

## Cloning, Characterization, and Inactivation of the *Bacillus brevis lon* Gene

KIMIO ITO, SHIGEZO UDAKA, AND HIDEO YAMAGATA\*

Department of Food Science and Technology, Faculty of Agriculture,  
Nagoya University, Chikusa-ku, Nagoya 464, Japan

Received 26 July 1991/Accepted 25 January 1992

A gene of *Bacillus brevis* HPD31 analogous to the *Escherichia coli lon* gene has been cloned and characterized. The cloned gene (*B. brevis lon* gene) encodes a polypeptide of 779 amino acids with a molecular weight of 87,400 which resembles *E. coli* protease La, the *lon* gene product. Fifty-two percent of the amino acid residues of the two polypeptides were identical. The ATP-binding sequences found in *E. coli* protease La were highly conserved. The promoter of the *B. brevis lon* gene resembled that recognized by the major RNA polymerase of *Bacillus subtilis* and did not contain sequences homologous to the *E. coli* heat shock promoters. The *B. brevis lon* gene was inactivated by insertion of the neomycin resistance gene. A mutant *B. brevis* carrying the inactivated *lon* gene showed diminished ability for the degradation of abnormal polypeptides synthesized in the presence of puromycin.

The *lon* gene of *Escherichia coli* is the structural gene for the ATP-dependent serine protease La, which is thought to catalyze rate-limiting steps in the degradation of many short-lived or abnormal proteins (8, 10, 30, 38). Protease La is a tetrameric enzyme composed of identical subunits (10, 28). The enzyme is one of the heat shock proteins whose synthesis is under the control of the *rpoH* (*htpR*) gene (13). The *rpoH* gene encodes a specific  $\sigma$  factor necessary for the recognition of heat shock gene promoters by RNA polymerase (11). The *lon* gene has been cloned, and its nucleotide sequence has been determined. The deduced amino acid sequence of the subunit of protease La comprises 783 amino acids with a molecular weight of 87,000 (9). *E. coli* mutants deficient in protease La show mucoidy (24), a filamentous form, increased sensitivity to UV light (3, 41) or methyl methanesulfonate (MMS [15]), and an inability to lysogenize some bacteriophages (37). These properties result from a failure to rapidly degrade proteins that regulate the corresponding cellular processes (30, 38). The mutants also show a diminished rate of degradation of abnormal proteins such as puromycin polypeptides, amino acid analog-containing polypeptides (10), and products of heterologous genes (6).

Since *E. coli* protease La plays such important roles in protein metabolism and the regulation of various cellular processes, it is of interest to examine whether or not the enzyme is conserved in other organisms such as bacilli. Bacilli are gram-positive bacteria with cell structures quite different from those of *E. coli*. They undergo sporulation and spore germination, which include many processes of protein turnover. In addition, comparison of proteases that resemble La and are of different origins should provide further information on the structure and function of this unique type of protease.

In this paper, we describe the cloning and characterization of a gene of *Bacillus brevis* HPD31 analogous to the *E. coli lon* gene. *B. brevis* HPD31 is a bacterium used as a host for heterologous protein production (40, 43). We also describe the construction and properties of mutants containing insertion mutations in the *lon* gene of *B. brevis* HPD31.

### MATERIALS AND METHODS

**Bacteria, bacteriophages, media, and transformation.** The bacterial strains used were *B. brevis* HPD31 (35) and *E. coli* JM103 (44) and Q358 (20). *B. brevis* was grown in T2 (39) or M2 (22) medium at 30°C. To grow plasmid-carrying *B. brevis* strains, 10  $\mu$ g of erythromycin, 20  $\mu$ g of tetracycline, or 60  $\mu$ g of neomycin per ml was added to the medium. *E. coli* JM103 was used as a host for DNA manipulation, and Q358 was used to grow  $\lambda$ 3B6 (21). Transformation of *B. brevis* was performed by the Tris-polyethylene glycol method (36).

**DNA and RNA preparation.** Chromosomal DNA was prepared from *B. brevis* HPD31 as described by Saito and Miura (32). Phage and plasmid DNAs were prepared as described by Sambrook et al. (33) and Birnboim (5), respectively. Total RNA was prepared from *B. brevis* HPD31 carrying pTA200SBG by the hot-phenol method of Aiba et al. (4).

**Southern blot analysis of the *B. brevis* genome.** A nonradioactive DNA labeling and detection kit (Boehringer, Mannheim, Germany) was used to detect the *B. brevis lon* gene. Hybridization was carried out according to the manufacturer's standard protocol except that the temperature was lowered from 68 to 55°C.

**DNA sequence determination.** The dideoxy chain termination method (34) was employed after subcloning of appropriate restriction fragments into the pUC and M13 vectors (44). All the ends of restriction fragments used overlapped one another, and the sequences of both strands were determined.

**Primer extension assay of *B. brevis* RNA.** The method of primer extension described by Sambrook et al. (33) was used. RNA (100  $\mu$ g) prepared from *B. brevis* HPD31 carrying pTA200SBG was incubated with 50 ng of primer oligonucleotides in 30  $\mu$ l of hybridization buffer for 10 min at 85°C and then for 12 h at 55°C. The extension reaction was carried out in 40  $\mu$ l of reverse transcriptase buffer for 10 min at 37°C using 50 U of reverse transcriptase (Takara, Kyoto, Japan) and 20  $\mu$ Ci of [<sup>32</sup>P]dCTP. The primer oligonucleotide 5'-CGGAACGTTCCGCCAAGCGG-3' was synthesized at the Center for Gene Research of Nagoya University.

