JOURNAL OF BACTERIOLOGY, July 2000, p. 4104–4107 Vol. 182, No. 14
0021-9193/00/$04.00+0

Expression of the Neutral Protease Gene from a Thermophilic *Bacillus* sp. BT1 Strain in *Bacillus subtilis* and Its Natural Host: Identification of a Functional Promoter

BRANISLAV VEČEREK[1]* AND GERARD VENEMA[2]

Laboratory of Enzyme Technology, Institute of Microbiology CAS, Prague 4, 142 20, Czech Republic,[1] and Department of Genetics, University of Groningen, 9751 NN Haren, The Netherlands[2]

Received 6 March 2000/Accepted 20 April 2000

The expression of the neutral protease gene (*npr*) from the thermophilic *Bacillus* sp. BT1 strain was studied in its natural host and in mesophilic *Bacillus subtilis*. In the thermophilic BT1 strain, the transcription of the protease gene is initiated from its own promoter, just 5′ to the gene. In contrast, in heterologous *B. subtilis* this thermophilic *npr* promoter does not function, and expression of the *npr* gene results from transcription originating upstream of an adjacent gene, open reading frame X (ORF X). A functional promoter was identified 5′ to ORF X that is required for efficient expression of the *npr* gene in *Bacillus subtilis* as verified by primer extension, reverse transcription-PCR, and 5′ rapid amplification of cDNA ends experiments. These data suggest that transcriptional signals used in thermophilic *Bacillus* sp. BT1 strain are different from those used in *B. subtilis*.

Many species of the genus *Bacillus* produce a variety of thermostable extracellular enzymes, some of which are industrially important. Among these enzymes, the neutral proteases (Npr) have been extensively studied not only for industrial production but also for the elucidation of mechanisms involved in thermostability of enzymes (6, 7, 16). Several *npr* genes from thermophilic bacilli have been cloned and expressed in *Bacillus subtilis* (4, 8, 12, 15) to study the gene and its product in a well-characterized host. However, relatively little is known about the regulation of the expression and promoter utilization of these thermophilic genes in *B. subtilis* and in natural hosts as well.

The thermostable neutral protease-encoding gene (*npr*) from a thermophilic strain of *Bacillus* sp. BT1 was cloned and sequenced (17). The gene was subcloned in *B. subtilis* on the 4.3-kb *Sau*3AI fragment and subsequently confined to the shorter 2.2-kb *Eco*RI-*Sau*3AI fragment. To further enhance the protease production in *B. subtilis* the *npr* gene was placed into the high-copy-number plasmid pGDV1. When cloned into pGDV1, the 4.3-kb fragment (pGBR5) produced about 15-fold more protease than did cloning of the 2.2-kb fragment in pGDV1 (pGBR50) in the protease-deficient *B. subtilis* DB117, indicating that important information for the expression of the *npr* gene resided in the 2.1-kb region upstream of the *npr* gene (17).

To characterize this region several plasmid derivatives of pGBR5 were constructed by generating deletions from the 5′ direction on the 4.3-kb *Sau*3AI fragment (Fig. 1). The proteolytic activity encoded by plasmid derivatives was determined by using Azocoll (Sigma) as a substrate (1). One unit of proteolytic activity was arbitrarily defined as the amount of enzyme causing an increase in A_{520} of 0.5 per 30 min. Deletions up to first *Bsa*BI site from *Sau*3AI end did not affect the protease production (pGBR80 and pGBR75), whereas the deletion of *Sau*3AI-*Msc*I fragment abolished the enhanced protease production (pGBR70). Therefore, the nucleotide sequence of the 1.5-kb *Bsa*BI-*Eco*RI fragment was determined and coupled with the already reported sequence of the 2,231-bp *Eco*RI-*Sau*3AI fragment containing the *npr* gene (GenBank accession number U07824 [17]). Within this region, only one large open reading frame (ORF) composed of 1,245 bp was found, and this was designated ORF X (Fig. 1). The ORF X sequence continues through the *Eco*RI site and ends 112 bp upstream from the *npr* gene. A potential Shine-Dalgarno (SD) sequence, AGATGGGAGGTGA, extensively complementary (ΔG = −23.4 kcal/mol) to the 3′ terminus of *B. subtilis* 16S RNA (10) was found upstream from the probable translation start codon ATG (Fig. 2). When comparing the ORF X sequence with database entries, no significant homology to other known genes was found.

