Cloning, sequencing and expression of the gene encoding the cell-envelope-associated proteinase from *Lactobacillus paracasei* subsp. *paracasei* NCDO 151

ASKILD HOLCK* and HELGA NÆS

MATFORSK, Norwegian Food Research Institute, Oslouv. 1, N-1430 Ås, Norway

(Received 18 November 1991; revised 10 February 1992; accepted 16 March 1992)

The gene encoding the cell-envelope-associated proteinase of *Lactobacillus paracasei* subsp. *paracasei* NCDO 151 (formerly *Lactobacillus casei* NCDO 151) was cloned and sequenced. The gene was located on the chromosome and encoded a polypeptide of 1902 amino acids. The proteinase is N-terminally cleaved upon maturation. It shows extensive homology to the *Lactococcus lactis* subsp. *cremoris* Wg2 proteinase. Similar to the situation in *Lactococcus*, a maturation gene was found upstream of the proteinase gene. The cloned proteinase gene was expressed in *Lactobacillus plantarum*. However, no expression was observed when the gene was cloned in *Lactococcus lactis*.

Introduction

The lactic acid bacteria are of great economic importance in the dairy industry. They contribute to the flavour and texture of fermented milk products (Fox, 1989; Law & Kolstad, 1983; Thomas & Mills, 1981). Much is known about the proteolytic systems of the lactococci. Several of the proteinases have been purified and their genes sequenced (for a review see Kok, 1990). The genes characterized so far appear to be located on plasmids.

Strains of *Lactobacillus casei* very often occur in hard and semihard cheeses as adventitious bacteria and in many cases contribute to flavour development during the ripening process. Consequently the proteolytic activity of *Lb. casei* is of interest to the dairy industry. *Lb. casei* has been tested for use in mixed starter cultures and shown to contribute to the production of superior-quality cheese (Girgis et al., 1983; Bianchi-Salvadori & Sacco, 1981; Ramos et al., 1981). There has also been interest in employing heat-shocked cells or cell lysates of *Lb. casei* to accelerate cheese ripening (El Abboudi et al., 1991; Trepanier et al., 1991).

Knowledge of the proteolytic systems of the lactobacilli is limited. Some reports on characterization and partial purification of *Lactobacillus* proteinases have been published (Argyle et al., 1976; Ezzat et al., 1985, 1988; Khalid & Marth, 1990; El Soda et al., 1986; Zevaco & Gripon, 1988). We have purified to homogeneity and characterized a cell-envelope-associated proteinase of *Lactobacillus paracasei* subsp. *paracasei* NCDO 151 (formerly named *Lactobacillus casei* NCDO 151) (Næs & Nissen-Meyer, 1992). This serine proteinase, designated Lp151, shared some properties with known lactococcal proteinases. Very little is known about the *Lactobacillus* proteinase genes. By Southern blotting a gene with homology to the *Lactococcus lactis* subsp. *cremoris* SK11 proteinase gene has been observed in *Lactobacillus lactis* (Kok & Venema, 1988). Kojic et al. (1991) recently reported the physical map of a proteinase gene from *Lb. casei* HN14 with similarity to the lactococcal genes.

Here we describe the cloning, sequencing and expression of the *Lb. paracasei* NCDO 151 proteinase gene. Knowledge of the genetics of the proteolytic system in lactobacilli should make it possible to construct strains with genetic characteristics important to the food fermentation industry.

Methods

**Bacterial strains and plasmids.** *Lactobacillus paracasei* subsp. *paracasei* NCDO 151 (formerly *Lactobacillus casei* NCDO 151) was obtained from the INRA Centre, Jouy-en-Josas, France. *JEML3*, *λgt10*, λ-
Nucleotide sequencing. Nucleotide sequence analysis was performed by the dideoxy chain-terminating method on single-stranded M13 DNA or directly on denatured plasmid DNA (Kraft et al., 1988). Both strands were sequenced by primer walking. Oligonucleotide sequencing primers were synthesized on an Applied Biosystems synthesizer, model 381A. The computer analyses were performed on an IBM personal computer employing the DNASIS sequence analysis program (Pharmacia) and on a microVAX 3400 computer employing the GCG program package (Devereux et al., 1984).