**Construction of the fusion of the *E. coli lacZ* and *B. brevis lon* genes.** The *Hpa*II 440-bp fragment (from the *Hpa*II site located upstream of the *Mfl*I site [see Fig. 3] to the one at

\* Corresponding author.

nucleotide 337 [see Fig. 2]) was isolated. After filling in of its cohesive ends by treatment with Klenow fragment of DNA polymerase I in the presence of four deoxynucleoside triphosphates, the fragment was inserted into the *Bam*HI site of pMC1403 (7) with the aid of 8-mer *Bam*HI linkers (CG GATCCG). On the resultant plasmid (pLL1), a 6-bp sequence, GGCGAT, connected in frame the fifth codon downstream from TTG (nucleotides 324 to 326; see Fig. 2) to the eighth codon of the  $\beta$ -galactosidase gene. The *Sma*I-*Dra*I 4-kb fragment encompassing the fused gene was isolated from pLL1 and inserted into the *Sma*I site of pHY300PLK (17). The resultant plasmid, pLL4, was used to transform *B. brevis* HPD31.

**$\beta$ -Galactosidase assay.** *B. brevis* HPD31 carrying pLL4 was grown in T2 medium supplemented with 20  $\mu$ g of tetracycline per ml. When the  $A_{660}$  reached 1.0,  $\beta$ -galactosidase activity was measured. The assay conditions and the definition of an activity unit were given by Miller (29).

**Measurement of degradation of puromycyl polypeptides.** *B. brevis* HPD31 or its mutant, L13, was grown at 37°C in M2 medium supplemented with all amino acids except leucine and valine until the  $A_{660}$  reached 0.4. Puromycin was added to a final concentration of 40  $\mu$ g/ml. In the control culture, no puromycin was added. After 15 min of incubation at 37°C, [<sup>3</sup>H]leucine was added to 20  $\mu$ Ci/ml. After 5 min, cells were collected and washed by centrifugation; they were then grown at 37°C in M2 medium supplemented with all amino acids. Portions of the cells were taken at intervals and precipitated with 5% trichloroacetic acid. Radioactivity in the acid-soluble or insoluble fraction was measured as described by Goldberg (14).

**Other analytical methods.** The sensitivity of bacteria to UV was measured as described by Uretz and Markovitz (41). *B. brevis* strains were grown in T2 medium and exposed to UV of various doses, which gave surviving efficiencies ranging from 95 to 0.01% as determined by plating on T2 agar plates. Sensitivity to MMS was measured as described by Gottesman and Zipser (15). *B. brevis* strains grown in T2 medium were plated on T2 agar plates containing various concentrations of MMS (0.01 to 0.1%), which gave efficiencies of plating ranging from 100% to less than 0.001%.

**Nucleotide sequence accession number.** The nucleotide sequence shown in Fig. 2 was submitted to the DDBJ data base (accession number D00863).

## RESULTS

### Cloning and nucleotide sequence of the *B. brevis lon* gene.

By comparison of Kohara's restriction map of the *E. coli* genome (21) with the nucleotide sequence of the *E. coli lon* gene (9),  $\lambda$ 3B6 (21) was deduced to contain the *E. coli lon* gene. Therefore, we subcloned the *Eco*RI-*Hind*III 0.6-kb and *Hind*III-*Sph*I 3-kb fragments from  $\lambda$ 3B6 onto pUC119 and pKEN403 (25), respectively (pUC119ECLN and pKEN403ECLC; Fig. 1). Nucleotide sequence analysis proved that the subcloned fragments contain the *E. coli lon* gene. These fragments were used as hybridization probes in Southern blot analysis of *B. brevis* genomic DNA. The analysis showed that the *B. brevis* genome contains a sequence (*B. brevis lon* gene) homologous to the *E. coli lon* gene. The restriction map of the *B. brevis* genome around the *lon* gene is shown in Fig. 1.

The *Eco*RI-*Bgl*II 1.6-kb, *Hind*III-*Eco*RI 2.3-kb, *Bam*HI-*Hind*III 1.3-kb, and *Hind*III-*Sac*I 2.8-kb fragments containing various portions of the *B. brevis lon* gene shown in Fig. 1 were isolated and cloned onto pUC18 or pUC19 (pUC18B

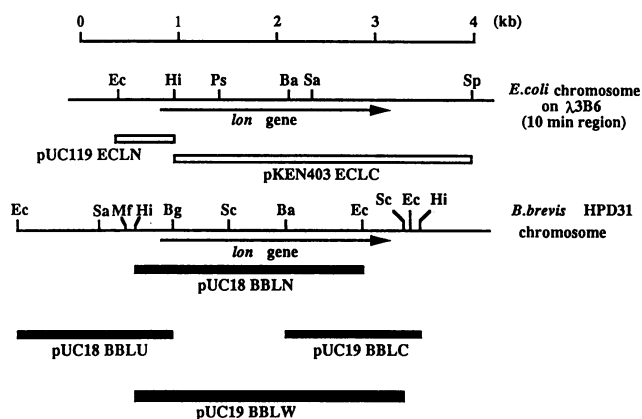


FIG. 1. Restriction maps around the *lon* gene of *E. coli* and *B. brevis* chromosomes and the DNA fragments cloned and used in this study. Restriction sites; Ec, *Eco*RI; Hi, *Hind*III; Ps, *Pst*I; Ba, *Bam*HI; Sa, *Sal*I; Sp, *Sph*I; Mf, *Msp*I; Bg, *Bgl*II; Sc, *Sac*I.

