JOURNAL OF BACTERIOLOGY, Dec. 1988, p. 5557–5563
0021-9193/88/125557-07$02.00/0

Vol. 170, No. 12

Gene Encoding a Minor Extracellular Protease in *Bacillus subtilis*

ALAN SLOMA,* ABDUL ALLY, DELPHINE ALLY, AND JANICE PERO

BioTechnica International, Inc., 85 Bolton Street, Cambridge, Massachusetts 02140

Received 27 June 1988/Accepted 12 September 1988

The gene for a minor, extracellular protease has been identified in *Bacillus subtilis*. The gene (*epr*) encoded a primary product of 645 amino acids that was partially homologous to both subtilisin (Apr) and the major internal serine protease (ISP-1) of *B. subtilis*. Deletion analysis indicated that the C-terminal 240 amino acids of Epr were not necessary for activity. This C-terminal region exhibited several unusual features, including a high abundance of lysine residues and the presence of a partially homologous sequence of 44 amino acids that was directly repeated five times. The *epr* gene mapped near *sacA* and was not required for growth or sporulation.

The gram-positive, sporeforming bacterium *Bacillus subtilis* produces and secretes proteases, esterases, and other kinds of exoenzymes at the end of the exponential phase of growth (15). The principal extracellular proteolytic enzymes, the alkaline (subtilisin) and neutral (metallo-) proteases, are encoded by the *apr* and *npr* genes, respectively (19, 21, 23). Double mutants bearing null mutations in both genes are substantially blocked in extracellular protease production but still secrete 2 to 4% of the wild-type level of protease (10; our unpublished results). We have been interested in identifying additional protease genes responsible for the residual extracellular protease production by Apr^-, Npr^- double mutants. Here we report the cloning of a new protease gene whose product, a protease sensitive to phenylmethylsulfonyl fluoride (PMSF) and EDTA, shows significant homology to subtilisin.

MATERIALS AND METHODS

Bacterial strains and plasmids. *B. subtilis* GP208 (Δ*aprA* Δ*nprE* Δ*ispA metC amyE*) is a derivative of the *B. subtilis* 168 strain IS75. Plasmids pBD64 or pUC19 and pBR322 were used for cloning in *B. subtilis* and *Escherichia coli*, respectively. The *cat* gene was obtained from plasmid pMI1101 (24). *B. subtilis* strains were grown on tryptose blood agar base (Difco Laboratories) or minimal glucose medium and were made competent by the procedure of Anagnostopoulos and Spizizen (1). *E. coli* JM107 was grown and made competent by the procedure of Hanahan (7). Plasmid DNA from *B. subtilis* and *E. coli* was prepared by the alkaline lysis method of Birnboim and Doly (2). Plasmid DNA transformation in *B. subtilis* was performed as described by Gryczan et al. (6).

Enzymes. Restriction enzymes, T4 DNA ligase, *Bal*31 exonuclease, and calf intestinal alkaline phosphatase were obtained from Boehringer Mannheim.

DNA isolation and gene library characterization. Isolation of *B. subtilis* chromosomal DNA was done as described previously (4). Total DNA from GP208 (100 μg) was digested with *Sau*3A (4 U) at 37°C for 10 min. The partially digested DNA was run on an 0.8% agarose gel. Fragments of 3 to 7 kilobases (kb) were electroeluted and ligated to *Bgl*II-digested pEc224 that had been treated with alkaline phosphatase (1 U) for 30 min at 37°C. The ligation was done at a ratio of insert to vector DNA of 4:1. The ligation mix was incubated at 16°C overnight and transformed into JM107.

Restriction fragments to be sequenced were ligated into appropriate sites of M13 vectors. DNA sequencing was performed by the dideoxy chain termination method (17).

Mapping of *epr* gene. Mapping of the *epr* locus was performed by PBS1 transduction (9) with a lysate from *B. subtilis* GP208(pNP7 integrated) Cm^r. Cm^r transductants were scored for linkage to several loci from the Dedonder set of reference strains (3).

Protease assays. Protease assays were performed by adding 100 μl of culture supernatant to 900 μl of 50 mM Tris (pH 8)–5 mM $CaCl_2$, with 10 mg of Azocoll (Sigma). The solutions were incubated at 37°C for 30 min with constant shaking. The reactions were then centrifuged to remove the insoluble Azocoll, and the A_{520} of the solution was determined. Inhibitors were preincubated with the reaction mix for 5 min at 37°C.

[^{14}C]casein was used to measure low amounts of residual protease. Culture supernatant (100 μl) was added to 100 μl of 50 mM Tris (pH 8)–5 mM $CaCl_2$ containing 10^5 cpm of [^{14}C]casein (New England Nuclear). The solutions were incubated at 37°C for 30 min. The reaction mixes were then placed on ice, and 20 μg of bovine serum albumin was added as carrier protein. Cold 10% trichloroacetic acid (600 μl) was added, and the mix was kept on ice for 10 min. The solutions were centrifuged to spin out the precipitated protein, and the radioactivity in the supernatants was counted in a scintillation counter. Up to 20 to 30% degradation remained in the linear range of the assay.

RESULTS

Cloning an additional protease gene from *B. subtilis*. To identify additional extracellular protease-encoding genes, we used a high-copy-number vector to construct a library of cloned *B. subtilis* DNAs from a triple mutant that contained deletions of the genes for the three known proteases: subtilisin (*apr*), neutral protease (*npr*), and the major intracellular serine protease (*isp*). We reasoned that amplification of the additional protease gene(s) on a high-copy-number vector would increase protease levels sufficiently to produce a halo around colonies containing a cloned protease gene.