Pulsed-field gel electrophoresis. The bacterial cells were cast and lysed within blocks of agarose (inserts) according to Lillegaard et al. (1991). Lysis was performed with 4 mg lysosome ml⁻¹ and 400 U mutanolysin ml⁻¹. Restriction enzyme digestion was executed overnight. Pulsed-field gel electrophoresis was performed on a CHEF-DR II (Bio-Rad). Electrophoreses were run under two different conditions in a 1:1% (w/v) agarose gel (in 22 mM-Tris-base, 22 mM-boric acid, 0.5 mM-EDTA) at 14 °C: either 200 V, 18 h, pulse time increasing from 15 s to 70 s, or 200 V, 24 h, pulse time 1 s to 200 s, according to the manufacturer's instructions.

Results

Identification and chromosomal location of the proteinase gene

To establish whether there was any similarity between the Lb. paracasei NCDO 151 proteinase gene and the known lacticoccal proteinase genes, total DNA from Lb. paracasei NCDO 151 was subjected to restriction enzyme analysis and hybridized to a probe containing the central BamHI–EcoRI fragment of the Lc. lactis subsp. cremoris SK11 proteinase gene. The probe gave one unique signal, indicating the presence of a single-copy Lb. paracasei gene with homology to the lacticoccal SK11 gene (Fig. 1). This was confirmed by the restriction pattern obtained from several Southern blots (not shown). Similar experiments were performed with a prtM-specific probe from Lc. lactis Wg2 (the HindIII–ClaI fragment), and a Lb. paracasei prtM gene was detected close to the proteinase gene.

Lb. paracasei NCDO 151 harbours two plasmids. Neither of the plasmids hybridized to the lacticoccal probes (not shown). To determine whether the proteinase gene was chromosomally located or resided on a large plasmid which co-purified with the chromosomal DNA rather than with the plasmid DNA fraction, the DNA was subjected to pulsed-field gel electrophoresis under various conditions (Fig. 2). The proteinase apparently resided on a 120 kb NorI fragment and a SfiI fragment of 100 kb. Moreover, the proteinase gene remained with these fragments, indicating that the proteinase gene was chromosomally located.

Lc. lactis Wg2 and SK11 each contain the insertion sequence ISS7/ downstream of both prtM and prtP...
Sequence of Lactobacillus proteinase gene

Cloning and sequencing of the \textit{prtP} and \textit{prtM} genes

A \textit{Lb. paracasei} NCDO 151 \textlambda EMBL3 library was constructed as described in Methods and \(3 \times 10^4\) plaques were screened with the central \textit{BamH}I–\textit{EcoRI} fragment of the \textit{Lc. lactis} SK11 proteinase gene as a probe. DNA from various positive clones was isolated and characterized. Various fragments containing the 5' and 3' regions of the proteinase gene were subcloned into M13 phage vector, or into pUC19 or pGEM-7zf(+) plasmid vectors. None of the clones, however, contained the maturation gene \textit{prtM}. A \textit{Lb. paracasei} \textit{Agtl}0 DNA library was screened with a \textit{ClaI}–\textit{HindIII} fragment containing the \textit{prtM} gene of \textit{Lc. lactis} Wg2. A 4.0 kbp \textit{EcoRI} fragment from one of the positive clones was subcloned into pGEM-7zf(+). Appropriate subclones containing the \textit{prtP} and the \textit{prtM} genes were sequenced. The sequence obtained was very similar to the sequence of the \textit{Lc. lactis} Wg2 proteolytic system, but apparently carried a deletion in the promoter region common to \textit{prtM} and \textit{prtP}. A 6.7 kbp chromosomal HindIII fragment of \textit{Lb. paracasei} NCDO 151 containing both the \textit{prtP} and \textit{prtM} genes in pVS2 was obtained after screening a partial chromosomal library of HindIII fragments cloned in \textit{Lb. plantarum} NC8. This construction was named pHN1. Upon sequencing of the HindIII fragment, the deletion in the \textit{Agtl}0 clone was shown to comprise 14 base pairs encompassing the proposed –10 region of the \textit{prtP} gene. The complete sequence of both the \textit{prtP} and the \textit{prtM} genes is presented in Fig. 3. Two divergent open reading frames, ORF1 and ORF2, were detected, encoding polypeptides of 299 and 1902 amino acids, respectively. The N-terminal sequence of the mature proteinase has been determined by Edman degradation to be X-A-K-A-N-S-M-A-N (Næs & Nissen-Meyer, 1992); this indicates the cleavage of the precursor proteinase at position +187 in the nascent polypeptide chain encoded by ORF2. This cleavage would yield an active proteinase, unless otherwise modified, of 1715 amino acid residues and an \(M_r\) of 180659. The smaller ORF1 encodes the proposed maturation protein. A typical signal sequence of lipoproteins (von Heijne, 1989) was found at the N-terminus of the primary translation product. A presumed cleavage between amino acid 23 and 24 would give a polypeptide of 276 amino acid residues and an \(M_r\) of 30675.