BLU, pUC18BBLN, pUC18BBLC, and pUC19BBLW, respectively; Fig. 1). The 3,146-bp nucleotide sequence from the *Msp*I to the *Hind*III site encompassing the *B. brevis lon* gene was determined (Fig. 2). The sequence from nucleotides 306 to 316 was homologous to the 3'-terminal sequence of *B. brevis* 16S rRNA (23), suggesting that the sequence acts as a ribosomal binding sequence (Shine-Dalgarno sequence [SD]). Starting from TTG at nucleotide 324 (8 bp downstream from SD), a long open reading frame of 2,337 bp was found. TTG is used as an initiation codon more frequently in *B. brevis* than in *E. coli* (2). Sixty-one percent of the nucleotides of the open reading frame were identical to those of the *E. coli lon* gene.

**Promoter and transcription start sites of the *B. brevis lon* gene.** To localize the promoter region of the *B. brevis lon* gene, a promoter probe vector, pTA200 (1), was used. pTA200 can replicate in *B. brevis* and contains a promoter-deficient *Bacillus licheniformis*  $\alpha$ -amylase gene and the erythromycin resistance gene. Only when a DNA fragment with promoter activity is inserted into the *Bam*HI or the *Eco*RI site located immediately upstream of the  $\alpha$ -amylase gene can it direct the synthesis of  $\alpha$ -amylase in *B. brevis*. Four DNA fragments, the *Hind*III-*Hpa*II, *Sal*I-*Hind*III, *Sal*I-*Msp*I, and *Sal*I-*Bgl*II fragments shown in Fig. 3, were isolated from the 5' flanking region of the *B. brevis lon* gene and then inserted into the *Bam*HI site of pTA200 (pTA200HHP, -SH, -SM, and -SBG, respectively). Plasmids pTA200SH and pTA200SBG introduced into *B. brevis* HPD31 directed efficient  $\alpha$ -amylase production, whereas pTA200HHP and pTA200SM did not (Fig. 3), suggesting that some essential element of the promoter of the *B. brevis lon* gene is located between the *Msp*I and *Hind*III sites. In this region, sequences TAGACA and TACAAT, homologous to the consensus -35 and -10 sequences (TTGACA and TATAAT, respectively) recognized by  $\sigma$ A( $\sigma^{43}$ )-RNA polymerase of *Bacillus subtilis* (31) were found (Fig. 3, bottom). The distance between the putative -35 and -10 sequences was 17 bp, which is common in many promoters. No sequences analogous to the heat shock promoters of *E. coli* (11) were found.

To determine the transcription start site, RNA prepared from *B. brevis* HPD31 carrying pTA200SBG was analyzed by the primer extension method, because the detection of transcripts from the chromosomal *lon* gene was difficult. The

