APPLIED AND ENVIRONMENTAL MICROBIOLOGY, Dec. 1995, p. 4490–4493 Vol. 61, No. 12
0099-2240/95/$04.00+0

Alkalophilic *Bacillus* sp. Strain LG12 Has a Series of Serine Protease Genes

BRIAN F. SCHMIDT,[1,2]* LESLIE WOODHOUSE,[1] ROBIN M. ADAMS,[1] TONY WARD,[1] STANLEY E. MAINZER,[1] AND PUSHKARAJ J. LAD[1]

Genencor International, Inc.,[1] and Khepri Pharmaceuticals, Inc.,[2] South San Francisco, California 94080

Received 19 May 1995/Accepted 5 October 1995

Four tandem subtilisin-like protease genes were found on a 6,854-bp DNA fragment cloned from the alkalophilic *Bacillus* sp. strain LG12. The two downstream genes (*sprC* and *sprD*) appear to be transcribed independently, while the two upstream genes (*sprA* and *sprB*) seem to be part of the same transcript.

Bacteria isolated from alkalophilic and/or thermophilic environments may produce subtilisins with the characteristics needed in commercial applications (11, 14, 16, 18, 30, 33). In addition, these enzymes may provide sequence and biochemical information helpful in the design of new subtilisins (5, 8, 13, 17, 27, 34). To this end, we have cloned and sequenced a series of protease genes from *Bacillus* sp. strain LG1Z.

Cloning of the protease genes. An alkalophilic *Bacillus* sp. strain, LG12, was isolated from an alkaline creek bed and grown in Luria-Bertani medium with the pH adjusted to 8.5. This organism was found to produce at least two different alkaline proteases. Purified chromosomal DNA from *Bacillus* sp. strain LG12, digested with restriction enzymes (Boehringer Mannheim), was fractionated on 0.8% agarose gels and Southern blotted (29) on Nytran filters (Schleicher & Schuell). The nick-translated (24) subtilisin BPN′ gene from *Bacillus amyloliquefaciens* (35) was used as a probe. Chromosomal DNA fragments of ca. 1 kilobase pair (kb) from a *Hin*dIII digest were cloned into pBR322 (2) and used to transform *Escherichia coli* MM294 (12), and colony hybridization was used to detect positive clones (9). The nucleotide sequence (25) of a 990-bp DNA insert from a positive clone was homologous to the 3′ portion of the subtilisin BPN′ gene. A ca. 6-kb *Bgl*II fragment was cloned and sequenced in order to obtain the remaining upstream sequences of the protease gene. Three additional subtilisin-like protease genes were discovered on this fragment.

The protease genes. The four open reading frames coding for subtilisin-like serine proteases have been designated *sprA*, *sprB*, *sprC*, and *sprD* (Fig. 1). The upstream portion of the *sprA* gene, coding for presumably the first 70 or so amino acids, is missing on the cloned *Bgl*II fragment. Another downstream open reading frame (ORF) may exist, but there is not enough sequence on the cloned fragment to determine if it codes for an additional protease.

The nucleotide sequence of the cloned region is shown in Fig. 2. Inverted repeat sequences, consistent with known rho-independent terminators (22), were found immediately downstream of the *sprB*, *sprC*, and *sprD* genes. Additional inverted repeats were also found upstream of the *sprC*, *sprD*, and ORF genes and may have regulatory functions. Putative promoter sites, −10 and −35 regions (Fig. 2), were discovered upstream of the *sprC*, *sprD*, and ORF genes by comparison with promoter consensus sequences typical of *B. subtilis* σ^A recognition sequences (19).

