Molecular Cloning and Sequence Analysis of a Gene Encoding an Extracellular Proteinase from *Lactobacillus helveticus* CP790

Naoyuki YAMAMOTO,† Tadashi SHINODA, and Toshiaki TAKANO

R&D Center, Calpis Co., Ltd., 11-10, 5-Chome, Fuchinobe, Sagamihara, Kanagawa 229-0006, Japan

Received December 16, 1999 Accepted February 11, 2000

A 2.6-kilobase *Hae*III DNA fragment corresponding to an extracellular proteinase gene (*prtY*) was cloned from chromosomal DNA of *Lactobacillus helveticus* CP790 in *Escherichia coli* using a pKK223-3 vector. The transformant expressed a 48-kDa protein that reacts with monoclonal antibodies specific to the proteinase and seemed to be a pro-proteinase, but had no proteolytic activity. About 1.6 kilobases of the 2.6-kilobase DNA fragment, which contained the complete gene for the proteinase was sequenced. Sequence analysis found an open reading frame with a capacity to encode a protein of 449 amino acids. The coding region contained a Gram-positive-type signal peptide of 30 amino acids. The *N*-terminal sequences of the proproteinase and the mature proteinase have been observed in the polypeptide at position +31 and +38. The putative amino acid sequence showed a significant similarity to a surface layer protein of *L. helveticus* and *Lactobacillus acidophilus* in the amino terminal signal sequence and carboxyl terminus.

**Key words:** extracellular proteinase gene; *Lactobacillus helveticus*; surface layer protein

Lactic acid bacteria are used in the food industry for production of a variety of fermented milk products. These bacteria produce proteolytic enzymes to hydrolyze milk proteins for their growth.1,2 The extracellular proteinase is important in the degradation of casein, which is a major milk protein. The extracellular proteinases have been well characterized and classified into two types, designated as PI and PIII on the basis of their cleavage specificity towards casein molecules.3 In most of the lactococci, the proteinase genes are located on the plasmids. A number of plasmid-located proteinase genes of lactococci have been cloned and sequenced.4,5 Detailed analyses found that two genes, designated *prtP* and *prtM*, are necessary for the expression of proteolytic activity. The *prtP* gene encodes a cell wall-associated serine proteinase of approximately 200-kDa, which is activated by a *prtM*-dependent maturation step.6,7 *Lactobacillus helveticus* has a strong proteolytic activity and shows a high casein hydrolysis activity in culture medium.8 An antihypertensive effect specific to the milk fermented by *L. helveticus* was reported.9 The extracellular proteinase (prtY) was purified from *L. helveticus* CP790, and was found to be a serine type enzyme with a molecular weight of 45,000.10 The active enzyme was generated from a 46-kDa precursor protein, by excision of seven amino acids from the amino terminus.11 A different type of a proteinase with a molecular weight of 170,000 was reported.12 These proteinases from *L. helveticus* were divided into two types13 by an immunological analysis. Recently, the gene corresponds to the 170-kDa proteinase was cloned and sequenced from *L. helveticus* CNRZ32.14 However, there is no genetic information on the 45-kDa proteinase of *L. helveticus*.

In this paper, we describe here the gene cloning by use of a monoclonal antibody specific to the 45-kDa extracellular proteinase, and the sequence of the enzyme from *L. helveticus* CP790. This report provides the first findings on the structure of the 45-kDa proteinase gene.

**Materials and Methods**

*Bacterial strains, plasmid, and media. Lactobacillus helveticus* CP790 was used from our stock culture.15 *Escherichia coli* strain HB101 was from Takara Shuzo (Kyoto, Japan). *E. coli* HB101 was used as the host for transformation test. A pKK223-3 vector was from Pharmacia Biotech (Tokyo, Japan). *L. helveticus* CP790 was grown in MRS broth (Difco Laboratories, Detroit, MI) at 37°C.

*Recombinant DNA techniques. General procedures for cloning and DNA manipulation were done essentially as described by Maniatis et al.15 Chromosomal DNA from *L. helveticus* was isolated ac-
According to the method of Leenhouts et al.[16] For genomic screening of the protease gene (prY), chromosomal DNA isolated from *L. helveticus* CP790 was partially digested with *Hae*III, and the resultant fragments were ligated into the Smal site of pKK223-3 expression vector. The ligation mixture was used to transform *E. coli* HB101. Enzymes used for DNA manipulations were purchased from Takara Shuzo.