GGATCTGGCATCTTCGCCATTTTCCTTGTAAACACTCGGAGCGCATGCAAAATTAGACATCGGTTTGCCTTTCGGCTACAATGGAAAAAGCTTCTGCTG 100  
 NflI -35 -10 HindIII  
CAAAGAGCTTGGGCCATGCCGCTTTGCTTTTGGGGGCGCCGTGCAGGAGCGAAACGGCTGTTTCGGCGTTTTCATGTGCGACTATCGAAGGATT 200  
 SD M P L C F W G R R A G A K T A V S A V F A C A T I E G I  
 GACGTGACGAGAGAGTGTAAAAGCGGACGAGGCTACGCCCTGCCACTGGTGTAGCCGAGTTTCTCATTAGCTGACTCTTGGTAAGAGGGATAGAT 300  
 D V Y A D D G K S A T R L R R P T G V A E F S H \*  
TCAACAAGGACAGCTCATCGCTTGGGCGAACCTTCGGTAAAGAGAGAATTACCGCTTCTCCGTTGCGAGGCTTCTTATATCCGACTATGTTCT 400  
 SD M G E R S G K R E L P L L P L R G L L V Y P T H V L  
 HpaII  
 TCATTGGACGTTGGACGGGAAAAAGTCCATTGGTGCATTGGAGCAAGCGATGCTGACGACAAACAAGATCTTCTGCGACGCGAGGAAGAATACACATA 500  
 H L D V G R E K S I R A L E Q A M V D D N K I L L A T Q E E V H I  
 BglII  
 GAAGAACCTGATGCTGAACAAATCTACAGCATCGGAACCTTGGCGCGTAAACAAATGCTGAAGCTGCCAACCGTACGATCCGTATTGTTGGGAAG 600  
 E E P D A E Q I Y S I G T V A R V K Q H L K L P N G T I R V L V E G  
 CCTGCGCGCTGCCAAAATCGAAGAAATCTTCAAAAAGAAGATTCTGCTGTCATTACATACTTGAAGGAAGAAAAAGCGGAAAGAAACGAAT 700  
 L Q R A K I E E Y L Q K E D Y F V V S I T Y L K E E K A E E N E V  
 GGAAGCGTTGATCGCTCTTTGCTGACCCATTTCGAGCAATATCAAGCTGTCGAAAAAGTTTCTCCGAGACGCTCACATCCGTGACGAGCATCGAA 800  
 A L M R S L L T H F E Q Y I K L S K K V S P E T L T S V Q D I E  
 GAACCGAGGAGCGTGGCTGATGTCATCGCTCCCATGCGCGTCAAAATGAAGBATAAACAGGAGATTCTGGACAGTCAATATCCAGGAGCGTTGG 900  
 E P G R L A D V I A S H L P L K H K D R Q E I L E T V N I Q E R L E  
 AAATCTTTTGACCATTTGAACAACGAGCGGAAATTCGAGCTGGAGCGAAAAATCGCAATCGCGTGAAGAAGCAGATGGAGCGCACGCAAAAGGA 1000  
 I L L T I L N N E R E V L E L E R K I G N R V K K Q M E R T Q K E  
 ATATTACCTCGTAGCAAAATGAAGCGGATCCAGAAAGAGCTCGGGGACAGGAGCGGCGACGGGCGAAGTCGATGAGCTCGCGCTCACTGGAAG 1100  
 Y Y L R E Q M K A I Q K E L G D K D C R Q G E V D E L R A Q L E K  
 SacI  
 TCCGACGACCGGAGCGCATCAAAGCCAGATCCGAAAGAGCTGGAGCGCTGGAGAAAATCGCGTCTACTTTCGACAGGGGCTCTGTATCCGCACCT 1200  
 S D A P E R I K A K I E K L E R L E K H P S T S A E G S V I R T Y  
 ACATCGATACGCTGTTTCGCTGCCGTGGACCAAGACGACAGAAGCAATCTCGACATCAAGCATCGGGAAGAAGTCTGGATGAAGATCACTACGGACT 1300  
 I D T L F A L P W T K T T E D N L D I K H A E E V L D E D H Y G L  
 GGA AAAACCGAAGAGCGGCTACTGGAGTACTTGGCTGACAGAAGCTGGTCAATTCCATCGCGCGTCCGATTCTCGCTGGTGGTCTCCAGGGGGT 1400  
 E K P K E R V L E Y L A V Q K L V N S M R G P I L C L V G P P G V  
 GGA AAAACCTCTCGCGCGCTCTGTTGACAGCGCTGGT 1500  
 G K T S L A R S V A R A L G R E F V R I S L G G V R D E A E I R G H  
 ATCGCGTACGTCAGTAGGGCTGTCGCGGACCGCATTATCAAGGGTGAAGCAAGCGGGTACGATCAATCCCGTGTCTCGCTGGATGAGATCGACAA 1600  
 R R T Y V G A L P G R I I Q G H K Q A G T I N P V F L L D E I D K  
 GCTCGCGCTGATTTCCGCGGATCCAGCCTCAGCTTTCGCTGGAAGTCTCGACCCGAAACAAAATGACAAGTTCAGCGATCATTATAGAAGAAAG 1700  
 L A S D F R G D P A S A L L E V L D P N Q N D K F S D H Y I E E T  
 BamHI  
 TATGACTTGACGAACTGATGTTTACGACCCCAAGCGCTCGACACCAATCCCGGACCAATGCTCGACCGGATGGAAGTCAATTGATTTCCGGCT 1800  
 Y D L T N V H F I T T A N S L D T I P R P L L D R H E V I S I S G Y  
 ACACGAGCTGGA AAAACTGAATATCTCCGCGCTACCTGCTCGGACGAGATGAAGATCATGCTCGCAAGGATAAGCTGCAAAATGAACGAGGA 1900  
 T E L E K L N I L R G Y L L P K Q H E D H G L G K D K L Q M N E D  
 TGCCATGCTTAAGCTGCTCCGCTGTACACCGCGAAGCGGTGTGCGCAACCTGACAGGGAAGCGGCAATGCTGTGCAAAAAGCGGCAAGATCATC 2000  
 A H L K L V R L Y T R E A G V R N L N R E A A N V C R K A A K I I  
 GTCGCGGAGAGAAAGCGTGTAGTGTGACGGCCAAAACGCTGGAAGCCCTCCTCGAAAACCTCGCTACCGCTACGGCTGAGAAAAAGAC 2100  
 V G G E K K R V V V T A K T L E A L L G K P R Y R Y G L A E K K D Q  
 AGGTGCGCTGTATCGGCTGACACAGGCTGGCGGGATACGCTGAACGTGGAAGTCAAGCTCCTTGCAGGGAAGGCAAGCTGACCTTGAC 2200  
 V G S V T G L A W T Q A G G D T L N V E V S I L A G K G K L T L T  
 TGGACAGCTGGCGATGTGATGAAGGAATCCGCTCAGCGGCAATTCAGTATATTCGTTCCCGTGGAGCGAATGGGAATCGATCCGAGTTCACGAA 2300  
 G Q L G D V H K E S A Q A A F S Y I R S R A S E W G I D P E F H E  
 AAGAAGCATCCATATCCAGCTCCGGAAGCGCCATTCAAAGGATGGTCCATGCGAGGATACGATGGCGACTCGCTGATTCGCGCTGACCG 2400  
 K N D I H I H V P E G A I P K D G P S A G I T H A T A L V S A L T G  
 GAATTCCTGTGAAAAAGGAGTCCGAATGACAGGTGAAATCACCTCGCGCGCGGCTGCTTCAATCGCGGACTGAAAGAAAAATGCATGTCGCTCA 2500  
 I P V K K E V G H T G E I T L R G R V L P I G G L K E K C H S A H  
 EcoRI  
 TGGGCGAGGACTCAGCACCATCTTTGCCAAAAGACAATGAGAAAGACATCGAAGATATCCAGAAAGCGTACGGGAAGCCCTCAGCTTCTATCCGGT 2600  
 R A G L T T I I L P K D N E K D I E D I P E S V R E A L T F Y P V  
 GAGCATTGGACGAAGTCTCGCGATGCCCTGACGAAACAGCCAGTAAAGTGAACAAAAATGAAAGTAACTTCAATCCGAGTTTGTATCATGCTGTAGG 2700  
 E H L D E V L R H A L T K Q P V G D K K \*  
 ACCCAAGCAGTACCGGAGGACGGCTGACAGGATGCCCTCGTGGCGCTTCCAACGTAGGTAAGTCTCTCTTTGAAACAAAATGATGAAATCCGAAAG 2800  
 GGCTGGCGCATCAGTTCGCGCTGCAAAAACGACAGCTGAAATTAATTCGCGCTCAACAACATGCTCTATTTGCTGACTTCCCTGGCTACGGCTA 2900  
 CGCAAAAGTGGCAAGACCATTAAGAGCAATGGGGCAAAATGATTGAAGGCTACCTCAAGAAATCGCAAAAGCTCCGTTTTGTGATCAATGTTGGACT 3000  
 SacI  
 ATTCGACCGCTCCGGAAGGACGATATGGCATGTAGGATGGTCAAGCAAAATCGAAATTCGACCGTGTAGTGGCTACCAAGGGGATAAGATTG 3100  
 EcoRI  
 CACCGGACGCTGGCTGACGATACGAAGGTATCCGCTCAAAGCTT 3146  
 HindIII

