Jour. Microbiol. March 1996 p. 68-73

Vol. 34, No. 1

Cloning and Molecular Characterization of a *nprX* gene of *Bacillus subtilis* NS15-4 Encoding a Neutral Protease

Seunghwan Lee, Ki-Hong Yoon[2], Hee-Sop Nam[3], Tae Kwang Oh[2], Seog Jae Lee[1], and Keon-Sang Chae*

Department of Molecular Biology, [1]Department of Chemistry, Chonbuk National University, Chonju 561-756
[2]Applied Microbiology Research Group, Korea Research Institute of Bioscience and Technology, P.O. Box 115, Yusung, Taejeon 305-600
[3]Research and Development Center, Nong Shim Co., Ltd., Kyung-ki 435-030, Korea

(Received September 15, 1995/Accepted January 12, 1996)

An *nprX* gene of *Bacillus subtilis* NS15-4 encoding a neutral protease was cloned and its molecular characteristics were analyzed. The complete nucleotide sequence indicated that there is an open reading frame (ORF) possibly encoding 521 amino acid polypeptide. The ORF used all codons except two cysteine and a proline having a codon bias index (CBI) of 0.09 in *Escherichia coli*. There were homologous sequences to the consensus sequence of −35 and −10 regions of *E. coli* promoters and to a Shine-Dalgarno (SD) sequence located 25 bp downstream of a major transcription initiation site. Moreover, there were also five minor transcription initiation sites at 6, 7, 8, 14 and 15 nt downstream of the major site. Northern blot analysis revealed the presence of about 1.8 kb mRNA transcript in *E. coli* having the *nprX* gene. The nucleotide sequence was identified in GenBank to be a gene for a neutral protease of *B. subtilis* with six nucleotide difference in the ORF region. The flanking regions of the NprX ORF showed much more differences from those of other neutral protease genes except the *nprE* gene of *B. subtilis*, which has the most homology to the *nprX* gene, and of which the flanking regions were identical to those of the *nprX* gene.

Key words: *Bacillus subtilis*, neutral protease transcription initiation site, *nprX* gene

Proteases are one class of the most industrially important enzymes. Among various sources of proteases, *Bacillus* spp. have long been regarded as attractive producers of them, becuase of their non-pathogenic properties to human and animals, high secretion capacity, and knowledges about the fermentation technology.

Bacillus subtilis has been known to have at least six proteases of which the genes have been cloned and characterized. They are the neutral protease A (18, 21), a subtilisin (17), an extracellular protease (19), a metalloprotease (14), the bacillopeptidase F (15, 20) and the minor protease Vpr (16). Among the six, the activities of the neutral protease A and the alkaline protease (subtilisin) account for about 70% and 20% of total protease activities in the culture medium, respectively (6, 17, 21).

A *Bacillus subtilis* strain, NS15-4, was isolated from soil for its excellent protease activities of industrial values. One of the protease produced by this bacterium was purified and characterized to be an alkaline protease (8). From this bacterium, we have tried to isolate both genes encoding an alkaline protease and a neutral protease in order to express them in large amounts, and to know molecular characteristics of the genes.

In this report, the *nprX* gene encoding a neutral protease was isolated in a 2.7 kb DNA fragment. The detailed molecular characteristics of this gene including the complete nucleotide sequence and the transcription initiation sites are described.

Materials and Methods

Strains and media

B. subtilis NS15-4 was isolated from nature and identified at first to be *B. amyloliquefaciens* by analyzing the cellular fatty acid profiles (8). Later, it was re-identified as *B. subtilis* by PCR and Southern hybridization (data not shown). *E. coli* JM83 and DH5α were used as hosts

* Corresponding author.

for transformation. Bacterial strains were grown on LB medium where ampicillin at 50 μg/ml was added, when necessary. For selection of protease-positive clones, 1% skim milk was added to the LB medium.