**Immunological detection of positive clones.** Positive clones were screened from the library of the genome of *L. helveticus* CP790 as follows. The *E. coli* transformants were plated on LB agar plates[15] containing 30 µg/ml ampicillin (Sigma, St Louis, MO). The obtained transformants were replicated on nitrocellulose filters and cultured on LB agar plates. The colonies grown on the nitrocellulose filter were lysed by gently shaking in 0.2 M NaOH and 1% (w/v) SDS at room temperature for 10 min. The filters neutralized in 1 M Tris hydrochloride (pH 7.0) containing 0.5 M NaCl were shaken gently in 20 mM sodium phosphate buffer (pH 6.1) containing 0.15 M NaCl and 0.5% (w/v) casein for 30 min. The filters were tested for the immunoreactivity by using proteinase-specific mouse monoclonal antibodies.[11]

**Immunoblotting.** Proteinase of *L. helveticus* and or proteinase expressed in recombinant *E. coli* strains were detected by immunoblotting analysis. Cells of *L. helveticus* CP790 or recombinants *E. coli* transformants were disrupted by a Sonicator 5203 (Ohtake works) at 100 W. Cell debris were removed by centrifugation, and the supernatants were analyzed by 1 mm thick SDS-9%polyacrylamide gel electrophoresis (SDS-9%PAGE) by the method of Laemmli.[17] Proteinase was detected by using proteinase-specific mouse monoclonal antibodies after immunoblotting as described previously.[11]

**Nucleotide sequence analysis.** Nucleotide sequence was analyzed by the dideoxy chain termination method of Sanger et al.[18] with a sequencing kit (Amersham). Sequence-specific primers were used for part of the sequence reactions.

**Amino acid analysis.** The 46-kDa proproteinase was purified by the method described previously.[11] The purified protein was hydrolyzed under HCl vapor at 110°C for 20 h. The resulting amino acids were measured at 250 nm by reversed-phase high-pressure liquid chromatography (Jasco amino acid analyzer, Tokyo, Japan).

**Results**

**Immunological screening of positive clones**

Two positive clones were screened from about 10⁶ transformants, using a monoclonal antibody against the proteinase[11] as described in Materials and Methods, and designated *E. coli* EPI and EPII. Analysis of plasmid DNAs in two positive clones by 0.8% agarose gel electrophoresis, *E. coli* EPI and EPII, showed that the pKK223-3 vector contained a 2.6-kb and a 2.4-kb insert. From restriction fragment analysis of both insert DNAs, the 2.4-kb DNA seemed to be a part of 2.6-kb DNA. These plasmids obtained from the *E. coli* transformants were designated pEPI and pEPII respectively. In the immunoblotting analysis of the transformants, a 45-kDa protein band which corresponds to the proteinase was observed in the extract from *L. helveticus* CP790 as reported previously[11] (Fig. 1, lane 1). On the other hand, a 48-kDa protein was observed in the extracts of both transformant cells (lane 2 and 3). This suggests that the 48-kDa protein is the pre-proteinase which is expressed from the cloned gene in *E. coli*. In the transformant cells of *E. coli*, no proteolytic activity were detected in the extracts of the transformants (data not shown). The 2.6-kb *Hae*III DNA fragment was used for more detailed study.

**Sequence of the cloned DNA**

To sequence the 2.6-kb *Hae*III DNA fragment, pEPI plasmid was digested with *Bam*HI and recloned into *Bam*HI site of M13mp18 vector, and used to transform *E. coli* JM105. The complete gene for the proteinase of *L. helveticus* CP790 was sequenced in

![Fig. 1. Immunoblotting of the Cell Extracts of E. coli Transformants and L. helveticus CP790.](image)