* Corresponding author.

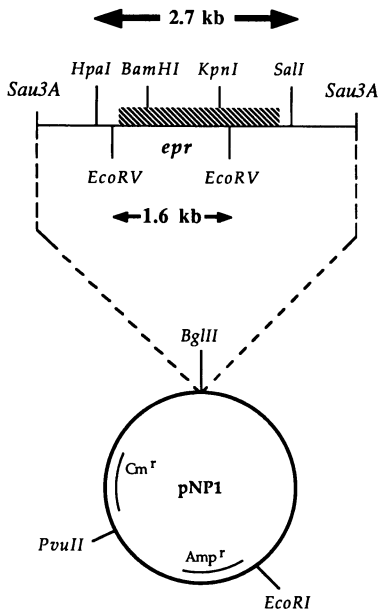

FIG. 1. Restriction map of pNP1 containing the *epr* gene cloned in pEc224.

To construct a gene bank, total chromosomal DNA was isolated from strain GP208 (Δ*apr* Δ*npr* Δ*isp*) and partially digested with *Sau*3A. Size-fractionated fragments (3 to 7 kb) were cloned into the *Bgl*II site of pEc224, a shuttle vector made from pBR322 and pBD64 joined at the *Eco*RI and *Pvu*II sites. Transformation of competent cells of *E. coli* JM107 resulted in more than 1,200 colonies. Plasmid screening showed that 90% of the colonies contained plasmids with inserts of an average size of about 4 kb; however, none of the *E. coli* transformants produced halos on casein plates.

To screen for protease activity in *B. subtilis*, the 1,200 *E. coli* colonies were pooled into groups of 100 (G1 to G12). The groups were grown in liquid culture (LB and ampicillin [50 μg/ml]), and miniprep plasmid DNA prepared from each group was used to transform *B. subtilis* GP208. Approximately 5% of the transformants from one of the pools (G11) produced halos on casein plates.

Cleavage of the plasmid DNA with restriction enzymes showed that all the transformants producing halos had a plasmid with the same-sized DNA insert (about 4 kb). A restriction map of one such plasmid, pNP1, is shown in Fig. 1. The protease secreted by GP208 cells containing pNP1 was sensitive to both PMSF (1 mM) and EDTA (10 mM).

Nucleotide and deduced amino acid sequences of the protease gene. Subcloning and deletion experiments established that all or most of the protease gene was contained on the 1.6-kb *Eco*RV fragment (Fig. 1). Determination of the nucleotide sequence of the 1.6-kb *Eco*RV fragment revealed an open reading frame (ORF) which covered almost the entire fragment, starting 450 base pairs (bp) from the left end and proceeding through the right end (Fig. 2). Comparison of the deduced amino acid sequence with other amino acid sequences in GENBANK indicated that the protein encoded by the ORF had strong homology (approximately 40%) to both subtilisin (19) and ISP-1 (11) from *B. subtilis* 168 (Fig. 3A and B). The most probable initiation codon for this protease gene is the ATG at position 1 in Fig. 2. This ATG (second codon in the ORF) was preceded by an excellent *B. subtilis* ribosome-binding site (AAAGGAGATGA), which had a calculated ΔG of −16.4 kcal (20). In addition, the first 26 amino acids following this Met resembled a typical *B. subtilis* signal sequence, with a short sequence containing two positively charged amino acids, followed by 15 hydrophobic amino acids, a helix-breaking proline, and a typical Ala-X-Ala signal peptidase cleavage site (13). Even though the 1.6-kb *Eco*RV fragment was sufficient for protease activity, the ORF continued past the end of the downstream *Eco*RV site. To locate the 3′ end of the gene, the DNA sequence of the overlapping *Kpn*I-*Sal*I fragment was determined (Fig. 1). As shown in Fig. 2, the end of the ORF was found 717 bp downstream from the *Eco*RV site, and the entire *epr* gene was found to encode a 645-amino-acid protein, of which approximately the first 380 amino acids were homologous to subtilisin (Fig. 3). The final approximately 240 amino acids were apparently not essential for proteolytic activity, since subcloning experiments with the 1.6-kb *Eco*RV fragment indicated that the information encoded in the first 405 amino acids was sufficient for protease activity.

Structure of the Epr protein. In vitro transcription-translation experiments were used to confirm the size of the gene. Addition of *B. subtilis* or *E. coli* plasmid DNA containing the 2.7-kb *Hpa*I-*Sal*I fragment with the entire *epr* gene (pNP3 or pNP2) to an S30-coupled transcription-translation system (New England Nuclear) resulted in the synthesis of a protein of approximately 75,000 daltons (Fig. 4A). (Additional proteins of 60,000 and 34,000 daltons were also seen and presumably represented processed or degraded forms of the 75,000-dalton protein.) This size was in reasonable agreement with the predicted molecular weight for the primary product of 69,702.