FIG. 2. Nucleotide and deduced amino acid sequences of the *B. brevis lon* gene. The numbers on the right are nucleotide positions. The transcription start site is indicated by an arrowhead at the top. The -35 and -10 sequences of the promoter, ribosome-binding sites (SD), and restriction sites used in this study are underlined. The termination codons of the *lon* gene and a short open reading frame located upstream of the *lon* gene are indicated by asterisks.

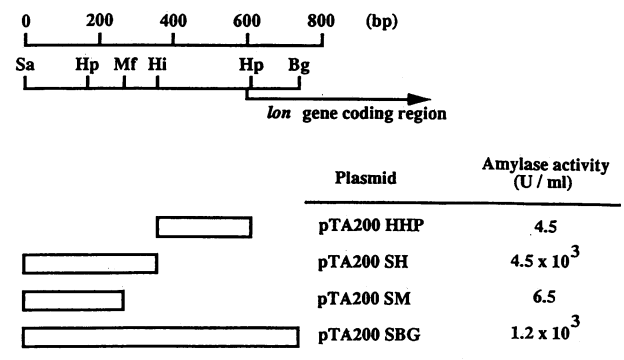


FIG. 3. Localization of the promoter of the *B. brevis lon* gene. The restriction map of the 5' flanking region of the *lon* gene is at the top. The DNA fragments inserted in the promoter probe vector pTA200 and the  $\alpha$ -amylase production by *B. brevis* due to the resulting plasmids are shown in the middle. The nucleotide sequence of the *MflI*-*HindIII* fragment containing the promoter is shown at the bottom. The -35 and -10 sequences are boxed, and the major transcription start site determined by primer extension assays, as shown in Fig. 4, is indicated by an asterisk.  $\alpha$ -Amylase activity was determined as described previously (42). Sa, *Sall*; Hp, *HpaII*. The *MflI* (Mf), *HindIII* (Hi), and *BglII* (Bg) sites are those shown in Fig. 2 at nucleotides 1 to 6, 88 to 93, and 466 to 471, respectively.

*Sall*-*BglII* 740-bp fragment inserted in pTA200SBG contained the 600-bp 5' flanking region and the 140-bp region encoding the putative NH<sub>2</sub>-terminal part of the *B. brevis lon* gene product. A 20-bp oligonucleotide complementary to the sequence from immediately downstream of SD to the putative fifth codon of the *B. brevis lon* gene (nucleotides 320 to 339 in Fig. 2) was synthesized and used as a primer. The 5' ends of the transcripts from the *B. brevis lon* promoter were mapped at the A cluster located 5 bp downstream from the above-described putative -10 sequence (Fig. 4 and nucleotides 85 to 88 in Fig. 2). Major transcripts seemed to start from A at nucleotide 86. The distance between the transcription start site and the putative initiation codon TTG was 237 bp.

**Translation initiation site of the *B. brevis lon* gene.** The fifth codon (nucleotides 336 to 338, Fig. 2) downstream from TTG at nucleotides 324 to 326 of the *B. brevis lon* gene was connected in frame to the eighth codon of the *E. coli*  $\beta$ -galactosidase gene on pMC1403 (7), a vector for the detection of translation initiation, as described in Materials and Methods. The fused gene introduced into *B. brevis* HPD31 using pHY300PLK (17) as a vector directed the synthesis of  $\beta$ -galactosidase (7 U/ml). *B. brevis* HPD31 carrying no plasmid synthesized no detectable  $\beta$ -galactosidase. This confirmed that the translation initiation site of the *B. brevis lon* gene is located upstream of nucleotide 338. Between the transcription start site described above and nucleotide 338, there are two combinations of sequences homologous to *B. brevis* 16S rRNA (23) and initiation codons at appropriate distances. One is AAAGAGGT at nucleotides 102 to 109, with ATG 8 bp downstream, and the other is the above-described AAGGAGAGGTG at nucleotides 306 to