Cells were disrupted by sonication and the proteins in the supernatant were solubilized with 10 mM Tris hydrochloride (pH 8.0) containing 1% SDS, subjected to SDS-9% PAGE, and transferred to a nitrocellulose membrane. Lane 1: *L. helveticus* CP790, lane 2: *E. coli* PEI, lane 3: *E. coli* PEPII. Proteinase was detected with monoclonal antibody LHP-1. The following molecular marker proteins were used: phosphorylase b (94 kDa), bovine serum albumin (67 kDa), ovalbumin (43 kDa), and carbonic anhydrase (30 kDa).
both directions, by using sequence-specific primers. An about-1.6 kb DNA fragment of the 2.6-kb DNA fragment shown in Fig. 2 was sequenced. Within this region, an open reading frame (ORF) was found. The ORF starts with an ATG start codon at position 93, and first stop codon (TAA) is located at position 1,442. The ORF could encode a polypeptide with a translated molecular weight of 47,795 (449 codons). The N-terminal sequences of the proproteinase and the mature proteinase analyzed previously were detected at positions +31 and +38 in the ORF respectively (Fig. 2). Polypeptide of 30 amino acids in N-terminus of the primary translation product had a consensus sequence of the signal sequence, and the cleavage between 37 and 38 would give a mature proproteinase as suggested in our previous report. Moreover, the amino acid composition of the deduced proproteinase from the DNA sequence showed similarity to the composition of the 46-kDa proproteinase described in Materials and Methods (Table 1). The typical feature of the amino acid composition is the large number of hydrophobic amino acids and amino acids with a hydroxyl group. Upstream of the ATG start codon, a putative Shine-Dalgarno ribosomal-binding site (AGGAGG) and −10 and −35 proposed promoter sites conserved in

![DNA Sequence](image)

**Fig. 2.** DNA Sequence of the Proteinase Gene of *L. helveticus* CP790 and deduced Amino Acid Sequence.

The N-terminal amino acid sequences of the proproteinase and the mature protein are double underlined, and the ribosome binding sites (RBS) are indicated. The cleavage site of the signal peptide is between amino acids 30 and 31, and of the pro-sequence between 37 and 38, as indicated by vertical arrows. The predicted −10 and −35 regions of promoters, and transcriptional terminator are indicated. Ser and Asp involved in the active sites are shown by boxes. For details, see the text. The sequence for the proteinase gene has been submitted to GenBank and assigned accession no. AB026985.
prokaryote are present. A putative transcription terminator sequence (ΔG, -17.3 kcal/mol) can be found downstream of the stop codon of the proteinase gene (Fig. 2).

**Similarity**

The putative proteinase of *L. helveticus* CP790 showed similarity with three proteins in the GenBank sequence database in September 1999. The deduced amino acid sequence of *L. helveticus* CP790 showed 56.1% identity in 449 amino acids of the surface layer proteins of CNRZ 89219 and CNRZ 1269,20 and 46.1% to a protein of *Lactobacillus acidophilus* ATCC 4356,20 respectively. The overlap was highest with the surface layer protein from *L. helveticus* CNRZ 892 and CNRZ 1269 out of three surface layer proteins, as shown in Fig. 3. The nucleotide sequence of surface layer protein from *L. helveticus* CNRZ 892 was identical to the sequence of CNRZ 1269. Surface layer protein is a major protein in the cell to maintain the cell shape.19,20 Extreme similarity to the sequence of CNRZ 892 was observed in N-terminus containing a signal sequence of the ORF, extending over 33 amino acid residues with 97.0% match. A second region of similarity, with 81.5% match over a stretch of 168 amino acids, is found in the C-terminus of the protein (Fig. 4).

From the similarity comparison with the reactive center regions of *L. helveticus* CNRZ32 proteinase (pH), *Bacillus subtilis* Carlsberg subtilisin, to the *L. helveticus* CP790 proteinase (pH), and *L. lactis* subsp. cremoris Wg2 proteinase (pH) over 10 amino acids, two regions, 68–77 and 96–105 which contains Ser and Glu, showed 40.0% and 40.0% similarities respectively (Fig. 5). However, there was no similarity to the His-containing active center region, and the active sites of the proteinase are still not clear.