Enzymes, reagents and nucleic acids

Sequenase (ver 2.0) and Moloney Murine Leukemia Virus (MMLV) reverse transcriptase were obtained from USB Co. and GibcoBRL (MD, USA), respectively. All enzyme reactions were carried out according to the recommendation of manufacturers, unless indicated otherwise. For amplification and sequencing of an inserted DNA in pUC19, the sense (NEB#1233) and the antisense (NEB#1224) primers were used. Also, five primers synthesized by DNA International Inc., (OR, USA) were used for sequencing some parts of the gene. The sequences of primer 1, 2, 3, 4 and 5 were 5′-AAACAAACACAGGACAAT-3′ (−86～−69), 5′-AACAGCGTCTGGTGATTT-3′ (517～500), 5′-AACAAAAGACGGCAGCTA-3′ (568～585), 5′-CGCTGTCACATCTAATGA-3′ (1126～1109) and 5′-CCAAGAAACAGCCAACTT-3′ (1151～1168), respectively.

DNA manipulation

Plasmid DNAs from *E. coli* were obtained by the modified alkaline-lysis procedure (12). Chromosomal DNA from *B. subtilis* NS15-4 was isolated by the spooling procedure of Marmur (9). The genomic DNA of the bacterium was digested partially with *Sau*3AI and fractionated on a sucrose gradient by ultracentrifugation. DNA fragments of 8 to 12 kb were ligated with pUC19 linearized by *Bam*HI to make a genomic library. All other DNA manipulations were carried out essentially as described by Sambrook *et al.* (12).

Polymerase chain reaction (PCR)

All subclones were identified by PCR. PCR was carried out under the conditions described separately with minor modifications (2). A small colony paste of the recombinant subclone grown on a LB plate was directly used as a template for PCR.

Determination of the nucleotide sequence

Nucleotide sequence of the *nprX* gene was determined by the chain termination method (13). The sequencing templates were prepared by asymmetric PCR. PCR conditions were same as those for ordinary PCR except the amount of primers and the concentration of dNTP. The amount of one primer was reduced to one-50th of the other one in order to obtain a single-stranded DNA template, and the dNTP concentration to 50 mM

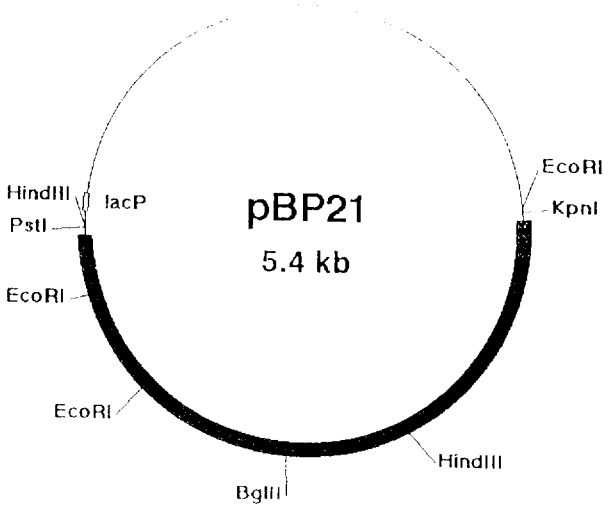

Fig. 1. Restriction map of pBP21. The single line and the filled box indicate pUC19 vector and *B. subtilis* DNA, respectively.