The homology of the *epr* protease to subtilisin throughout the amino-terminal half of Epr suggests that Epr might also be produced as a preproenzyme with a prosequence of similar size to that found for subtilisin (70 to 80 amino acids). If true, and if there were no additional processing, this would argue that the mature Epr enzyme has a molecular weight of ca. 58,000. Examination of culture supernatants, however, indicated that the protein had a molecular weight of about 34,000. Comparison by electrophoresis on sodium dodecyl sulfate (SDS)-polyacrylamide gels of the proteins secreted by *B. subtilis* strain GP208 containing a plasmid with the *epr* gene (pNP3 or pNP5) or just the parent plasmid alone (pBs81/6) showed that the 2.7-kb *Hpa*I-*Sal*I fragment (Fig. 1) cloned in pNP3 directed the production of proteins of about 34,000 and 38,000 daltons (Fig. 4B), whereas the 1.6-kb *Eco*RV fragment cloned in pNP5 in the same orientation (Fig. 1) directed production of just the 34,000-dalton protein (data not shown). The two proteins appeared to be different forms of the Epr protease, resulting from either processing or proteolytic degradation. Clearly, the 1.6-kb *Eco*RV fragment, which lacks the 3′ third of the *epr* gene, is capable of directing the production of an active protease similar in size to that observed when the entire gene is present. This suggests that the protease normally undergoes C-terminal processing.