To analyze the possible role of the ORF X or its gene product in the protease production, plasmids carrying deletions extending into the ORF X gene were constructed from pGBR5 (Fig. 1). Deletion of the major part of the ORF X, which, in addition, changed the reading frame of ORF X (620-bp *Msc*I-*Eco*RI fragment; pGBR65) had no effect on the enhanced production of the protease, indicating that the potential protein encoded by ORF X is not necessary for activation. However, deletion of the 5′-terminal region of the ORF X (458-bp *Ear*I fragment; pGBR60) probably containing the ORF X regulatory sequences reduced the protease production to the level produced by DB117(pGBR50).

To characterize the expression of the protease gene in *B. subtilis* more precisely, the transcription start sites of the *npr* and ORF X genes were mapped by primer extension. Total RNA was isolated from DB117(pGBR5) cells grown in Luria-Bertani medium to early stationary phase using the RNeasy purification kit (Qiagen). The primers NPREX (5′-CGAACG GCCAAGCCATCAATC-3′) and ORFEX (5′-GGCAATAT GGCACAGTTCTCG-3′) complementary to 5′ end of the *npr* (from nucleotides [nt] 443 to 423 in the published sequence) and ORF X (from nt 652 to 632) genes, respectively, were end labeled with [γ-^{32}P]ATP (Amersham) and extended by avian myeloblastosis virus (AMV) reverse transcriptase (Promega) using total RNA extracted from DB117(pGBR5) as a tem-

* Corresponding author. Mailing address: Czech Academy of Sciences, Institute of Microbiology, Vídeňská 1083, Prague 4-Krč, 142 20, Czech Republic. Phone: 420 2 4752237. Fax: 420 2 4727021. E-mail: vecerek@biomed.cas.cz.

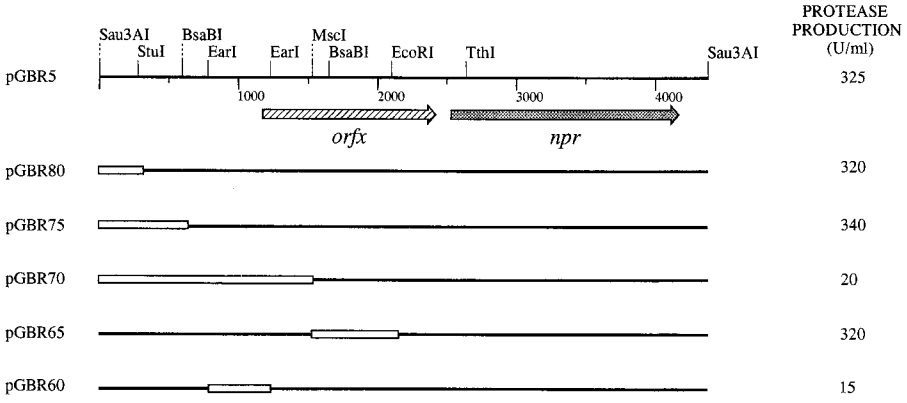

FIG. 1. Localization of the ORF X and *npr* genes on the cloned 4.3-kb *Sau*3AI fragment. The position and direction of transcription of the genes are indicated by large arrows. The structure of pGBR5 derivatives is shown below the physical map of the cloned fragment. Black lines indicate the DNA from *Bacillus* sp. BT1 strain; the open boxes represent deletions of the specific fragments. The protease production encoded by the corresponding plasmid is indicated on the right side.