Possible translational start sites are shown in Fig. 2 for the *sprB*, *sprC*, *sprD*, and ORF genes. The *sprC* coding region starts with GUG rather than AUG, which is not unusual for *Bacillus subtilis* genes (32). Two potential translational start sites were found for the *sprB* gene (Fig. 2); one 27 nucleotides downstream of the putative *sprA* termination codon with a recognizable ribosome binding site (RBS) (32), and the other nine nucleotides downstream without a discernible RBS (Fig. 2). Since there is such a small distance between the *sprA* and *sprB* genes and there are not obvious transcriptional start or termination signals, it seems likely that they are part of the same transcript. If this is the case, then the two genes might be translationally coupled, avoiding the need for a good RBS (32).

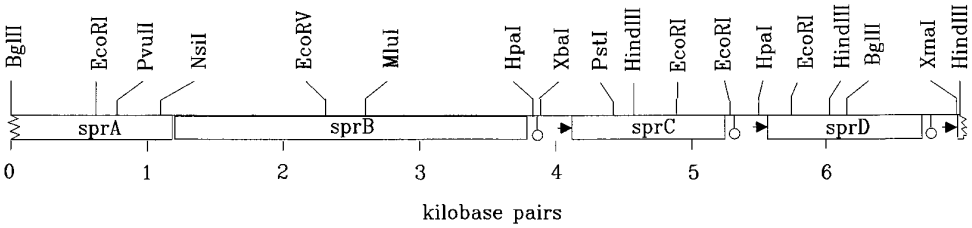

FIG. 1. Restriction endonuclease map of the DNA fragment carrying the four serine protease genes cloned from *Bacillus* sp. strain LG12. ○, putative terminator; ▶, promoter.

Generally, the codon usage for the four protease genes from *Bacillus* sp. strain LG12 (not shown), as in *B. subtilis* (26),

* Corresponding author. Mailing address: Khepri Pharmaceuticals, Inc., 260 Littlefield Ave., South San Francisco, CA 94080. Phone: (415) 794-3500. Fax: (415) 794-3599. Electronic mail address: schmidt@khepri.com.