Location of *epr* on the *B. subtilis* chromosome. To map *epr* on the *B. subtilis* chromosome, we introduced a drug resistance marker into the chromosome at the site of the *epr* gene and used phage PBS1-mediated transduction to determine the location of the insertion. A 1.3-kb *Eco*RI fragment containing a chloramphenicol acetyltransferase (*cat*) gene was cloned into the unique *Eco*RI site on an *E. coli* plasmid containing the *epr* gene (pNP2, Fig. 5). The resulting plasmid (pNP7) was used to transform *B. subtilis* GP208, and chloramphenicol-resistant transformants were selected. Since the

```
-446 ATCCGAGCTTATCGGCCCACTCGTTCCCAAACACACTCGCCATGAAATCAGCATACCCC
     GGAATCGGCAAGCTCGTTAAAATCAAGAAGACAGACCCGATAATAATCAGCGGCATGGT
     CAGGATAATCCGTCACGCAAAGCGCTGAGATGCCGCTGCCCGGCAATTTTCCCGGCGAC
     AGGCATTATTTTTTCCTCCATCACCCGAGTGAATGTGCTCATCTTAAAAACCCCCTTTT
     CTCATTGCTTTGTGAACAACCTCCGCAATGTTTTCTTTATCTTATTTTGAAAACGCTTA
     CAAATTCATTTGGAAAATTTCCTCTTCATGCGGAAAAAATCTGCATTTTGCTAAACAAC
     CCTGCCCATGAAAAATTTTTTCCTTCTTACTATTAATCTCTCTTTTTTTCTCCGATATA
     TATATCAAACATCATAGAAAAAGGAGATGAATC

+1   ATG AAA AAC ATG TCT TGC AAA CTT GTT GTA TCA GTC ACT CTG TTT
     met lys asn met ser cys lys leu val val ser val thr leu phe

46   TTC AGT TTT CTC ACC ATA GGC CCT CTC GCT CAT GCG CAA AAC AGC
     phe ser phe leu thr ile gly pro leu ala his ala gln asn ser

91   AGC GAG AAA GAG GTT ATT GTG GTT TAT AAA AAC AAG GCC GGA AAG
     ser glu lys glu val ile val val tyr lys asn lys ala gly lys

136  GAA ACC ATC CTG GAC AGT GAT GCT GAT GTT GAA CAG CAG TAT AAG
     glu thr ile leu asp ser asp ala asp val glu gln gln tyr lys

181  CAT CTT CCC GCG GTA GCG GTC ACA GCA GAC CAG GAG ACA GTA AAA
     his leu pro ala val ala val thr ala asp gln glu thr val lys
                BamHI
226  GAA TTA AAG CAG GAT CCT GAT ATT TTG TAT GTA GAA AAC AAC GTA
     glu leu lys gln asp pro asp ile leu tyr val glu asn asn val

271  TCA TTT ACC GCA GCA GAC AGC ACG GAT TTC AAA GTG CTG TCA GAC
     ser phe thr ala ala asp ser thr asp phe lys val leu ser asp

316  GGC ACT GAC ACC TCT GAC AAC TTT GAG CAA TGG AAC CTT GAG CCC
     gly thr asp thr ser asp asn phe glu gln trp asn leu glu pro

361  ATT CAG GTG AAA CAG GCT TGG AAG GCA GGA CTG ACA GGA AAA AAT
     ile gln val lys gln ala trp lys ala gly leu thr gly lys asn

406  ATC AAA ATT GCC GTC ATT GAC AGC GGG ATC TCC CCC CAC GAT GAC
     ile lys ile ala val ile asp ser gly ile ser pro his asp asp

451  CTG TCG ATT GCC GGC GGG TAT TCA GCT GTC AGT TAT ACC TCT TCT
     leu ser ile ala gly gly tyr ser ala val ser tyr thr ser ser

496  TAC AAA GAT GAT AAC GGC CAC GGA ACA CAT GTC GCA GGG ATT ATC
     tyr lys asp asp asn gly his gly thr his val ala gly ile ile

541  GGA GCC AAG CAT AAC GGC TAC GGA ATT GAC GGC ATC GCA CCG GAA
     gly ala lys his asn gly tyr gly ile asp gly ile ala pro glu

586  GCA CAA ATA TAC GCG GTT AAA GCG CTT GAT CAG AAC GGC TCG GGG
     ala gln ile tyr ala val lys ala leu asp gln asn gly ser gly

631  GAT CTT CAA AGT CTT CTC CAA GGA ATT GAC TGG TCG ATC GCA AAC
     asp leu gln ser leu leu gln gly ile asp trp ser ile ala asn

676  AGG ATG GAC ATC GTC AAT ATG AGC CTT GGC ACG ACG TCA GAC AGC
     arg met asp ile val asn met ser leu gly thr thr ser asp ser

721  AAA ATC CTT CAT GAC GCC GTG AAC AAA GCA TAT GAA CAA GGT GTT
     lys ile leu his asp ala val asn lys ala tyr glu gln gly val

766  CTG CTT GTT GCC GCA AGC GGT AAC GAC GGA AAC GGC AAG CCA GTG
     leu leu val ala ala ser gly asn asp gly asn gly lys pro val

811  AAT TAT CCG GCG GCA TAC AGC AGT GTC GTT GCG GTT TCA GCA ACA
     asn tyr pro ala ala tyr ser ser val val ala val ser ala thr

856  AAC GAA AAG AAT CAG CTT GCC TCC TTT TCA ACA ACT GGA GAT GAA
     asn glu lys asn gln leu ala ser phe ser thr thr gly asp glu

901  GTT GAA TTT TCA GCA CCG GGG ACA AAC ATC ACA AGC ACT TAC TTA
     val glu phe ser ala pro gly thr asn ile thr ser thr tyr leu

946  AAC CAG TAT TAT GCA ACG GGA AGC GGA ACA TCC CAA GCG ACA CCG
     asn gln tyr tyr ala thr gly ser gly thr ser gln ala thr pro

991  CAC GCC GCT GCC ATG TTT GCC TTG TTA AAA CAG CGT GAT CCT GCC
     his ala ala ala met phe ala leu leu lys gln arg asp pro ala

1036 GAG ACA AAC GTC CAG CTT CGC GAG GAA ATG CGG AAA AAC ATC GTT
     glu thr asn val gln leu arg glu glu met arg lys asn ile val
               KpnI
1081 GAT CTT GGT ACC GCA GGC CGC GAT CAG CAA TTT GGC TAC GGC TTA
     asp leu gly thr ala gly arg asp gln gln phe gly tyr gly leu

1121 ATC CAG TAT AAA GCA CAG GCA ACA GAT TCA GCG TAC GCG GCA[GCA
     ile gln tyr lys ala gln ala thr asp ser ala tyr ala ala[ala

1171 GAG CAA GCG GTG AAA AAA GCG GAA CAA ACA AAA GCA CAA ATC GAT
     glu gln ala val lys lys ala glu gln thr lys ala gln ile asp
          EcoRV
1216 ATC AAC AAA GCG CGA GAA CTC ATC AGC CAG CTG CCG AAC TCC GAC
     ile asn lys ala arg glu leu ile ser gln leu pro asn ser asp

1261 GCC AAA ACT GCC CTG CAC AAA AGA CTG GAT AAA GTA CAG TCA TAC
     ala lys thr ala leu his lys arg leu asp lys val gln ser tyr

1306 AGA AAT GTA AAA]GAT[GCG AAA GAC AAA GTC GCA AAG GCA GAA AAA
     arg asn val lys]asp[ala lys asp lys val ala lys ala glu lys

1351 TAT AAA ACA CAG CAA ACC GTT GAC ACA GCA CAA ACT GCC ATC AAC
     tyr lys thr gln gln thr val asp thr ala gln thr ala ile asn

1396 AAG CTG CCA AAC GGA ACA GAC AAA AAG AAC CTT CAA AAA CGC TTA
     lys leu pro asn gly thr asp lys lys asn leu gln lys arg leu

1441 GAC CAA GTA AAA]CGA TAC ATC GCG TCA AAG CAA[GCG AAA GAC AAA
     asp gln val lys]arg tyr ile ala ser lys gln[ala lys asp lys

1486 GTT GCG AAA GCG GAA AAA AGC AAA AAG AAA ACA GAT GTG GAC AGC
     val ala lys ala glu lys ser lys lys lys thr asp val asp ser

1531 GCA CAA TCA GCA ATT GGC AAG CTG CCT GCA AGT TCA GAA AAA ACG
     ala gln ser ala ile gly lys leu pro ala ser ser glu lys thr
          PstI
1576 TCC CTG CAG AAA CGC CTT AAC AAA GTG AAG]AGC ACC AAT TTG AAG
     ser leu gln lys arg leu asn lys val lys]ser thr asn leu lys

1621 ACG[GCA CAG CAA TCC GTA TCT GCG GCT GAA AAG AAA TCA ACT GAT
     thr[ala gln gln ser val ser ala ala glu lys lys ser thr asp

1666 GCA AAT GCG GCA AAA GCA CAA TCA GCC GTC AAT CAG CTT CAA GCA
     ala asn ala ala lys ala gln ser ala val asn gln leu gln ala

1711 GGC AAG GAC AAA ACG GCA TTG CAA AAA CGG TTA GAC AAA GTG AAG]
     gly lys asp lys thr ala leu gln lys arg leu asp lys val lys]

1756 AAA AAG GTG GCG GCG GCT GAA GCA AAA AAA GTG GAA ACT[GCA AAG
     lys lys val ala ala ala glu ala lys lys val glu thr[ala lys

1801 GCA AAA GTG AAG AAA GCG GAA AAA GAC AAA ACA AAG AAA TCA AAG
     ala lys val lys lys ala glu lys asp lys thr lys lys ser lys
               PstI
1846 ACA TCC GCT CAG TCT GCA GTG AAT CAA TTA AAA GCA TCC AAT GAA
     thr ser ala gln ser ala val asn gln leu lys ala ser asn glu

1891 AAA ACA AAG CTG CAA AAA CGG CTG AAC GCC GTC AAA]CCG AAA AAG
     lys thr lys leu gln lys arg leu asn ala val lys]pro lys lys

1936 TAA CCAAAAACCTTTAAGATTTGCATTCCAAGTCTTAAAGGTTTTTTTCATTCTAAGA
     ...

1994 ACACCACACACAACCTTTTTCCCATCCATTGT
```