**Table 1.** Amino Acid Composition of Pre-proteinase of *L. helveticus* CP790

<table>
<thead>
<tr>
<th>Amino acid</th>
<th>deduced (%)</th>
<th>determined (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gly</td>
<td>7.4</td>
<td>7.3</td>
</tr>
<tr>
<td>Ala</td>
<td>9.8</td>
<td>10.1</td>
</tr>
<tr>
<td>Val</td>
<td>10.5</td>
<td>10.2</td>
</tr>
<tr>
<td>Leu</td>
<td>5.7</td>
<td>6.5</td>
</tr>
<tr>
<td>Ile</td>
<td>2.4</td>
<td>3.0</td>
</tr>
<tr>
<td>Ser</td>
<td>7.1</td>
<td>7.6</td>
</tr>
<tr>
<td>Thr</td>
<td>12.4</td>
<td>12.1</td>
</tr>
<tr>
<td>Cys</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Met</td>
<td>1.2</td>
<td>0.8</td>
</tr>
<tr>
<td>Asp</td>
<td>6.9</td>
<td>14.5</td>
</tr>
<tr>
<td>Asn</td>
<td>7.9</td>
<td>—</td>
</tr>
<tr>
<td>Glu</td>
<td>3.3</td>
<td>4.0</td>
</tr>
<tr>
<td>Gln</td>
<td>1.4</td>
<td>—</td>
</tr>
<tr>
<td>Arg</td>
<td>1.4</td>
<td>2.4</td>
</tr>
<tr>
<td>Lys</td>
<td>11.2</td>
<td>10.2</td>
</tr>
<tr>
<td>His</td>
<td>0.7</td>
<td>0.6</td>
</tr>
<tr>
<td>The</td>
<td>2.9</td>
<td>3.0</td>
</tr>
<tr>
<td>Tyr</td>
<td>5.5</td>
<td>4.9</td>
</tr>
<tr>
<td>Trp</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Pro</td>
<td>2.2</td>
<td>2.9</td>
</tr>
</tbody>
</table>

* Predicted proproteinase as deduced from the DNA sequence of the proteinase gene (Fig. 2).

* As described in Materials and Methods.

1' MKRNLRIVSA AAAALLAVAP VAATAMPVNA ATTVTSSSNT NKPSTVLSG GDVSESDKGV
1' MKRNLRIVSA AAAALLAVAP VAATAMPVNA ATTVTSSSNT NKPSTVLSG DVTPSISAIA

61' NVTSPFPLTS EAKGDPPALQ GSIRASLGT STVADVREPA KPYTLZDGKG VAYYSDKNT
61' AVKESDTPMP --- AIDGGDPALQ GSIRASLGT STVADVREPA KPYTLZDGKG VAYYSDKNT

121' LTVKSLSDVK GDDYTMVESG VGPSFPGKGA GKVYTF---E LPEGVTFVEGA NNYKDDKVT
121' LTVKSLSDVK GDDYTMVESG VGPSFPGKGA GKVYTF---E LPEGVTFVEGA NNYKDDKVT

107' NTVKYABLEA DRAYTVTPD VSPHPGEHNA GKEITISSG PNVTPTKDGQ DQASTVTKV
107' NTVKYABLEA DRAYTVTPD VSPHPGEHNA GKEITISSG PNVTPTKDGQ DQASTVTKV

167' LQGDQVAKLIS SQVQKNVYAI DTVTNSHNVF YDVTGATVQ TAVSNIDN QOLINVTSSV
167' LQGDQVAKLIS SQVQKNVYAI DTVTNSHNVF YDVTGATVQ TAVSNIDN QOLINVTSSV

237' KALNDKTEAM QFRDSLFOVY KWNTADDVK AELEKAGKY DAANNFAPFD TTTTILMKM
237' KALNDKTEAM QFRDSLFOVY KWNTADDVK AELEKAGKY DAANNFAPFD TTTTILMKM

227' AANSSLYFAA QYDEKGLNVT TDDTFV---ATK DALKAKQIV KSGSFGKAPF TTFSTIKAIS
227' AANSSLYFAA QYDEKGLNVT TDDTFV---ATK DALKAKQIV KSGSFGKAPF TTFSTIKAIS

297' DVNGKTEALSP VTVTQYNGKES TPVPSQKSTI MINAYYODK AKEKQDZKTV RNAYTVTSMN
297' DVNGKTEALSP VTVTQYNGKES TPVPSQKSTI MINAYYODK AKEKQDZKTV RNAYTVTSMN

356' TTVKKGSTSY YS-IEGKAT GRYINADID GTRKLYKHA YYSKSKRA NIVDLLKGETE
356' TTVKKGSTSY YS-IEGKAT GRYINADID GTRKLYKHA YYSKSKRA NIVDLLKGETE

346' TTVKKGSTSY YS-IEGKAT GRYINADID GTRKLYKHA YYSKSKRA NIVDLLKGETE
346' TTVKKGSTSY YS-IEGKAT GRYINADID GTRKLYKHA YYSKSKRA NIVDLLKGETE

405' VTYGNYTPFY KNGKRYKIG ADKRYTVYR ENFD CP790 pHY
405' VTYGNYTPFY KNGKRYKIG ADKRYTVYR ENFD CP790 pHY

1' MKRNLRIVSA AAAALLAVAP VAATAMPVNA ATTVTSSSNT NKPSTVLSG GDVSESDKGV
1' MKRNLRIVSA AAAALLAVAP VAATAMPVNA ATTVTSSSNT NKPSTVLSG DVTPSISAIA