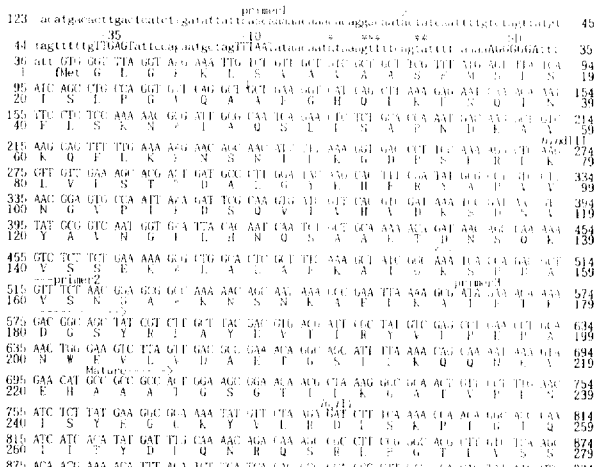

Fig. 2. The nucleotide sequence of the *nprX* gene. The asterisks above the sequences indicate the transcription initiation sites. Putative −35 and −10 regions and SD sequence are shown by underlines and bold-phase capitals, respectively. The vertical arrow indicates the site possibly processed by a signal peptidase. The N-terminus of a mature protein is indicated above the sequence. A sequence possibly forming a hairpin structure is shown by underlines from 1611 to 1636. The GenBank access number of this sequence is U30962.

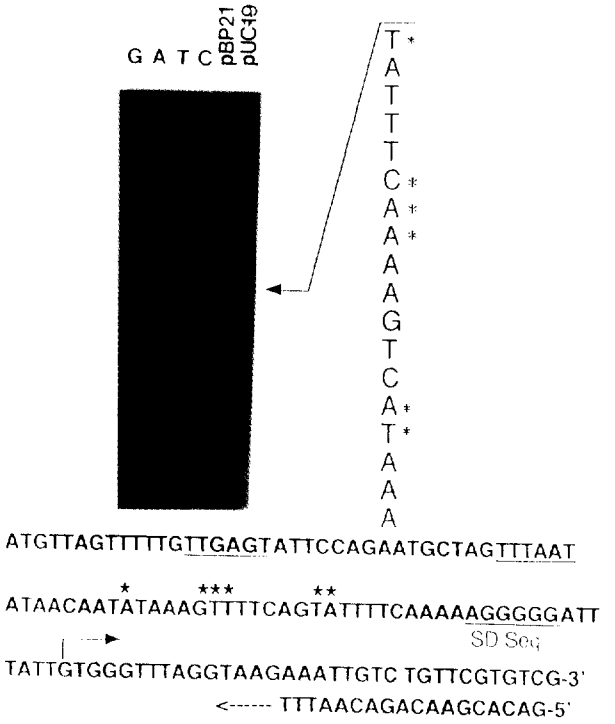

Fig. 3. Primer extension analysis to determine the transcription initiation sites. The asterisks above the sequences indicate the transcription initiation sites. Possible −35 and −10 regions are shown by underlines. The primer sequence for the primer extension analysis is shown under the *nprX* sequence. The translation initiation codon (GTG) is presented by a bent arrow.

to minimize mis-incorporation errors (10).

RNA preparation

Total RNA of *E. coli* was isolated by guanidine thiocyanate (GT)-CsCl ultracentrifugation method (3). The bacterial pellet was suspended and lyzed in GT buffer. The lysate was then laid on a CsCl cushion (5.7M CsCl) and ultracentrifugated at 180,000 g for 21 h. The RNA pellet was dissolved in 400 ml of GT buffer and RNA was reprecipitated with 3 volume of ethanol.

Primer extension analysis

Primer extension reaction was performed by a modification of the method of Guilfoile and Hutchinson (4). A synthetic 18-mer primer complementary to the sequence spanning 53~70 was used as a primer for reverse transcription (Fig. 3). After the reaction, the product labeled with [α-^{32}P]-dCTP was analyzed onto a 6% sequencing gel in parallel with sequencing products made from the same primer, followed by autoradiography.

Northern analysis

About 20 μg of total RNA was subjected to formaldehyde-denaturating gel electrophoresis, and blotted overnight onto a Hybond N+ membrane (Amersham Laboratories) by capillary transfer. The membrane was hybridized with labeled 0.6 kb *Bgl*II-*Eco*RI fragment according to the conditions recommended by the manufacturer.