The restriction pattern for the \textit{prtP} and the \textit{prtM} genes obtained after sequencing indicated that the \textit{Lb. paracasei} NCDO 151 proteinase was very similar to the recently identified \textit{Lb. casei} HN14 proteinase (Fig. 4).

When comparing the amino acid sequences of the polypeptides, the Wg2 and SK11 mature proteinases were found to differ from the \textit{Lb. paracasei} proteinase in 58 (96.6% homology) and 77 (95.5% homology) amino acids, respectively (Fig. 5). When comparing the mature \textit{prtM} polypeptides, the homology to the lactococcal maturation peptides was found to be 96.0%.

The DNA sequence homology between the \textit{Lb. paracasei} and the lactococcal proteinase genes ended abruptly just downstream of both the \textit{prtP} and \textit{prtM} genes. Thus the palindromic sequences downstream of the translation stop signal, which may be involved in transcription termination, showed no similarities with the corresponding sequences of the lactococcal counterparts.

Putative Shine–Dalgarno ribosome-binding sites and –10 and –35 proposed promoter sites were located.
A. Holck and H. Næs

Fig. 2. Southern blot analysis of total *Lb. paracasei* NCDO 151 DNA employing pulsed-field gel electrophoresis. (a) Ethidium-bromide-stained agarose gel, 200 V, 18 h, pulsetime 15–70 s; (b) autoradiography after blotting and hybridization with the *prtP*-specific probe. U, undigested DNA; N, *NcoI*-digested DNA; S, *SfiI*-digested DNA. \( \lambda \), lambda ladder; one \( \lambda \) unit represents 48.5 kbp.

upstream of both the *prtP* and *prtM* genes. This region also contained several palindromic sequences, two of which encompassed the two suggested -35 regions. These sequences, being able to form hairpin loops, may be involved in transcription regulation of the genes of the proteolytic system.

**Expression of the cloned proteinase**

No cross-hybridization was observed when *Lb. plantarum* NC8 DNA was hybridized to a *prtP* specific probe. The weakly proteolytic *Lb. plantarum* NC8 used for expression of the cloned proteinase thus contained a proteinase very different from that of *Lb. paracasei* (not

<table>
<thead>
<tr>
<th>Strain</th>
<th>Clotting time (d)*</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Lb. plantarum</em> NC8(pHN1)</td>
<td>4 ± 1</td>
</tr>
<tr>
<td>transformant 6</td>
<td></td>
</tr>
<tr>
<td><em>Lb. plantarum</em> NC8(pHN1)</td>
<td>4 ± 1</td>
</tr>
<tr>
<td>transformant 47</td>
<td></td>
</tr>
<tr>
<td><em>Lb. plantarum</em> NC8(pVS2)</td>
<td>12 ± 2</td>
</tr>
<tr>
<td><em>Lb. plantarum</em> NC8</td>
<td>12 ± 2</td>
</tr>
<tr>
<td><em>Lc. lactis</em> MG1363(pHN1)†</td>
<td></td>
</tr>
<tr>
<td><em>Lc. lactis</em> MG1363</td>
<td>–</td>
</tr>
<tr>
<td><em>Lc. lactis</em> WG2</td>
<td>1 ± 0</td>
</tr>
</tbody>
</table>