plate. Hybridization and extension procedures were carried out essentially as described by Sambrook et al. (13). The synthesized cDNA was precipitated with ethanol, dissolved in gel loading buffer, and analyzed on a 6% polyacrylamide-urea sequencing gel. No apparent signal was detected using the *npr*-specific primer, whereas as shown in Fig. 3, two cDNA products were obtained using ORF X-specific primer. The obtained products correspond to initiation either at an A or G residue separated by 16 or 15 bp from the presumed ORF X start codon (Fig. 2). The initiation sites are preceded by a plausible −10 region TATAAT, which perfectly matches the consensus for the major *B. subtilis* σ^A factor (10). In addition, an AGGATA sequence 17 bp upstream probably represents the −35 region, with three matches out of six with σ^A consensus. Also, the sequence ATATA found at positions −18 to −14 (relative to the transcription initiation site) resembles the less-specific sequence prototype PuTPuTG, proposed for the −18 to −14 positions of *B. subtilis* strong promoters (6, 10). Also, the presence of an AT-rich sequence (87.5 mol%) positioned between −50 and −35 bases indicates that the thermophilic ORF X promoter represents a functional promoter showing all the necessary features to be efficiently utilized by σ^A-RNA polymerase of mesophilic *B. subtilis* (2, 11). Further confirmation of the results obtained from primer extension experiments was done by testing the promoter activities of the ORF X and *npr* genes using *lacZ* transcriptional fusions. The 5′-flanking regions of the *npr* and ORF X genes were inserted upstream of the *lacZ* in the vector pBTK2 to construct *lacZ* fusions. As the integration vector pBTK2 (D. van Sinderen, unpublished data) carries the promoterless *spoVG-lacZ* indicator gene fusion flanked by 5′ and 3′ parts of the *B. subtilis amy* gene, the fusions were stably integrated via double recombination with the *amy* gene into the chromosome of *B. subtilis* DB117. To construct the *npr-lacZ* fusion, the 489-bp *Eco*RI-*Tth*I fragment, including the promoter region and the first 115 bp of the *npr*, was gel purified, filled in with Klenow DNA polymerase, and ligated to linearized plasmid pBTK2. For that purpose, pBTK2 was digested with *Eco*RI and the protruding ends were filled in. To construct an ORF X-*lacZ* fusion, a 458-bp *Ear*I fragment encompassing the regulatory region and the first 65 bp of

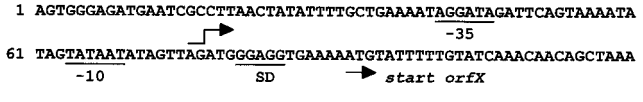

FIG. 2. The promoter region of the ORF X gene. The presented region corresponds to nt 481 to 600 in the deposited ORF X sequence. The ORF X coding region starts at the possible initiation codon ATG as marked by the short arrow under the sequence. The suggested SD and promoter (−10 and −35 regions) sequences are underlined. The black arrow above the sequence denotes the doublet of ORF X transcriptional start sites determined in this study.