Results and Discussions

Cloning the *nprX* gene

E. coli JM83 was transformed with a genomic library of *B. subtilis* NS15-4, and plated on LB plates containing 50 μg/ml of ampicillin and 1% skim milk. After incubation at 37℃ overnight, a clear zone and a creamy halo were observed around four transformants. The four colonies were picked and plasmids from the transformants were isolated. One plasmid, labeled as pBP2, produced a large clear zone and a creamy halo again, when transforming *E. coli* DH5α. The size of the inserted DNA in pBP2 was measured to be about 9.3 kb. The gene encoding this protease activity, of which the nucleotide sequence was identified in GenBank to be a gene for a neutral protease (see below), was named as the *nprX*. The *nprX* gene was localized within a 2.7 kb *Sau*3AI-fragment of the inserted DNA in pBP2, yielding a smaller pBP21. Fig. 1 shows a restriction map of pBP21.

Nucleotide sequence of the *nprX* gene

Full-length sequence of the *nprX* gene was completely determined in both direction. An 1,563 nt open reading frame (ORF) possibly encoding 521 amino acid polypeptide was found as analyzed by DNASIS™ program (Hitachi, Japan) (Fig. 2). The N-terminus of the ORF seemed to have a 27 amino acid signal sequence as shown by the vertical arrow. The mature NprX protein may be 300 amino acid polypeptide as indicated in Fig. 2. To determine the transcriptional initiation site of the *nprX* gene, total RNA was isolated from *E. coli* having the *nprX* gene and primer extension analysis was performed. One major transcriptional initiation site at +1 and five minor transcriptional initiation sites at 6, 7, 8, 14, 15 were detected (Fig. 3). Upstream of these sites, putative promoter sequences were found in good agreement with a consensus sequence of *E. coli* promoters and that for sigma factor functioning on the same sequence during vegetative growth of *B. subtilis* (5, 7, 11). Also, a putative SD sequence was located between the transcription initiation sites and a putative translation initiation codon, as in other cases of *B. subtilis* genes for a neutral proteases (21). The distance between the SD sequence and the translation initiation codon GTG was 6 nt. These structural characteristics indicated that this gene can be

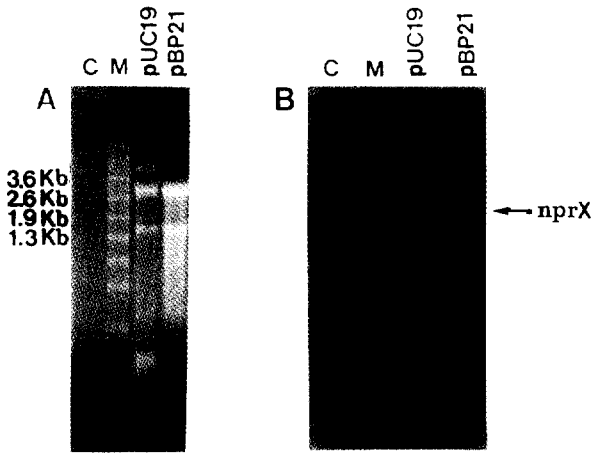

Fig. 4. Northern blot analysis identifying the *nprX* mRNA. (A) About 20 μg of total RNA was subjected to formaldehyde-denaturating gel electrophoresis, and blotted overnight onto a Hybond N+ membrane by capillary transfer. (B) The nylon membrane was hybridized with a labeled 0.6 kb *Bgl*II-*Eco*RI fragment of pBP21. Lane C, M, pUC19, and pBP21 mean a positive control of the probe, RNA marker, RNAs isolated from *E. coli* DH5α having pUC19 and pBP21, respectively.

Codon usage analysis

The NprX ORF was comprised of 521 amino acid codons. It used all amino acid codons except two cysteine codons, as other related genes, and one proline codon (Table 1). The codon bias index of the *nprX* gene was calculated to be 0.09 in *E. coli*, suggesting that this gene may not efficiently be translated in *E. coli* because it has been known that the transcription level was partly depended on the codon bias index (1). But the *nprX* gene in *E. coli* was expressed to some extent. It could be inferred from the results that the promoter region showed a strong homology to the consensus sequence of *E. coli* promoters and the *nprX* gene was cloned in a high copy plasmid such as pUC19.