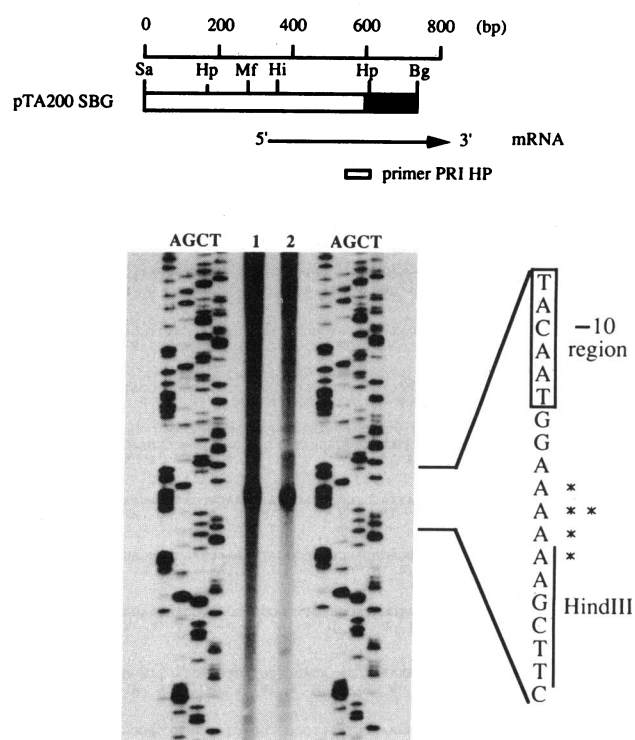


FIG. 4. Primer extension assay of *B. brevis lon* gene transcripts. A restriction map of the DNA fragment contained by pTA200SBG is at the top. Open and closed bars indicate noncoding and coding regions, respectively. The position of the primer used in the assay is also indicated (PRI HP). Two independent preparations of RNA from *B. brevis* HPD31 carrying pTA200SBG were assayed (lanes 1 and 2). The lengths of the extension products were determined by electrophoresis on a sequencing gel with the products of the dideoxy chain termination reaction (AGCT), where pTA200SBG and PRI HP were used as the template and primer, respectively. The positions of the 5' ends of the extension products are shown by asterisks to the right of the sequence. The positions of the 5' ends of major products are shown by double asterisks.

316, with TTG 8 bp downstream. The open reading frame starting from the former initiation codon ends at TAG at nucleotides 273 to 275. Therefore, TTG at nucleotides 324 to 326 should be the translation initiation site of the *B. brevis lon* gene.

**Amino acid sequence of the *B. brevis lon* gene product.** The *B. brevis lon* gene encodes a polypeptide of 779 amino acids with a molecular weight of 87,400 which is quite similar to *E. coli* protease La (783 amino acids with a molecular weight of 87,000). The two amino acid sequences are compared in Fig. 5. The homology extends over almost the full lengths of the two sequences, 52% of the amino acids being identical. There are two regions where the two amino acid sequences show very low homology (amino acids 254 to 317 and 539 to 573 in the *B. brevis* sequence, the homology being 10 and 8%, respectively). The nucleotide sequences corresponding to these regions show 52 and 57% homology. The low homology in the amino acid sequence in these regions resulted from insertions or deletions in the nucleotide sequence that shifted the reading frame. The reading frame was shifted again, back to the original frame, at the ends of these regions. Between the two nonhomologous regions, there is a region that shows high homology (67%; amino



similar to La, the gene product of the *E. coli lon* gene, might be conserved in a variety of organisms, as has been shown for another ATP-dependent intracellular protease, Clp, of *E. coli* (16, 26). Fifty-two percent of the amino acids of the *B. brevis lon* product are identical to those of protease La. The identical residues are spread out along the amino acid sequence except in two narrow regions, where the homology is very low. Moreover, ATP-binding sequences found in protease La are found in the *B. brevis lon* product at exactly the same positions, suggesting that the latter is also an ATP-dependent serine protease. The sequences (GDSGGP, GDSGG, TSMA, and AASMGGA) which are found in the active sites of many mammalian serine proteases (19), some bacterial serine proteases and trypsin (18), subtilisin BPN' and related enzymes (18), and *E. coli* Clp protease (26), respectively, could not be found in *E. coli* protease La or in the *B. brevis lon* product. *E. coli* protease La contains 43 serine residues, while the *B. brevis lon* product contains 33. Recently, a gene of *Myxococcus xanthus* analogous to the *E. coli lon* gene was cloned, and its nucleotide sequence was determined (21a). The *M. xanthus lon* product contains 29 serine residues. Assuming that this polypeptide also functions as a serine protease, an active-site serine residue is most probably one of the conserved serine residues among these three enzymes. These residues are Ser-35, -363, -367, -429, -437, -678, and -689 (amino acid numbers are for the *B. brevis* sequence). Ser-437 or Ser-678 may be a candidate for an active-site residue because each is surrounded by Gly, Pro, Ala, and Asp, which frequently appear at positions close to the catalytic serine residues in many serine proteases, as described above.

The promoter region of the *B. brevis lon* gene resembles those of the vegetative promoters of *B. subtilis* recognized by a major  $\sigma A$  RNA polymerase (31). Sequences analogous to those of the heat shock promoters of *E. coli* could not be found. Repeated attempts to detect a heat shock response in transcription from the *B. brevis lon* gene promoter were unsuccessful (data not shown). RNAs extracted from *B. brevis* carrying plasmid pTA200SBG were analyzed in these experiments. Increase in the gene copy number or fusion to a foreign gene might affect the regulation of expression of the *lon* gene, or the regulatory mechanism might be distinct from that in the *E. coli lon* gene.

A mutant of *B. brevis* containing an insertion in the *lon* gene showed a diminished rate of degradation of puromycin peptides. This indicates that the *B. brevis lon* gene plays a role similar to that of the *E. coli lon* gene in the degradation of abnormal and probably short-lived regulatory proteins. However, other phenotypes shown by *E. coli lon* mutants, such as increased sensitivities to UV and MMS, mucoidy, and a filamentous cell form, were not observed in *B. brevis lon* mutants. In *E. coli*, these properties result from a failure to rapidly degrade proteins that regulate filamentation and the expression of genes for capsular polysaccharide synthesis such as SulA (30) and RcsA (38). In *B. brevis*, the mechanism of regulation of these cellular processes and the properties of regulatory proteins involved in these processes might be different from those in *E. coli*. Further work is necessary to elucidate the physiological role of the *lon* gene in *B. brevis*. Comparison of the two proteases with respect to the recognition of abnormal proteins as substrates is particularly interesting. Are *E. coli* proteins recognized by the *B. brevis* protease as abnormal proteins and vice versa?