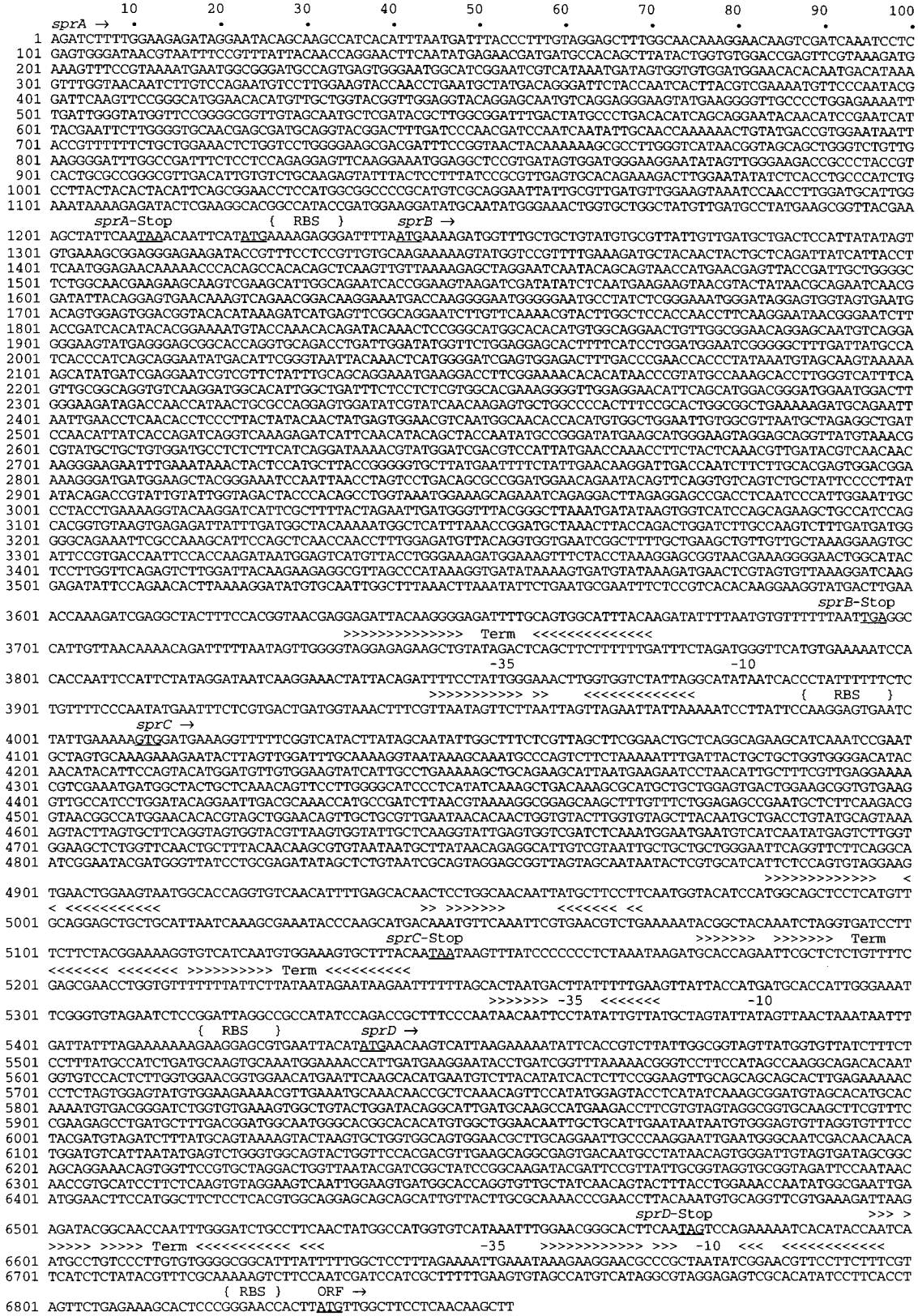

FIG. 2. Nucleotide sequence of the 6,854-bp DNA fragment containing the four *spr* genes (shown schematically in Fig. 1). The putative start and stop codons for the genes are underlined, and potential RBSs are labelled. Putative −10 and −35 promoter regions are indicated. Inverted repeats (> <) that could function as rho-independent terminators are labelled (Term). The beginning of another possible ORF is shown.