Homology search in GenBank

The nucleotide sequence of the *nprX* gene was identified in GenBank to be homologous to other genes encoding neutral proteases with 6 nucleotide differences in the ORF regions (Table 2). Sequences of flanking regions of the NprX ORF were more different from those of other genes except the *nprE* gene of *B. subtilis*. The *nprX* gene was identical to the *nprE* gene except 6 nucleotides in the ORF region, and the flanking regions were completely identical to each other in their nucleotide sequences. These differences may arise from several possibilities. First, there may be sequencing errors, if any in the *nprE* gene. Second, the strain of *B. subtilis* NS15-4 is a natural isolate, and thus possibly

Table 1. Codon usage analysis of the *nprX* gene of *B. subtitlis* NS15-4.[a]

codon	Count	%	codon	Count	%	codon	Count	%	codon	Count	%
TTT-Phe	8	1.53	TCT-Ser	15	2.87	TAT-Tyr	15	2.87	TGT-Cys	0	0.00
TTC-Phe	7	1.34	TCC-Ser	7	1.34	TAC-Tyr	13	2.49	TGC-Cys	0	0.00
TTA-Phe	9	1.72	TCA-Ser	12	2.29	TAA-Ter	1		TGA-Ter	0	
TTG-Phe	7	1.14	TCG-Ser	2	0.38	TAG-Ter	0		TGG-Trp	4	0.76
CTT-Leu	11	2.10	CCT-Pro	9	1.72	CAT-His	4	0.76	CGT-Arg	3	0.57
CTC-Leu	7	1.34	CCC-Pro	0	0.00	CAC-His	6	1.14	CGC-Arg	3	0.57
CTA-Leu	2	0.38	CCA-Pro	8	1.53	CAA-Gln	12	2.29	CGA-Arg	3	0.57
CTG-Leu	4	0.76	CCG-Pro	2	0.38	CAG-Gln	11	2.10	CGG-Arg	1	0.19
ATT-Ile	10	1.91	ACT-Thr	5	0.95	AAT-Asn	13	2.49	AGT-Ser	2	0.38
ATC-Ile	10	1.91	ACC-Thr	3	0.57	AAC-Asn	25	4.78	AGC-Ser	14	2.68
ATA-Ile	1	0.19	ACA-Thr	20	3.83	AAA-Lys	32	6.13	AGA-Arg	2	0.38
ATG-Met	3	0.57	ACG-Thr	9	1.72	AAG-Lys	8	1.53	AGG-Arg	1	0.19
GTT-Val	14	2.68	GCT-Ala	17	3.25	GAT-Asp	16	3.06	GGT-Gly	13	2.49
GTC-Val	12	2.29	GCC-Ala	15	2.87	GAC-Asp	15	2.87	GGC-Gly	15	2.87
GTA-Val	4	0.76	GCA-Ala	11	2.10	GAA-Glu	19	3.63	GGA-Gly	9	1.72
GTG-Val	6	1.14	GCG-Ala	9	1.72	GAG-Glu	3	0.57	GGG-Gly	1	0.19

[a]One of GTG valine codons was used as a translation initiation codon.

expressed in *E. coli*.

the mRNA transcript size is expected to be 1.6 kb. However, about 1.8 kb RNA transcript of the *nprX* gene was detected by northern blot analysis (Fig. 4). Thus, this result indicated that the putative hairpin structure is not a transcription terminator, and that about 190 nt 3′ untranslated region in the *nprX* mRNA may exist. At present, functions of the putative hairpin structure behind the NprX ORF are unknown.

Table 2. Differences of nucleotide sequences of the *nprX* gene from other related ones

Gene[a]	GenBank Acc. No.	No. of amino acids	No. of different nt		Positions[b]
			in the ORF	in flanking regions	
nprE	K01985	521	6	0	65, 66, 169, 170, 268, 269
npr	D10773	521	6	several tens	169, 556, 574, 658, 727, 1340
A10786	A10786	521	6	several tens	48, 169, 556, 574, 727, 1340

[a]All these three genes are *Bacillus subtilis* ones. [b]Positions of the *nprX* gene that are different from others are shown.