*B. brevis* HPD31, a strain recently isolated from soil, secretes vast amounts of proteins but exhibits little or no protease activity in the medium (35). A host-vector system

for the production of heterologous proteins with this bacterium as a host has been developed. Many bacterial and mammalian proteins have been produced efficiently with this system (40, 43). The use of the *B. brevis lon* mutant isolated here is expected to improve the efficiency of heterologous protein production in this system, as was the case for *E. coli lon* mutants (6).

#### ACKNOWLEDGMENTS

We are grateful to M. J. Casadaban and Y. Kohara for the plasmid and phages, T. Komano for the unpublished sequence of the *M. xanthus lon* gene, S. Ichihara and T. Sasaki for discussion, and S. Otsuka for preparation of the manuscript.

#### REFERENCES

- Adachi, T., H. Yamagata, N. Tsukagoshi, and S. Udaka. 1989. Multiple and tandemly arranged promoters of the cell wall protein gene operon in *Bacillus brevis* 47. *J. Bacteriol.* 171:1010-1016.
- Adachi, T., H. Yamagata, N. Tsukagoshi, and S. Udaka. 1990. Use of both translation initiation sites of the middle wall protein gene in *Bacillus brevis* 47. *J. Bacteriol.* 172:511-513.
- Adler, H. I., and A. A. Hardigree. 1964. Analysis of a gene controlling cell division and sensitivity to radiation in *Escherichia coli*. *J. Bacteriol.* 87:720-726.
- Aiba, H., S. Adhya, and B. de Crombrughe. 1981. Evidence for two functional *gal* promoters in intact *Escherichia coli* cells. *J. Biol. Chem.* 256:11905-11910.
- Birnboim, H. C. 1983. A rapid alkaline extraction method for the isolation of plasmid DNA. *Methods Enzymol.* 100:243-255.
- Buell, G., M.-F. Schulz, G. Selzer, A. Chollet, N. R. Movva, D. Semon, S. Escanez, and E. Kawashima. 1985. Optimizing the expression in *Escherichia coli* of a synthetic gene encoding somatomedin-C (IGF-I). *Nucleic Acids Res.* 13:1923-1938.
- Casadaban, M. J., J. Chou, and S. N. Cohen. 1980. In vitro gene fusions that join an enzymatically active  $\beta$ -galactosidase segment to amino-terminal fragments of exogenous proteins: *Escherichia coli* plasmid vectors for the detection and cloning of translational initiation signals. *J. Bacteriol.* 143:971-980.
- Charette, M. F., G. W. Henderson, and A. Markovitz. 1981. ATP hydrolysis-dependent protease activity of the *lon* (*capR*) protein of *Escherichia coli* K12. *Proc. Natl. Acad. Sci. USA* 78:4728-4732.
- Chin, D. T., S. A. Goff, T. Webster, T. Smith, and A. L. Goldberg. 1988. Sequence of the *lon* gene in *Escherichia coli*. A heat-shock gene which encodes the ATP-dependent protease La. *J. Biol. Chem.* 263:11718-11728.
- Chung, C. H., and A. L. Goldberg. 1981. The product of the *lon* (*capR*) gene in *Escherichia coli* is the ATP-dependent protease, protease La. *Proc. Natl. Acad. Sci. USA* 78:4931-4935.
- Cowing, D. W., J. C. A. Bardwell, E. A. Craig, C. Woolford, R. W. Hendrix, and C. A. Gross. 1985. Consensus sequence for *Escherichia coli* heat shock gene promoters. *Proc. Natl. Acad. Sci. USA* 82:2679-2683.
- Downs, D., L. Waxman, A. L. Goldberg, and J. Roth. 1986. Isolation and characterization of *lon* mutants in *Salmonella typhimurium*. *J. Bacteriol.* 165:193-197.
- Goff, S. A., L. P. Casson, and A. L. Goldberg. 1984. Heat shock regulatory gene *htpR* influences rates of protein degradation and expression of the *lon* gene in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* 81:6647-6651.
- Goldberg, A. L. 1972. Degradation of abnormal proteins in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* 69:422-426.
- Gottesman, S., and D. Zipser. 1978. Deg phenotype of *Escherichia coli lon* mutants. *J. Bacteriol.* 133:844-851.
- Gottesman, S., C. Squires, E. Pichersky, M. Carrington, M. Hobbs, J. S. Mattick, B. Dalrymple, H. Kuramitsu, T. Shiroza, T. Foster, W. P. Clark, B. Ross, C. L. Squires, and M. R. Maurizi. 1990. Conservation of the regulatory subunit for the Clp ATP-dependent protease in prokaryotes and eukaryotes. *Proc. Natl. Acad. Sci. USA* 87:3513-3517.