* Experiments were performed twice, with two replicates each time. The results are means ± range. –, No clotting observed.
† Eight transformants containing the proteinase gene were tested.
Fig. 3. Nucleotide sequence of the pratP and the pratM genes. The antisense strand (mRNA-like) is shown for the pratP gene, whereas the sense strand is shown for the pratM gene. Restriction sites are indicated by horizontal lines above the sequence. Proposed promoter sites and ribosome-binding sites are underlined. Dyad symmetries are indicated by convergent arrows, with the centres of symmetry shown by dots. The upward-pointing arrow indicates the cleavage site to yield the mature proteinase. The downward pointing arrow indicates the cleavage site of the proposed signal sequence of the maturation polypeptide. Amino acids comprising the active site of the serine proteinase are boxed. The deduced amino acid sequences were compared to those of Lc. fuctis subsp. cremoris Wg2 pratP and pratM. Different amino acids of the Wg2 polypeptides are shown below the amino acid sequences of the Lb. paracasei polypeptides. Downward pointing arrowheads mark the beginning and the end of DNA sequence homology. The broken line indicates amino acids determined by N-terminal sequencing.

**HindIII**

6841

7081

7201

7321

7441

7561

7681

7801

1340
Sequence of Lactobacillus proteinase gene

(a) C BH CC BA E HH BA
(b) C H C BA E B E H C
(c) EH B C BA E B E BH C E

Fig. 4. Restriction enzyme maps of regions containing the proteinase and maturation genes of (a) L. lactis subsp. cremoris Wg2, (b) Lb. casei HN14 and (c) Lb. paracasei NCDO 151. The divergent arrows mark the positions of the \textit{prtP} and \textit{prtM} genes. The restriction sites for (a) and (c) were determined by sequence analysis; those for (b) are from Kojic \textit{et al.} (1991). C, CfaI; B, BgfII; H, HindIII; BA, BamHI; E, EcoR1.

Fig. 5. Comparison of the amino acid sequences of the primary transcripts of the proteinase genes from (a) Lb. paracasei NCDO 151, (b) L. lactis Wg2, (c) L. lactis NCDO 763 and (d) L. lactis SK11. Differences between the \textit{Lactobacillus} and the lactococcal proteinases are shown by vertical lines. The large horizontal box indicates the Lp151 proteinase with the signal sequence (SS, stippled) and the pro-sequence (PRO, hatched). Black boxes mark regions with homology to subtilisins. Active-site amino acids Asp, His and Ser are indicated. Horizontal lines at the C-terminus mark the transmembrane 'anchoring' site. The horizontal arrow indicates the C-terminal duplication of the SK11 proteinase. The bar (top left) represents 100 amino acids.

shown). Plasmid pHN1, containing the 6.7 kp HindIII fragment with the entire \textit{prtM} gene and a slightly truncated \textit{prtP} gene, was transformed into the plasmid-free \textit{Lb. plantarum} NC8 and the proteolytically negative \textit{Lc. lactis} MG1363. Untransformed cells, and cells transformed with pVS2 vector or pHN1 were tested for activity in milk-clotting experiments (Table 1). The results showed a marked increase in proteolytic activity of \textit{Lb. plantarum} NC8 containing pHN1, whereas no expression of \textit{prtP} was observed in \textit{Lc. lactis} MG1363.