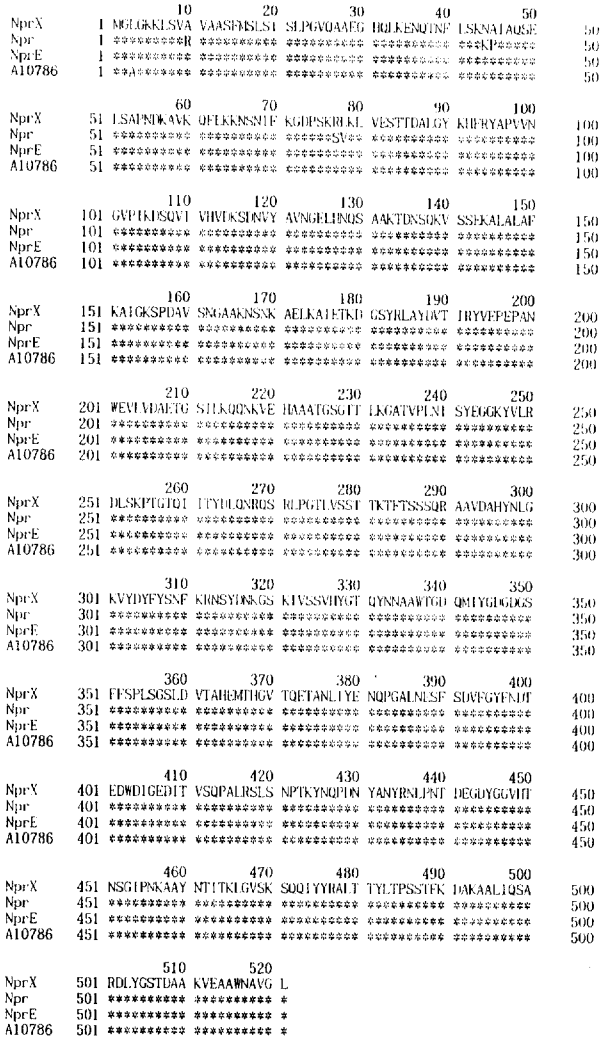

Fig. 5. Amino acid sequence comparison of the NprX ORF with other related *B. subtilis* genes for neutral proteases. Asterisks indicate identical amino acids and only amino acids different from the NprX sequence are shown.

As Yang *et al.* (21) showed a sequence possibly forming a hairpin structure just downstream the NprE ORF, the same sequence was found downstream of the NprX ORF. If the structure acts as a transcription terminator, differs from other *B. subtilis* strains in the nucleotide sequence of some genes. However, it was unexpected that the nucleotide sequences of the NprX and the NprE ORFs were different by six nucleotides although the 350 bp flanking regions of them were identical to each other.

The deduced amino acid sequence of the *nprX* gene showed an identity to a neutral protease ORF (Npr ORF) of *B. subtilis,* and differed by only one or five amino acids to other two ORF, the NprE ORF and the ORF in A10786 (Fig. 5). Noticeable were that only the amino acid sequence of the NprX ORF was identical to that of Npr ORF, but the nucleotide sequence in the NprX ORF differed from the *npr* gene by 6 nucleotides and the flanking regions of the ORFs differed more significantly, and that the *nprX* gene showed 5 amino acid difference from the *nprE* gene in the amino acid sequences with identity in the nucleotide sequences of the flanking regions (Fig. 5).

Acknowledgments

This work was supported by a grant of Genetic Engineering Research Program of The Ministry of Education in 1989 to K.-S. Chae.