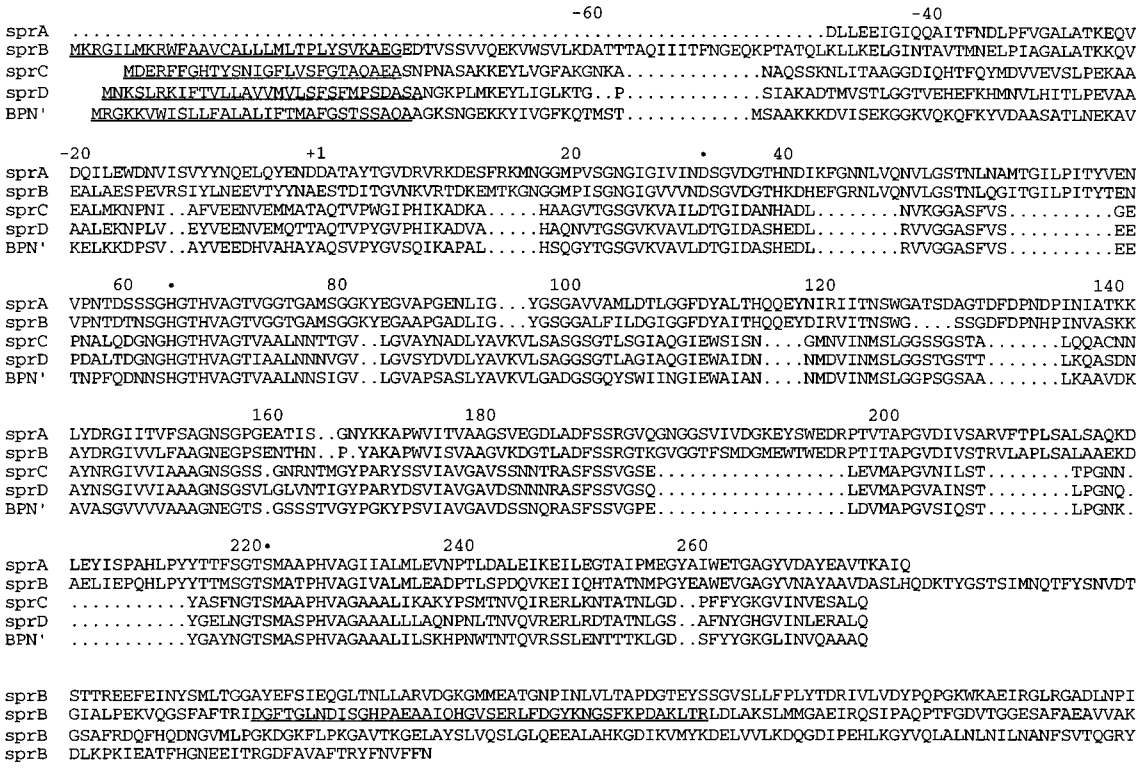

FIG. 3. Computer-generated (pileup program, Genetics Computer Group) amino acid sequence alignment of the Spr proteins, with the subtilisin BPN′ sequence added for comparison. Numbering is based on the BPN′ sequence (35). Putative signal sequences at the beginning of the Spr proteases are underlined. The amino acid sequence in the C-terminal extension of SprB having homology to the S-layer-like repeat sequences in the cellulosome from *C. thermocellum* (7) is also underlined. The catalytic aspartate, histidine, and serine residues are indicated (·).

favors codons with A or T in the third position (an exception is TCC). This reflects the overall G+C content of 44%, which is consistent with a number of *Bacillus* species (23). Unlike in *B. subtilis* and *Escherichia coli*, there does not seem to be a strong preference against ATA as an isoleucine codon. Like the other members of the subtilisin family (4), all four of these genes use TCN as the codon for the active-site serine.

Finding a series of four subtilisin-like protease genes was surprising since in *B. subtilis* the protease genes map to very different loci on the chromosome (21). Tandem gene duplications can arise by homologous recombination of repeated sequences (15), and chromosomal amplifications of antibiotic genes can be induced in *B. subtilis* by increasing the selective pressure (1, 37). However, no obvious repeated sequences, indicative of previous recombination events, were found in the sequenced region.

Protease expression. For protease production, the *sprC* and *sprD* genes were cloned in the replicating vector pBN2, a hybrid plasmid of pUC18 (36) and pUB110 (10), as ca. 1.7-kb *Xba*I-*Hpa*I and ca. 2.5-kb *Hin*dIII-*Xma*I fragments, respectively (Fig. 1). Both the genes were expressed in *B. subtilis* BG2036 (Δ*npr* Δ*apr*) utilizing their own intrinsic transcriptional and translational signals. Protease activity was detected by the formation of clearing zones around colonies on skim milk plates and assayed by using the synthetic substrate *N*-succinyl–L-Ala–L-Ala–L-Pro–L-Phe–*p*-nitroanilide (6) (data not shown).

Protease sequence features. The predicted protein sequences of SprB, SprC, and SprD have typical signal sequences indicative of secreted proteins (20), although the precise signal peptidase cleavage sites have yet to be located (Fig. 3). All four proteases appear to have propeptides (20), with the promature junction in SprC and SprD most likely identical to that of subtilisin BPN′ (35). The SprA and SprB junctions are also likely to be near this region (Fig. 3), but as yet the N-terminal sequences of the mature proteases have not been determined.