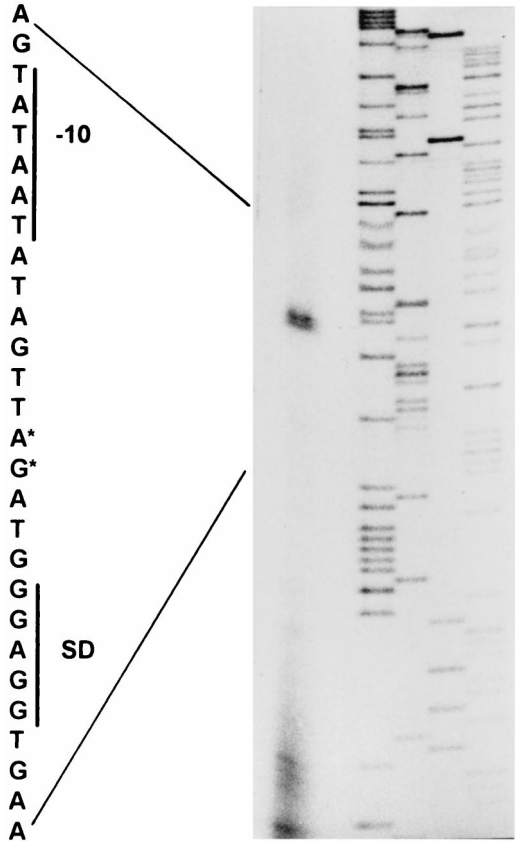

FIG. 3. Identification of the ORF X transcription start sites by primer extension analysis. The primer extension product (the leftmost lane) was analyzed by electrophoresis along with a dideoxy sequencing ladder (in order A, C, G, and T) obtained with the same primer used for primer extension. The sequence of the coding strand is shown with the transcriptional starts (*) and the −10 and SD sequences (vertical lines).

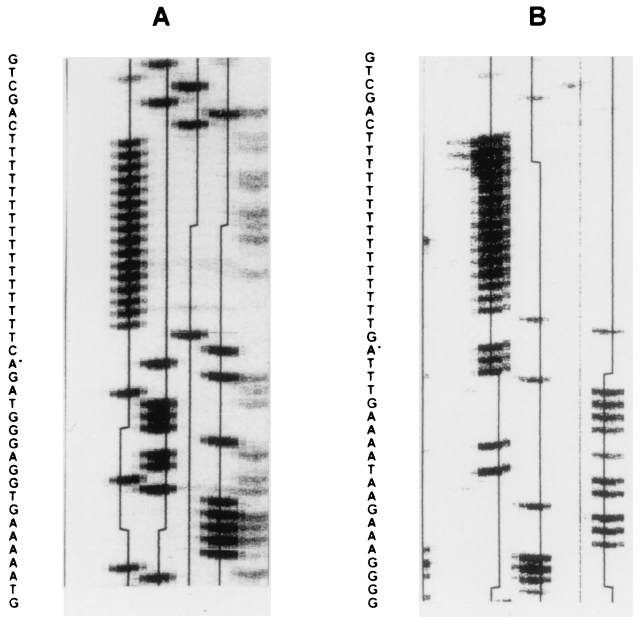

FIG. 4. Sequence analysis of 5′ RACE products. The very 5′ ends of *npr*-specific transcripts were obtained using total RNA from *B. subtilis* DB117(pGBR5) (A) and total RNA from thermophilic *Bacillus* sp. BT1 strain (B). The fluorogram images of the sequence ladder were created using Vistra DNA Sequencer 725 software (Amersham). The sequences displayed are complementary to that determined by automated DNA sequencing; the corresponding transcriptional starts are marked with an asterisk.

ORF X was purified, filled in with Klenow DNA polymerase, and ligated with pBTK2 obtained as described above. The ligation mixes were used to transform competent cells of *Escherichia coli* DH5α (5) to ampicillin and kanamycin resistance. Several clones were screened for plasmid constructs carrying the inserted ORF X or *npr* promoter DNA in the proper orientation to the *lacZ* reporter gene. The correct plasmid constructs were linearized with *Sca*I enzyme and transformed into *B. subtilis* DB117 to allow the single-copy integration of transcriptional fusions into the chromosome (14). Kanamycin-resistant clones were tested for the amylase-deficient phenotype by checking halo formation on agar plates containing 1% starch. Such clones carried a single copy of the specific fusion as verified by Southern hybridization. The expression of the *lacZ* fusions was monitored by assaying β-galactosidase activity of cells during the growth in Luria-Bertani medium at 37°C as described by Miller (9). The ORF X-*lacZ* fusion integrant strain expressed about 130 Miller units (U) of β-galactosidase, and, as expected for the σ^A-dependent promoter, the ORF X-directed β-galactosidase activity was found to be induced during the exponential phase of growth. Apparently, the promoter of the thermophilic gene ORF X displays a transcriptional activity in *B. subtilis*. In contrast, the activity of *npr-lacZ* fusion was only about 2 U of β-galactosidase, which was in fact, in the range of the background activity produced by the strain DB117 (1 to 2 U) (data not shown), indicating that the DNA upstream from *npr* gene did not exhibit promoter activity. This finding is consistent with the negative result of the *npr* promoter mapping by primer extension. The protease production of pGBR50 carrier could be then explained as background expression resulting from nonspecific transcription, probably originating in high-copy vector sequences.