FIG. 2. Nucleotide and deduced amino acid sequences of the *epr* gene. Nucleotides are numbered starting with the A of the presumed initiation codon ATG. The putative ribosome-binding site and a putative transcription terminator are underlined. The partially homologous, directly repeated sequences in the C-terminal domain are indicated in brackets. The translational stop codon (· · ·) is indicated.

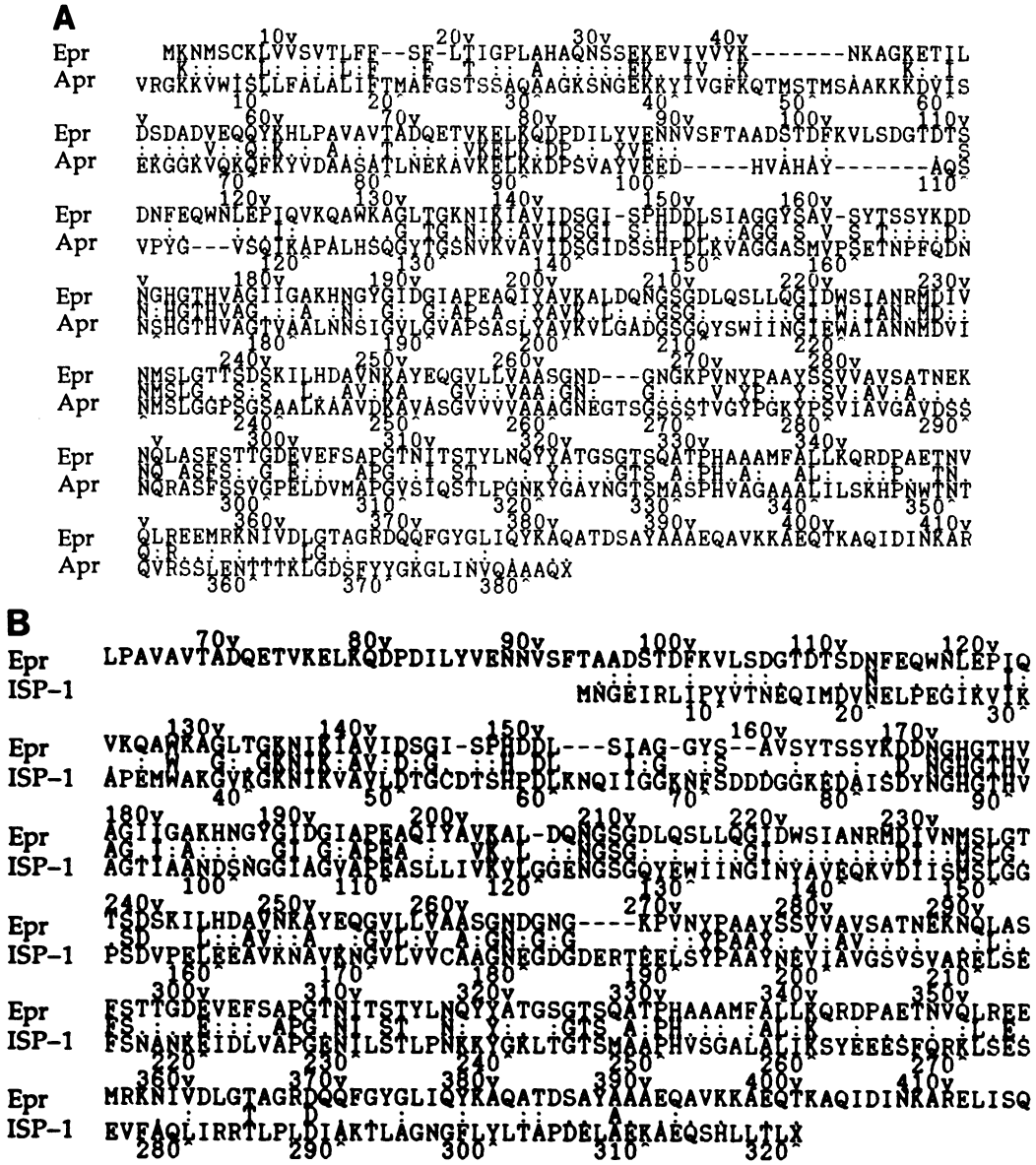

FIG. 3. (A) Comparison of the deduced amino acid sequence of the protease encoded by *epr* and that of subtilisin (*apr* gene product). Identical amino acids are indicated by the appropriate single-letter code; similar amino acids are designated with a colon. Unrelated amino acids are designated with a dot or blank. The final 235 amino acids of Epr are not shown. (B) Comparison of homology between the deduced amino acid sequences of *epr* and ISP-1. The first 61 and final 230 amino acids of Epr are not shown.

plasmid cannot replicate autonomously in *B. subtilis*, the Cm^r transformants were expected to arise by virtue of a single, reciprocal recombination event between the cloned *epr* gene on the plasmid and the chromosomal copy of the gene. Southern hybridization confirmed that the *cat* gene had integrated in the chromosome at the site of the cloned *epr* gene. Mapping experiments indicated that the inserted *cat* gene and *epr* gene were tightly linked to *sacA321* (77% cotransduction), weakly linked to *purA16* (5% cotransduction), and unlinked to *hisA1*. These findings suggest that the *epr* gene is situated near *sacA* in an area of the genetic map not known to contain any other protease genes.