The propeptides of SprB, SprC, and SprD are not as highly charged as other subtilisins (17 charged residues for SprD versus 25 for BPN′). It has been suggested that propeptides with a large net negative charge are indicative of subtilisins from alkalophilic *Bacillus* species (31), and although the propeptides of SprB, SprC, and SprD do have predicted net negative charges, the values are not very large (the most negative is −4).

The predicted SprC and SprD mature proteases are closely related to subtilisin BPN′ (Table 1). However, the SprA and SprB proteases are not as closely related to BPN′ and also are not as similar (<40% identity) to other subtilisins, including the minor extracellular proteases, Epr, Bpr, and Vpr, from *B. subtilis* (21). SprB has a long C-terminal extension of over 350 amino acids (Fig. 3) like Epr, Bpr, and Vpr, but there does not appear to be any homology among the sequences of the C-terminal extensions. The function of C-terminal extensions is

TABLE 1. Percent identities among mature protease sequences

Protease	% Identity with protease			
	BPN′	SprA	SprB	SprC
SprA	32			
SprB	40	69		
SprC	65	34	36	
SprD	65	32	37	77

unknown, but in the case of Vpr the extension may function as a membrane anchor (28). SprB may also be associated with the cell surface, since there is some sequence homology (ca. 35% identity in a 43-amino-acid sequence) in the C-terminal extension (Fig. 3) to that of the S-layer-like repeat sequences in the cellulosome from *Clostridium thermocellum* (7).

Besides the catalytic triad (D32, H64, and S221) and oxyanion-binding residue (N155), SprA, SprB, SprC, and SprD have additional amino acids reported to be highly conserved in subtilisins (G23, G34, H39, G65, T66, G70, G83, S125, G127, G146, G154, G219, T220, and P225) (27). The side chains of the calcium binding site, Q2, D41, and N77, are conserved in SprC and SprD. Consistent with the BPN′ structure (27), SprD would be expected to form the salt bridges K136-D140 and R170-E195, and both SprC and SprD have the side chains necessary for a E197-R247 salt bridge.

There are one cysteine in the signal sequence of SprB, one cysteine in the mature sequence of SprC, and no cysteines in SprA or SprD. Thus, as in other bacterial subtilisins (27), there are no internal disulfide bridges in these proteases.

A number of subtilisins from thermophilic and/or alkalophilic bacteria have a deletion of four amino acids relative to BPN′ at the P1 binding site in the region of amino acid 160 (31). However, SprC, SprD, and the alkaline protease from the thermophilic species *Bacillus smithii* (18) do not have this deletion; in fact, SprD has an insertion of one amino acid at this site (Fig. 3). Deletion of these four amino acids has been found to lower the catalytic efficiency of BPN′ (3). The insertions in SprA and SprB, relative to BPN′ (Fig. 3), are consistent with the variable regions noted for other subtilisins (27).

Of the amino acid substitutions known to increase the alkaline activity (change of Y to F at position 104 [Y104F]) or stability (M50F, I107V, and K213R) of BPN′ (5, 34), only F50 is present in SprC and SprD (Fig. 3). Twenty-nine residues have been found to be unique in subtilisins isolated from alkalophilic bacteria compared with proteases from mesophiles (30). Of these residues, only six are conserved in both SprC and SprD (A15, G25, P55, A108, Q109, and R170).

Nucleotide sequence accession number. The GenBank accession number for the DNA sequence identified in this study is U39230.

We thank Sienna Yoast for critical reading of the manuscript.