Fig. 3. Similarity Comparison of Amino Acid Sequences of the Proteinase of *L. helveticus* CP790 (upper) and the Surface Layer Protein of *L. helveticus* CNRZ32 (lower).
Dots indicate identical amino acids, and the horizontal lines indicate gaps. Ser and Asp involved in the active sites are shown by boxes.
Proteinase Gene of Lactobacillus helveticus

411

Fig. 4. Sequence Similarity Comparison of Proteinases and Surface Layer Protein.
CNRZ32 prtH: the proteinase (prtH) reported from L. helveticus CNRZ32, CP790 prtY: the proteinase (prtY) from L. helveticus CP790, and CNRZ892 SLP: the surface layer protein (SLP) from L. helveticus CNRZ 892.

Fig. 5. Similarity Comparison of Active Centers of Proteinases.
Sequence similarity of the proteinases from L. helveticus CP790 prtY, Bacillus subtilis Carlsberg subtilisin, L. lactis subsp. cremoris Wg2 prtP, and L. helveticus CNRZ32 prtH are shown. Identical residues are shown by white, Ser and Asp involved in the active sites are shown by boxes.

Discussion

Recently, two genes encoding extracellular proteinases were cloned from chromosomal DNA of Lactobacillus paracasei subsp. paracasei NCDO 1512 and L. helveticus CNRZ3224 based on a Lactococcus lactis subsp. cremoris SK11 proteinase sequence.13 First, we also tried to find if the lactococcus proteinase gene kindly provided by Dr. Jan Kok (University of Groningen, The Netherlands) hybridized to L. helveticus CP790 chromosomal DNA in Southern hybridization analysis. However, there was no homologous signal in the Southern hybridization analysis (data not shown). So, the proteinase gene was cloned from chromosomal DNA of L. helveticus CP790 by use of the monoclonal antibodies specific to the extracellular proteinase.11) The monoclonal antibody reacted to a 48-kDa gene product in an E. coli transformant carrying pEP1 or pEP2 which contained 2.6-kb or 2.4-kb DNA fragments. Amino terminal pro-sequence and mature sequence were observed in the putative amino acid sequence of the 2.6-kb DNA sequence. Moreover, the amino acid composition of the deduced pro-sequence resembled the amino acid composition of the proteinase in this study. From these results, we concluded the gene

in the cloned 2.6-kb DNA is a gene for the extracellular proteinase of L. helveticus CP790.

Protein and nucleic acid similarity searches using the SWISSPROT and GenBank/EMBL data bases found a high rate of similarity of the proteinase to the surface layer proteins of L. helveticus21 and L. acidophilus22 in both of the C-termini. The surface layer protein was also similar to L. helveticus CNRZ32 proteinase23 as shown in Fig. 4. The surface layer proteins have so far been reported in several bacteria,19-23 and exist attaching to the cell wall as a major protein in the surface layer. These C-terminal sequences of surface layer proteins of lactobacilli are thought to be responsible for cell-wall attachment,24 but the role of the surface layer protein is not fully known. The relationship between the proteinase of L. helveticus CP790 and the surface layer proteins is still not clear.

Recently, a maturation factor that accelerates the maturation process, that is conversion of the 46-kDa proproteinase to the 45-kDa mature enzyme, was suggested in L. helveticus CP790.20 There are some important factors that will affect the expression of the proteinase activity in the transformant cell. First, the expression of the maturation factor may be needed for the conversion of pro-enzyme to the active enzyme. Second, the translocation of the pro-enzyme to outer membrane before the activation may be necessary to prevent limited growth of the E. coli transformant. Third, excision of the signal peptide may affect the efficiency of the conversion of the proproteinase to the activated enzyme. Considering these difficulties in expressing the proteinase activity in E. coli transformant cell, gene disruption of the proteinase gene by introduction of the deletant gene of the proteinase to L. helveticus CP790 strain is now being addressed to confirm the function of the gene product.

Acknowledgments

We thank Prof. Kazuo Shishido (Tokyo Institute of Technology) for helpful suggestions and Naomi Kaya for technical assistance for this study. We wish to thank Dr. Jan Kok for provision of the lactococcus proteinase gene.

References

4) Kiwaki, M., Ikemura, H., Shimizu-Kadota, M., and