These results are suggestive of an operon structure for the ORF X and *npr* genes. Therefore, this possibility was verified by identification of the very 5′ end of the *npr*-specific mRNA in *B. subtilis* using the RT-PCR and 5′RACE methods with aid of the 5′/3′RACE kit (Boehringer Mannheim). In both cases, total RNA isolated from strain DB117(pGBR5) was used together with the NPREX primer for the synthesis of the first-strand cDNA. The resulting single-stranded DNA was first amplified in a reverse transcription (RT)-PCR experiment utilizing the NPREX primer and ORF X-specific ORFW primer 5′-CGGCATTGAATTCTTGAA-3′, hybridizing with the 1543 to 1560 region of the ORF X sequence. The product of the PCR had the expected size of 450 bp (data not shown), indicating that ORF X and *npr* genes are part of the same operon. In the rapid amplification of cDNA ends (RACE) experiment, the obtained *npr*-specific single-stranded cDNA was amplified according to the manufacturer's protocol. Terminal transferase was used to add a homopolymeric A tail to the 3′ end of the cDNA, and the tailed cDNA was amplified by PCR using the ORF X-specific primer XREV (5′-GCCTTGTGGTCATATT TTC-3′; nt 876 to 858) and the oligo(dT) anchor primer to form the second strand of DNA. The obtained PCR product

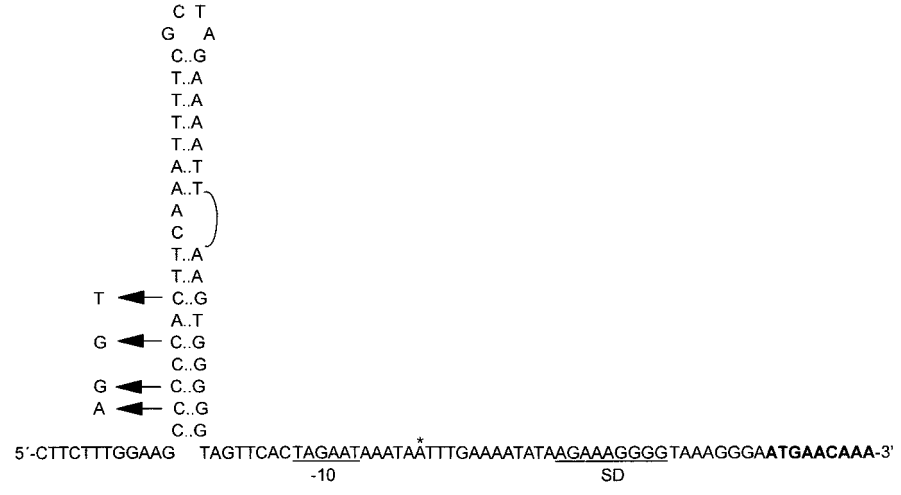

FIG. 5. Hairpin-like structure in the promoter region of the *npr* gene. The transcriptional start is marked with an asterisk; the −10 and SD sequences are underlined. The coding sequence of the *npr* gene is shown in bold letters. The base substitutions introduced by the mutagenesis are shown by arrows and letters.