17. Ishiwa, H., and H. Shibahara-Sone. 1986. New shuttle vectors for *Escherichia coli* and *Bacillus subtilis*. IV. The nucleotide sequence of pHY300PLK and some properties in relation to transformation. *Jpn. J. Genet.* **61**:515-528.
18. James, M. N. G. 1976. Relationships between the structures and activities of some microbial serine proteases. II. Comparison of the tertiary structures of microbial and pancreatic serine proteases, p. 125-142. In D. W. Ribbons and K. Brew (ed.), *Proteolysis and physiological regulation*. Academic Press, Inc., Orlando, Fla.
19. Jurasek, L., R. W. Olafson, P. Johnson, and L. B. Smillie. 1976. Relationships between the structures and activities of some microbial serine proteases. I. Purification, enzymic properties and primary structures of *Streptomyces griseus* proteases A, B and trypsin, p. 93-123. In D. W. Ribbons and K. Brew (ed.), *Proteolysis and physiological regulation*. Academic Press, Inc., Orlando, Fla.
20. Karn, J., H. W. D. Matthes, M. J. Gait, and S. Brenner. 1980. A new selective phage cloning vector, lambda 2001, with sites for *Xba*I, *Bam*HI, *Hind*III, *Eco*RI, *Sst*I, and *Xho*I. *Gene* **32**:217-224.
21. Kohara, Y., K. Akiyama, and K. Isono. 1987. The physical map of the whole *E. coli* chromosome: application of a new strategy for rapid analysis and sorting of a large genomic library. *Cell* **50**:495-508.
- 21a. Komano, T. Personal communication.
22. Konishi, H., T. Sato, H. Yamagata, and S. Udaka. 1990. Efficient production of human  $\alpha$ -amylase by a *Bacillus brevis* mutant. *Appl. Microbiol. Biotechnol.* **34**:297-302.
23. Kop, J., A. M. Kopylov, L. Magrum, R. Siegel, R. Gupta, C. R. Woese, and H. F. Noller. 1984. Probing the structure of 16S ribosomal RNA from *Bacillus brevis*. *J. Biol. Chem.* **259**:15287-15293.
24. Markovitz, A. 1964. Regulatory mechanisms for synthesis of capsular polysaccharide in mucoid mutants of *Escherichia coli* K12. *Proc. Natl. Acad. Sci. USA* **51**:239-246.
25. Matsuyama, S., and S. Mizushima. 1985. Construction and characterization of a deletion mutant lacking *micF*, a proposed regulatory gene for OmpF synthesis in *Escherichia coli*. *J. Bacteriol.* **162**:1196-1202.
26. Maurizi, M. R., W. P. Clark, S.-H. Kim, and S. Gottesman. 1990. ClpP represents a unique family of serine proteases. *J. Biol. Chem.* **265**:12546-12552.
27. McKenzie, T., T. Hoshino, T. Tanaka, and N. Sueoka. 1986. The nucleotide sequence of pUB110: some salient features in relation to replication and its regulation. *Plasmid* **15**:93-103.
28. Menon, A. S., and A. L. Goldberg. 1987. Protein substrates activate the ATP-dependent protease La by promoting nucleotide binding and release of bound ADP. *J. Biol. Chem.* **262**:14929-14934.
29. Miller, J. H. 1972. *Experiments in molecular genetics*. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
30. Mizusawa, S., and S. Gottesman. 1983. Protein degradation in *Escherichia coli*: the *lon* gene controls the stability of SULA protein. *Proc. Natl. Acad. Sci. USA* **80**:358-362.
31. Moran, C. P., Jr., N. Lang, S. F. J. LeGrice, G. Lee, M. Stephens, A. L. Sonenshein, J. Pero, and R. Losick. 1982. Nucleotide sequences that signal the initiation of transcription and translation in *Bacillus subtilis*. *Mol. Gen. Genet.* **186**:339-346.
32. Saito, H., and K. Miura. 1963. Preparation of transforming DNA by phenol treatment. *Biochem. Biophys. Acta* **72**:619-629.
33. 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.
34. Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74**:5463-5467.
35. Takagi, H., K. Kadowaki, and S. Udaka. 1989. Screening and characterization of protein-hyperproducing bacteria without detectable exoprotease activity. *Agric. Biol. Chem.* **53**:691-699.
36. Takahashi, W., H. Yamagata, K. Yamaguchi, N. Tsukagoshi, and S. Udaka. 1983. Genetic transformation of *Bacillus brevis* 47, a protein-secreting bacterium, by plasmid DNA. *J. Bacteriol.* **156**:1130-1134.
37. Takano, T. 1971. Bacterial mutants defective in plasmid formation: requirement for the *lon*<sup>+</sup> allele. *Proc. Natl. Acad. Sci. USA* **68**:1469-1473.
38. Torres-Cabassa, A. S., and S. Gottesman. 1987. Capsule synthesis in *Escherichia coli* K-12 is regulated by proteolysis. *J. Bacteriol.* **169**:981-989.
39. Udaka, S. 1976. Screening for protein-producing bacteria. *Agric. Biol. Chem.* **40**:523-528.
40. Udaka, S., N. Tsukagoshi, and H. Yamagata. 1989. *Bacillus brevis*, a host bacterium for efficient extracellular production of useful proteins. *Biotechnol. Genet. Eng. Rev.* **7**:113-146.
41. Uretz, R. B., and A. Markovitz. 1969. Dominance of ultraviolet radiation resistance in partial diploids of *Escherichia coli* K-12. *J. Bacteriol.* **100**:1118-1120.
42. Yamagata, H., T. Adachi, A. Tsuboi, M. Takao, T. Sasaki, N. Tsukagoshi, and S. Udaka. Cloning and characterization of the 5' region of the cell wall protein gene operon in *Bacillus brevis* 47. *J. Bacteriol.* **169**:1239-1245.
43. Yamagata, H., K. Nakahama, Y. Suzuki, A. Kakinuma, N. Tsukagoshi, and S. Udaka. 1989. Use of *Bacillus brevis* for efficient synthesis and secretion of human epidermal growth factor. *Proc. Natl. Acad. Sci. USA* **86**:3589-3593.
44. 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.