***epr* is a nonessential gene.** To examine the function of the *epr* gene, we constructed a deletion in the 5′ end of the cloned gene and then replaced the wild-type gene in the chromosome with the in vitro-created deletion mutation. A deletion mutation was created in *epr* by cleavage at the unique *Bam*HI site in pNP2 and digestion with *Bal*31 (Fig. 5). After religation and transformation, one plasmid (pNP8) was selected which contained a 2.3-kb insert instead of the original 2.7-kb insert. (As predicted, deletion of the 400 bp of *epr* generated a fragment that, when cloned on a *Bacillus* plasmid, no longer produced an active protease, as measured by the appearance of a halo on casein plates.)

To introduce the deleted gene into the *B. subtilis* chromosome, the *cat* gene described above was introduced into the *Eco*RI site on pNP8 to create pNP9. This *E. coli* plasmid was used to transform *B. subtilis* GP208 to Cm^r, and 70% of the resulting transformants had a very slight halo and 30% had no halos on casein plates. The absence of a halo is the result of a double crossover event between homologous chromosomal DNA and a concatemer of the plasmid DNA (19);

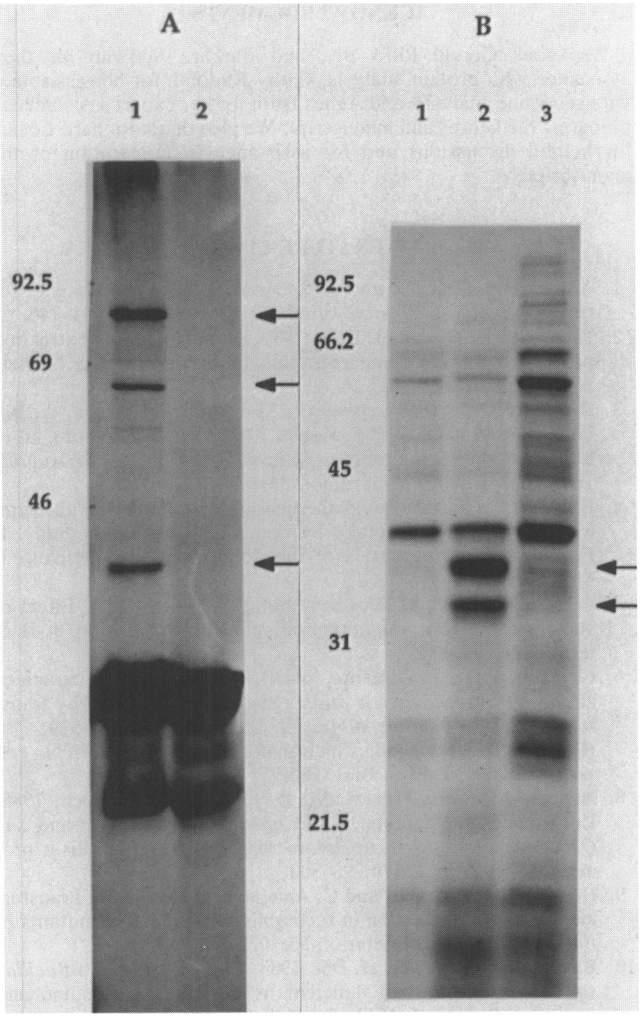

FIG. 4. Analysis of protein products by SDS-polyacrylamide gel electrophoresis. (A) Autoradiograph of [^{35}S]methionine-labeled proteins synthesized in an S30-coupled transcription-translation system and electrophoresed on a 7.5% polyacrylamide–SDS gel. Lanes: 1, pNP2; 2, pUC19. The arrows indicate the 75-kDa, 60-kDa, and 34-kDa proteins synthesized from pNP2. Molecular mass standards are given in kilodaltons. The thickest band in both lanes corresponds to β-lactamase of pUC19. (B) Electrophoresis of culture supernatants. Cultures were grown in LB with chloramphenicol (5 μg/ml). Three hours after the onset of stationary phase, culture supernatants were removed, boiled in SDS sample buffer, electrophoresed on a 7.5% polyacrylamide–SDS gel, and stained with Coomassie brilliant blue. Lanes: 1, pBs81/6; 2, pNP3; 3, pNP6 (pNP1 with the 1.6-kb *Eco*RV fragment deleted). The arrows indicate the 38-kDa and 34-kDa proteins synthesized from pNP3. Molecular mass standards are given in kilodaltons.