was further amplified by a second PCR using a nested, ORF X-specific primer, ORFEX, and the PCR anchor primer to increase the product specificity. The specific amplified product of size determined by the primers (136 bp) was isolated and subcloned for sequencing. From the sequence analysis of the RACE product (Fig. 4A) an A residue could be identified as the start of transcription, thus confirming the transcription initiation site of the ORF X gene. This confirmed that the *npr* transcript is initiated at the ORF X promoter in *B. subtilis*. On the basis of this data, the promoter of the *npr* gene is probably not active or not efficiently utilized in *B. subtilis*. However, we are aware of the fact that in the original strain the expression of the *npr* gene might be regulated by a different mechanism. Therefore, we have searched for the *npr* promoter also in the thermophilic strain BT1 using RT-PCR and 5′ RACE methods. The RT-PCR was performed under the same PCR conditions and with the same primers as in the case of RNA preparation from *B. subtilis* (see above); however, as a template the total RNA of the BT1 strain was used. Interestingly, in contrast to the result with RNA from *B. subtilis*, no product was obtained with RNA isolated from the BT1 strain. This indicated that in the thermophilic strain the ORF X gene is not transcribed together with *npr* gene and that the *npr* gene might be expressed from its own promoter. Accordingly, the start point of transcription of the protease gene was determined by 5′ RACE. Total RNA of the BT1 strain and the *npr*-specific primer NPRACE2 (nt 848 to 830) were used for the first-strand cDNA synthesis. Tailed cDNA was then amplified by two following PCR experiments using the nested *npr*-specific primers NPRACE3 (nt 647 to 638) and NPREV (nt 496 to 476). Sequencing of the specific PCR product revealed a transcription start point (Fig. 4B), which identified a −10 region of the *npr* promoter, TAGAAT. These experiments established that in the natural thermophilic host the transcription of *npr* gene is initiated at its own promoter. Furthermore, we have also confirmed that ORF X and *npr* genes do not comprise an operon in BT1 strain. Evidently, the read-through transcription from the ORF X promoter observed in *B. subtilis* does not represent the natural pattern of expression of the *npr* gene. The explanation for inefficiency of the *npr* promoter in *B. subtilis* could come from its own structure. The identified −10 region is located just 8 bp downstream of the hairpin-like structure present in the intergenic region between ORF X and *npr* genes (Fig. 5). Consequently, it is rational to suggest that the putative −35 region is a part of this structure. However, we could not identify any appropriately spaced −35-like sequence within the hairpin. As the best match to the conserved −35 sequence of a σ^A-type promoter, only a TAGAAA hexamer could be found. This region, however, is separated from the −10 region by 20 bp, instead of typical 17 or 18 bp. Presumably, such a promoter requires an additional regulatory factor to be recognized by σ^A-associated RNA polymerase. Nevertheless, the lack of activity of the *npr* promoter in *B. subtilis* might be due to sequestering of the −35 region in the hairpin, which might be absent at 60°C (the cultivation temperature of thermophilic strain BT1), but present at 37°C ($\Delta G = -29.4$ kcal/mol), thus making the promoter inaccessible for RNA polymerase in *B. subtilis*. To test this hypothesis, the symmetry in the GC-rich region creating the hairpin stem was reduced to destabilize the hairpin structure, and, therefore, four of seven C residues involved in the stem were replaced by site-directed mutagenesis (Fig. 5), leaving the region surrounding the −35 position unchanged. However, the destabilization of the hairpin by site-directed mutations did not cause significant improvement of the protease production in *B. subtilis* (data not shown). The identification of the *npr* promoter in the thermophilic BT1 strain might also suggest that this promoter is recognized by a positive transcription factor (activator) or by a novel holoenzyme form associated with a σ factor distinct from those available in *B. subtilis*. The protease genes of thermophilic origin that have been successfully cloned and characterized in *B. subtilis* were presumed to be expressed from specific promoters found immediately 5′ to protease genes. Apparently, our expression study does not prove this assumption, and on the contrary, it clearly demonstrates that transcriptional signals used in the mesophilic and thermophilic bacilli might be different. We assume that this observed transcriptional diversity may also be found with heterologous expression of other thermophilic genes.