References

1. **Bennetzen, J.L. and B.D. Hall,** 1982. Codon selection in yeast. *J. Biol. Chem.* **257**, 3026-3031.
2. **Chae, K.-S., K. Murakawa, K. Okubo, and K. Matsubara,** 1994. The use of overlapping and tailed short primers in the chromosomal assignment of short cDNAs by the polymerase chain reaction. *Gene* **142**, 199-205.
3. **Chirgwin, J.M., A.E. Przybyla, R.J. MacDonald, and W.J. Rutter,** 1979. Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. *Biochemistry* **18**, 5294-5299.
4. **Guilfoile, P.G. and C.R. Hutchinson,** 1992. Sequence and transcriptional analysis of the *Streptomyces glaucescens tcmAT* tetracenomycin C resistance and repressor gene loci. *J. Bacteriol.* **174**, 3651-3658.
5. **Henkin, T.M. and A.L. Sonenshein,** 1987. Mutations of the *Escherichia coli lacUV5* promoter resulting in increased expression in *Bacillus subtilis*. *Mol. Gen. Genet.* **209**, 467-474.
6. **Kawamura, F. and R.H. Doi,** 1984. Construction of a Bacillus subtilis double mutant deficient in extracellular alkaline and neutral proteases. *J. Bacteriol.* **160**, 442-444.

7. **Kenney, T.J. and C.P. Moran, Jr.,** 1991. Genetic evidence for interaction of σ_A with two promoters in *Bacillus subtilis*. *J. Bacteriol.* **173**, 3282-3290.
8. **Kim, H.-K., K.-H. Kim, J.-K. Lee, Y.-O. Kim, H.-S. Nam, and T.K. Oh,** 1995. Characterization of a thermostable protease from thermophilic *Bacillus amyloliquefaciens* NS 15-4. *Kor. J. Appl. Microbiol. Biotech.* **23**, 322-328.
9. **Marmur, J.,** 1961. A procedure for the isolation of deoxyribonucleic acid from microorganisms. *J. Mol. Biol.* **3**, 208-218.
10. **McCabe, P.C.,** 1990. Production of single-stranded DNA by asymmetric PCR, pp. 76-83. *In* (Innis, M.A., D.H. Gelfaud, J.J. Swinky, and T.J. White, (ed.), PCR Protocols. Academic Press Inc., San Diego, California, USA.
11. **Moran, Jr., C.P., 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.
12. **Sambrook, J., E.F. Fritsch, and T. Maniatis,** 1989. Molecular Cloning; A Laboratory Manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor.
13. **Sanger, F., S. Nicklen, and A.R. Coulson,** 1977. DNA sequencing with chain terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74**, 5463-5467.
14. **Sloma, A., A. Ally, D. Ally, and J. Pero,** 1988. Gene encoding a minor extracellular protease in *Bacillus subtilis*. *J. Bacteriol.* **170**, 5556-5563.
15. **Sloma, A., G.A. Rufo, Jr., C.F. Rudolph, B.J. Sullivan, K.A. Theriault, and J. Pero,** 1990. Bacillopeptidase F of *Bacillus subtilis*: purification of the protein and cloning of the gene. *J. Bacteriol.* **172**, 1470-1477.
16. **Sloma, A., G.A. Rufo, Jr., K.A. Theriault, M. Dwyer, S. Wilson, and J. Pero,** 1991. Cloning and characterization of the gene for an additional extracellular serine protease of *Bacillus subtilis*. *J. Bacteriol.* **173**, 6889-6895.
17. **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.
18. **Tran, L., X.-C. Wu, and S.-L. Wong,** 1991. Cloning and expression of a novel protease gene encoding an extracellular protease from *Bacillus subtilis*. *J. Bacteriol.* **173**, 6364-6372.
19. **Wong, S.-L., C.W. Price, D.S. Goldfarb, and R.H. Doi,** 1984. The subtilisin E gene of *Bacillus subtilis* in transcribed from a σ^{37} promoter *in vivo*. *Proc. Natl. Acad. Sci. USA* **81**, 1184-1188.
20. **Wu, X.-C., S. Nathoo, A.S.-H. Pang, T. Carne, and S.-L. Wong,** 1990. Cloning, genetic organization, and characterization of a structural gene encoding bacillopeptidase F from *Bacillus subtilis*. *J. Biol. Chem.* **265**, 6845-6850.
21. **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.