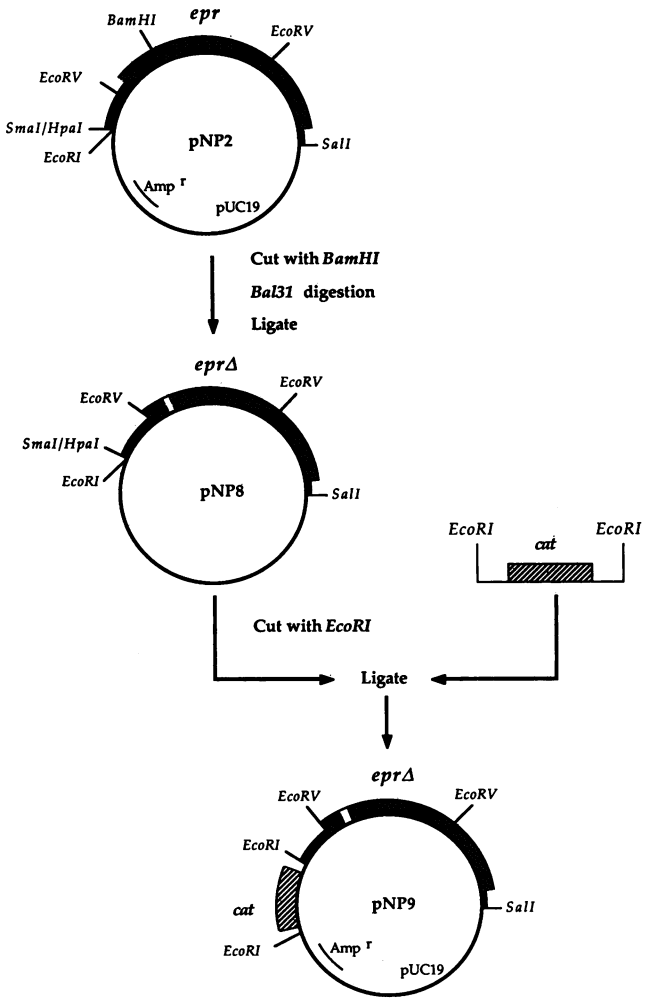

FIG. 5. Construction of plasmid pNP9 containing a deletion mutation of the *epr* gene.

these strains contained the *E. coli* replicon and *cat* gene flanked by two copies of the deleted *epr* gene. To screen for a strain that had undergone a recombination event between the two copies of the Δ*epr* and hence had jettisoned the *cat* gene and the *E. coli* replicon, a colony was selected and grown overnight in rich medium without drug selection. Individual colonies were then screened for drug resistance, and about 0.1% were found to be Cm^s. One such strain (GP216) was selected for further study.

First, Southern hybridization was used to confirm that the strain did indeed contain a deletion in the chromosomal *epr* gene (Fig. 6). Second, like the Cm^r parent strain, GP216 failed to produce a halo on casein plates (Fig. 7). In liquid cultures, however, [^{14}C]casein protease assays indicated that *epr* alone did not contribute significantly to the residual protease activity. A strain with deletion mutations in *epr*, *apr*, *npr*, and *isp* did not produce significantly less protease than a strain with mutations in just *apr*, *npr*, and *isp* in liquid culture. Apparently, this result is not reflected on a casein plate. Finally, both growth and sporulation were checked in standard laboratory media. No differences were seen in growth in LB medium. Similarly, no appreciable differences

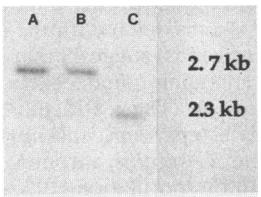

FIG. 6. Southern hybridization depicting deletion of the chromosomal *epr* gene. *B. subtilis* chromosomal DNA was digested with *Hin*dIII and probed with the nick-translated *epr* gene. (A) GP208; (B) GP208; (C) GP216 (Δ*epr*).

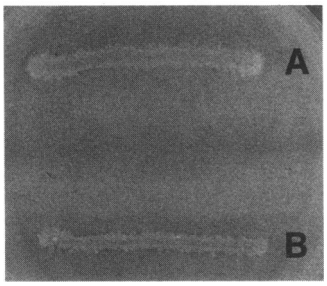

FIG. 7. Assay for protease activity on casein plates. A top layer containing 1% dry milk was added to tryptose blood agar base plates. (A) GP216 (Δ*epr*); (B) GP208.

were seen in sporulation frequency after growth on DSM medium (18) for 30 h (10^8 spores per ml for both GP208 and GP216).

DISCUSSION

The primary product of the *epr* gene is a 69,702-dalton protein that appears to be processed to yield a 34,000-dalton protease. Subclones of the gene encoding only the first 405 of the 640 amino acids produced the active 34,000-dalton protease, arguing that the enzyme normally undergoes C-terminal processing and that the final 240 amino acids are not required for activity. Several unusual features were observed in this C-terminal portion of the protein. First, the C-terminal domain was unusually rich in basic amino acids, with lysine accounting for approximately 24% of the amino acids. Second, a partially homologous 44-amino-acid sequence was repeated five times between nucleotides 1168 and 1927 (Fig. 2).

Two other extracellular proteases have been described which contain a large C-terminal domain that is processed. However, in both these cases (IgA protease from *Neisseria gonorrhoeae* [14] and a serine protease from *Serratia marcescens* [22]) the C-terminal domain is involved in excretion of the protease through the outer membrane of a gram-negative bacterium and is essential for enzyme production. The role of the C-terminal 240 amino acids of Epr remains to be elucidated.

