved estimation of secondary structure in ribonucleic acids. Nature (London) New Biol. **246**:40–41.
29. **Titani, K., M. A. Hermodson, L. H. Ericsson, K. A. Walsh, and H. Neurath.** 1972. Amino-acid sequence of thermolysin. Nature (London) New Biol. **238**:35–37.
30. **Vasantha, N., L. D. Thompson, C. Rhodes, C. Banner, J. Nagle, and D. Filpula.** 1984. Genes for alkaline protease and neutral protease from *Bacillus amyloliquefaciens* contain a large open reading frame between the regions coding for signal sequence and mature protein. J. Bacteriol. **159**:811–819.
31. **Wang, P.-Z., and R. H. Doi.** 1984. Overlapping promoters transcribed by *Bacillus subtilis* σ^{55} and σ^{37} RNA polymerase holoenzymes during growth and stationary phases. J. Biol.

Chem. **259**:8619–8625.
32. **Watson, M. E. E.** 1984. Compilation of published signal sequences. Nucleic Acids Res. **12**:5145–5164.
33. **Wells, J. A., E. Ferrari, D. J. Henner, D. A. Estell, and E. Y. Chen.** 1983. Cloning, sequencing, and secretion of *Bacillus amyloliquefaciens* subtilisin in *Bacillus subtilis*. Nucleic Acids Res. **11**:7911–7925.
34. **Winter, G., G. L. E. Koch, B. S. Hartley, and D. G. Barker.** 1983. The amino acid sequence of the tyrosyl-tRNA synthetase from *Bacillus stearothermophilus*. Eur. J. Biochem. **132**: 383–387.
35. **Wong, H. C., H. E. Schnepf, and H. R. Whiteley.** 1983. Transcriptional and translational start sites for the *Bacillus thuringiensis* crystal protein gene. J. Biol. Chem. **258**:1960–1967.
36. **Wong, S.-L., C. W. Price, D. S. Goldfarb, and R. H. Doi.** 1984. The subtilisin E gene of *Bacillus subtilis* is transcribed from a σ^{37} promoter *in vivo*. Proc. Natl. Acad. Sci. U.S.A. **81**:1184–1188.
37. **Yang, M. Y., E. Ferrari, and D. J. Henner.** 1984. Cloning of the neutral protease gene of *Bacillus subtilis* and the use of the cloned gene to create an in vitro-derived deletion mutation. J. Bacteriol. **160**:15–21.