REFERENCES

1. **Albertini, A. M., and A. Galizzi.** 1985. Amplification of a chromosomal region in *Bacillus subtilis*. J. Bacteriol. **162:**1203–1211.
2. **Bolivar, F., R. L. Rodriguez, P. J. Greene, M. C. Betlach, H. L. Heyneker, and H. W. Boyer.** 1977. Construction and characterization of new cloning vehicles. II. A multipurpose cloning system. Gene **2:**95–113.
3. **Bott, R., J. Dauberman, R. Caldwell, C. Mitchinson, L. Wilson, B. Schmidt, C. Simpson, S. Power, P. Lad, I. H. Sagar, T. Graycar, and D. Estell.** 1992. Using structural comparison as a guide in protein engineering. Ann. N.Y. Acad. Sci. **672:**10–19.
4. **Brenner, S.** 1988. The molecular evolution of genes and proteins: a tale of two serines. Nature (London) **334:**528–530.
5. **Cunningham, B. C., and J. A. Wells.** 1987. Improvement in the alkaline stability of subtilisin using an efficient random mutagenesis and screening procedure. Protein Eng. **1:**319–325.
6. **Estell, D. A., T. P. Graycar, and J. A. Wells.** 1985. Engineering an enzyme by site-directed mutagenesis to be resistant to chemical oxidation. J. Biol. Chem. **260:**6518–6521.
7. **Fujino, T., P. Béguin, and J.-P. Aubert.** 1993. Organization of a *Clostridium thermocellum* gene cluster encoding the cellulosomal scaffolding protein CipA and a protein possibly involved in attachment of the cellulosome to the cell surface. J. Bacteriol. **175:**1891–1899.
8. **Graycar, T. P.** 1991. Protein engineering of subtilisin, p. 257–284. *In* J. S. Dordick (ed.), Biocatalysts for industry. Plenum Press, New York.
9. **Grunstein, M., and D. Hogness.** 1975. Colony hybridization: a method for the isolation of cloned DNAs that contain a specific gene. Proc. Natl. Acad. Sci. USA **72:**3961–3965.
10. **Gryczan, T. J., S. Contente, and D. Dubnau.** 1978. Characterization of *Staphylococcus aureus* plasmids introduced by transformation into *Bacillus subtilis*. J. Bacteriol. **134:**318–329.
11. **Hakamada, Y., T. Kobayashi, J. Hitomi, S. Kawai, and S. Ito.** 1994. Molecular cloning and nucleotide sequence of the gene for an alkaline protease from the alkalophilic *Bacillus* sp. KSM-K16. J. Ferment. Bioeng. **78:**105–108.
12. **Hanahan, D.** 1983. Studies on transformation of *Escherichia coli* with plasmids. J. Mol. Biol. **166:**557–580.
13. **Heringa, J., P. Argos, M. R. Egmond, and J. de Vlieg.** 1995. Increasing thermal stability of subtilisin from mutations suggested by strongly interacting side-chain clusters. Protein Eng. **8:**21–30.
14. **Jarnagin, A. S., and E. Ferrari.** 1992. Extracellular enzymes: gene regulation and structure function relationship studies, p. 191–219. *In* R. Doi and M. McGloughlin (ed.), Biology of bacilli: applications to industry. Butterworth-Heinemann, Boston.
15. **Krawiec, S., and M. Riley.** 1990. Organization of the bacterial chromosome. Microbiol. Rev. **54:**502–539.
16. **Maciver, B., R. H. McHale, D. J. Saul, and P. L. Bergquist.** 1994. Cloning and sequencing of a serine proteinase gene from a thermophilic *Bacillus* species and its expression in *Escherichia coli*. Appl. Environ. Microbiol. **60:**3981–3988.
17. **Masui, A., N. Fujiwara, and T. Imanaka.** 1994. Stabilization and rational design of serine protease AprM under high alkaline and high-temperature conditions. Appl. Environ. Microbiol. **60:**3579–3584.
18. **Milano, A., P. L. Manachini, C. Parini, and G. Riccardi.** 1994. Sequence of the gene encoding an alkaline serine protease of thermophilic *Bacillus smithii*. Gene **145:**149–150.