Nucleotide sequence accession number. The nucleotide sequence of ORF X has been submitted to the GenBank Data Library under the accession number U23444.

We are grateful to Roy H. Doi for critical reading of the manuscript and helpful discussion. B.V. is very grateful to the members of Department of Genetics in Haren for encouragement throughout this work.

This work was carried out under an East European Fellowship and Short Term Fellowship from EMBO awarded to B.V. and was partially supported by postdoctoral grant C5020702 from the Grant Agency of the Czech Academy of Sciences.

REFERENCES

1. **Chavira, Jr., R., T. J. Burnett, and J. H. Hageman.** 1984. Assaying proteinases with azocoll. Anal. Biochem. **136:**446–450.
2. **Doi, R. H.** 1984. Genetic engineering in *Bacillus subtilis*, p. 121–153. *In* G. E. Russel (ed.), Biotechnology and genetic engineering reviews, vol. 2. Intercept, Newcastle Upon Tyne, United Kingdom.
3. **Eijsink, V. G. H., O. R. Veltman, W. Aukema, G. Vriend, and G. Venema.** 1995. Structural determinants of the stability of thermolysin-like proteinase. Nat. Struct. Biol. **2:**374–379.
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. **Hanahan, D.** 1983. Studies on transformation of *Escherichia coli* with plasmids. J. Mol. Biol. **166:**557–580.
6. **Helmann, J. D.** 1995. Compilation and analysis of *Bacillus subtilis* σ^A-dependent promoter sequences: evidence for extended contact between RNA polymerase and upstream promoter DNA. Nucleic Acids Res. **23:**2351–2360.
7. **Imanaka, T., M. Shibazaki, and M. Takagi.** 1986. A new way of enhancing the thermostability of proteases. Nature **324:**695–697.
8. **Kubo, M., and T. Imanaka.** 1988. Cloning and nucleotide sequence of the highly thermostable neutral protease gene from *B. stearothermophilus*. J. Gen. Microbiol. **134:**1883–1892.
9. **Miller, J. H.** 1972. Experiments in molecular genetics. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
10. **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.
11. **Mountain, A.** 1989. Gene expression systems for *Bacillus subtilis*, p. 73–114. *In* C. R. Harwood (ed.), Biotechnology handbooks, vol. 2. *Bacillus*. Plenum Press, New York, N.Y.
12. **Nishiya, Y., and T. Imanaka.** 1990. Cloning and nucleotide sequences of the *Bacillus stearothermophilus* neutral protease gene and its transcriptional activator gene. J. Bacteriol. **172:**4861–4869.
13. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
14. **Shimotsu, H., and D. Henner.** 1986. Construction of a single copy integration vector and its use in analysis of regulation of the *trp* operon of *Bacillus subtilis*. Gene **33:**103–119.
15. **van den Burg, B., H. G. Enequist, M. E. van der Haar, V. G. H. Eijsink, B. K. Stulp, and G. Venema.** 1991. A highly thermostable neutral protease from *Bacillus caldolyticus*: cloning and expression of the gene in *Bacillus subtilis* and characterization of the gene product. J. Bacteriol. **173:**4107–4115.
16. **van den Burg, B., V. G. H. Eijsink, O. R. Veltman, G. Venema, and G. Vriend.** 1998. Evolution reversed: engineering an enzyme to resist boiling. Proc. Natl. Acad. Sci. USA **95:**2056–2060.
17. **Večerek, B., and P. Kyslík.** 1995. Cloning and sequencing of the neutral protease-encoding gene from a thermophilic strain of *Bacillus* sp. Gene **158:**147–148.