The protease encoded by *epr* is a serine protease with homology (40%) to both subtilisin and ISP-1 (Fig. 3A and B). Epr does not appear to correspond to a previously described esterase, bacillopeptidase F (12, 16). The amino acid composition predicted for Epr does not resemble that deduced for bacillopeptidase F (16), and a strain carrying a deletion in the *epr* gene (GP216) did not produce less esterase activity than the parent strain GP208. R. Bruckner and R. Doi (personal communication) have independently identified a new protease gene that maps in the *sacA* region of the *B. subtilis* chromosome and is likely to be the same as *epr*.

The presence of residual protease activity in the culture fluids in *B. subtilis* strains deleted for the *epr* gene (as well as the *apr*, *npr*, and *isp* genes) suggests that *B. subtilis* contains additional genes encoding minor extracellular proteases, perhaps similar to *epr*. The extracellular protease genes characterized to date (*epr*, *apr*, and *npr*) are all located at different loci on the *B. subtilis* chromosome, and none are essential for growth or sporulation (19, 23). *apr* and *npr* are both regulated by the *sacQ*, *sacU*, *hpr*, *prtR*, and certain *spo0* genes (5, 8). It will be interesting to determine whether the *epr* gene is also regulated by these loci in the same manner as the *apr* and *npr* genes.

ACKNOWLEDGMENTS

We thank Gerald Rufo, Jr., and Barbara Sullivan for their assistance with protein analysis, Cathy Rudolph for her assistance with sequence analysis, and Agnes Huot for her expert assistance in preparing the figures and manuscript. We also thank Richard Losick for helpful discussions and for assistance in preparation of the manuscript.

LITERATURE CITED

1. **Anagnostopoulos, C., and J. Spizizen.** 1961. Requirements for transformation in *Bacillus subtilis*. J. Bacteriol. **81**:741–746.
2. **Birnboim, H. C., and J. Doly.** 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res. **7**:1513–1523.
3. **Dedonder, R. A., J. Lepesant, J. Lepesant-Kejzlarova, A. Billault, M. Steinmetz, and F. Kunst.** 1977. Construction of a kit of reference strains for rapid genetic mapping in *Bacillus subtilis* 168. Appl. Environ. Microbiol. **33**:989–993.
4. **Dubnau, D., and R. Davidoff-Abelson.** 1971. Fate of transforming DNA following uptake by competent *Bacillus subtilis*. I. Formation and properties of the donor-recipient complex. J. Mol. Biol. **56**:209–221.
5. **Ferrari, E., S. M. H. Howard, and J. A. Hoch.** 1986. Effect of stage 0 sporulation mutations on subtilisin expression. J. Bacteriol. **166**:173–179.
6. **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.
7. **Hanahan, D.** 1983. Studies on transformation of *Escherichia coli* with plasmids. J. Mol. Biol. **166**:557–580.
8. **Henner, D. J., E. Ferrari, M. Perego, and J. A. Hoch.** 1988. Location of the targets of the *hpr-97*, *sacU32*(Hy), and *sacQ36*(Hy) mutations in upstream regions of the subtilisin promoter. J. Bacteriol. **170**:296–300.
9. **Hoch, J. A., M. Barat, and C. Anagnostopoulos.** 1967. Transformation and transduction in recombination-defective mutants of *Bacillus subtilis*. J. Bacteriol. **93**:1925–1937.
10. **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.
11. **Koide, Y., A. Nakamura, T. Uozomi, and T. Beppu.** 1986. Cloning and sequencing of the major intracellular serine protease gene of *Bacillus subtilis*. J. Bacteriol. **167**:110–116.
12. **Mamas, S., and J. Millet.** 1975. Purification et propriétés d'une estérase excrétée pendant la sporulation de *Bacillus subtilis*. Biochimie **57**:9–16.
13. **Perlman, D., and H. O. Halvorson.** 1983. A putative signal peptidase recognition site and sequence in eucaryotic and procaryotic signal peptides. J. Mol. Biol. **167**:391–409.
14. **Pohlner, J., R. Halter, K. Beyreuther, and T. F. Meyer.** 1987. Gene structure and extracellular secretion of *Neisseria gonorrhoeae* IgA protease. Nature (London) **325**:458–462.
15. **Priest, F. G.** 1977. Extracellular enzyme synthesis in the genus *Bacillus*. Bacteriol. Rev. **41**:711–753.
16. **Roitsch, C. A., and J. H. Hageman.** 1983. Bacillopeptidase F: two forms of a glycoprotein serine protease from *Bacillus subtilis* 168. J. Bacteriol. **155**:145–152.
17. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
18. **Schaeffer, P., J. Millet, and J. Aubert.** 1965. Catabolite repression of bacterial sporulation. Proc. Natl. Acad. Sci. USA **54**: 704–711.
19. **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.
20. **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.
21. **Uehara, H., K. Yamane, and B. Maruo.** 1979. Thermosensitive,

extracellular neutral proteases in *Bacillus subtilis*: isolation, characterization, and genetics. J. Bacteriol. **139**:583–590.

22. **Yanagida, N., T. Uozumi, and T. Beppu.** 1986. Specific excretion of *Serratia marcescens* protease through the outer membrane of *Escherichia coli*. J. Bacteriol. **166**:937–944.
23. **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.
24. **Youngman, P. J., J. B. Perkins, and R. Losick.** 1984. Construction of a cloning site near one end of Tn*917* into which foreign DNA may be inserted without affecting transposition in *Bacillus subtilis* or expression of the transposon-borne *erm* gene. Plasmid **12**:1–9.