19. **Moran, C. P., Jr.** 1993. RNA polymerase and transcription factors, p. 653–668. *In* A. L. Sonenshein, J. A. Hoch, and R. Losick (ed.), *Bacillus subtilis* and other gram-positive bacteria: biochemistry, physiology, and molecular genetics. American Society for Microbiology, Washington, D.C.
20. **Nagarajan, V.** 1993. Protein secretion, p. 713–726. *In* A. L. Sonenshein, J. A. Hoch, and R. Losick (ed.), *Bacillus subtilis* and other gram-positive bacteria: biochemistry, physiology, and molecular genetics. American Society for Microbiology, Washington, D.C.
21. **Pero, J., and A. Sloma.** 1993. Proteases, p. 939–952. *In* A. L. Sonenshein, J. A. Hoch, and R. Losick (ed.), *Bacillus subtilis* and other gram-positive bacteria: biochemistry, physiology, and molecular genetics. American Society for Microbiology, Washington, D.C.
22. **Platt, T.** 1986. Transcription termination and the regulation of gene expression. Annu. Rev. Biochem. **55:**339–372.
23. **Priest, F. G.** 1993. Systematics and ecology of *Bacillus*, p. 3–16. *In* A. L. Sonenshein, J. A. Hoch, and R. Losick (ed.), *Bacillus subtilis* and other gram-positive bacteria: biochemistry, physiology, and molecular genetics. American Society for Microbiology, Washington, D.C.
24. **Rigby, P. W. J., M. Dieckmann, C. Rhodes, and P. Berg.** 1977. Labelling deoxyribonucleic acid to high specific activity *in vitro* by nick translation with DNA polymerase I. J. Mol. Biol. **113:**237–251.
25. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74:**5463–5467.
26. **Shields, D. C., and P. M. Sharp.** 1987. Synonymous codon usage in *Bacillus subtilis* reflects both translational selection and mutational biases. Nucleic Acids Res. **15:**8024–8040.
27. **Siezen, R. J., W. M. de Vos, J. A. M. Leunissen, and B. W. Dijkstra.** 1991. Homology modeling and protein engineering strategy of subtilases, the family of subtilisin-like serine proteinases. Protein Eng. **4:**719–737.
28. **Sloma, A., G. A. Rufo, Jr., K. A. Theriault, M. Dwyer, S. W. 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.
29. **Southern, E.** 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. **98:**503–517.
30. **Takami, H., T. Kobayashi, R. Aono, and K. Horikoshi.** 1992. Molecular cloning, nucleotide sequence and expression of the structural gene for a thermostable alkaline protease from *Bacillus* sp. no. AH-101. Appl. Microbiol. Biotechnol. **38:**101–108.
31. **van der Laan, J. C., G. Gerritse, L. J. S. M. Mulleners, R. A. C. van der Hoek, and W. J. Quax.** 1991. Cloning, characterization, and multiple chromosomal integration of a *Bacillus* alkaline protease gene. Appl. Environ. Microbiol. **57:**901–909.
32. **Vellanoweth, R. L.** 1993. Translation and its regulation, p. 699–712. *In* A. L. Sonenshein, J. A. Hoch, and R. Losick (ed.), *Bacillus subtilis* and other gram-positive bacteria: biochemistry, physiology, and molecular genetics. American Society for Microbiology, Washington, D.C.
33. **Völkl, P., P. Markiewicz, K. O. Stetter, and J. H. Miller.** 1994. The sequence of a subtilisin-type protease (aerolysin) from the hyperthermophilic archaeum *Pyrobaculum aerophilum* reveals sites important to thermostability. Protein Sci. **3:**1329–1340.
34. **Wells, J. A., and D. A. Estell.** 1988. Subtilisin—an enzyme designed to be engineered. Trends Biochem. Sci. **13:**291–297.
35. **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.
36. **Yanisch-Perron, C., J. Vieira, and J. Messing.** 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene **33:**103–119.
37. **Young, M.** 1984. Gene amplification in *Bacillus subtilis*. J. Gen. Microbiol. **130:**1613–1621.