**Discussion**

When proteinase was extracted from the cell wall of \textit{Lb. paracasei} NCDO 151 and purified, two distinct fractions of proteolytic activity were obtained (Næs \& Nissen-Meyer, 1992). They represented two proteins, of \(M_r\) 135000 and 110000, with an identical N-terminal amino acid sequence. This N-terminal sequence showed homology to the known lactococcal proteinases (Kiwaki \textit{et al.}, 1989; Kok \textit{et al.}, 1988; Nissen-Meyer \& Sletten, 1991; Vos \textit{et al.}, 1989a). The hybridization of total \textit{Lb. paracasei} NCDO 151 DNA clearly indicated the presence of only one chromosomal gene with homology to the lactococcal proteases. This supports the previous conclusion that the smaller protein is a cleavage product of the larger (\(M_r\), 135000) proteinase. Catalytically active cleavage products of homologous lactococcal proteinases have been described by Nissen-Meyer \& Sletten (1991). Ezzat \textit{et al.} (1988) have reported a further cell-wall-bound proteolytic activity of \textit{Lb. paracasei} NCDO 151. It is still not known whether this is another cleavage product of the Lp151 proteinase or represents a totally different proteinase.

The proteinase genes of the lactococci described so far appear to reside on plasmids. When \textit{NorI}-digested total DNA from \textit{Lb. paracasei} NCDO 151 was subjected to pulsed-field gel electrophoresis, the proteinase gene appeared on a fragment of approximately 120 kbp. Such large plasmids were not observed in \textit{Lb. paracasei} NCDO 151. Plasmids are known to show varying apparent \(M_r\),
values when subjected to pulsed-field gel electrophoresis, depending on the conditions used (S.-E. Birkeland, personal communication). For *Lb. paracasei* NCDO 151 the apparent fragment size remained unchanged, regardless of the electrophoretic conditions employed. Moreover, the *prtP*-specific probe hybridized to the undigestioned chromosomal DNA. These experiments strongly indicated that the *prtP* and *prtM* genes were chromosomally located. A recent report by Kojic et al. (1991) indicates a chromosomally located proteinase gene in *Lb. casei* HN14. Judged by the restriction map, the proteinase genes of *Lb. casei* HN14 and *Lb. paracasei* NCDO 151 appear to be very similar, but not identical.

Homologous insertion sequences or parts of insertion sequences are present downstream of both the lactococcal *prtP* and *prtM* genes (Haandrikman et al., 1990). The lactococcal genes thus appear to have been part of the same transposon. This is reflected in the sequence homology outside the coding regions of these genes. The *prtP* and *prtM* genes of *Lb. paracasei* NCDO 151 do not belong to this transposon-like structure. This was verified both by the lack of hybridization to the ISS/W element of the Wg2 proteolytic system and by the abrupt ending of the DNA sequence homology with lactococcal *prtM* and *prtP* downstream of the coding regions. It cannot be excluded, however, that the genes may be flanked by other IS elements with little homology to ISSI.

Sequence comparisons of homologous proteins are important in determining the amino acids responsible for the biochemical and biophysical properties of the proteins. Very few amino acid changes can lead to relatively large changes in protein properties. By construction of Wg2-SK11 hybrid proteinases it has been shown that perhaps as few as seven amino acids were responsible for the conversion of the PI proteinase to a PIII proteinase (Vos et al., 1991). For the distantly related alkaline serine proteinases, the subtilisins, it has been shown that only three amino acid changes may lead to an increase of 7°C in the denaturation temperature (Narhi et al., 1991). The deduced amino acid sequence of the *prtP* gene of *Lb. paracasei* NCDO 151 showed a high degree of similarity to the known sequences of the lactococcal proteinases, and much of what is known about the lactococcal proteinases is obviously valid for the *Lactobacillus* proteinase. It is a serine proteinase (as was indicated by its inhibition by diisopropyl fluorophosphate and phenylmethylsulphonyl fluoride (Næs et al., 1991). It is first synthesized as a pre-pro-proteinase which is cleaved upon maturation, as was indicated by the N-terminal sequence determination (Næs & Nissen-Meyer, 1992) and it contains the C-terminal hydrophobic anchor sequence that was reported for the SK11 proteinase (Vos et al., 1989a). Despite these similarities there are distinct differences between the properties of the *Lactobacillus* proteinase and the lactococcal proteinases. For example, when comparing the Lp151 and the 763 proteinases, the Lp151 proteinase showed a lower pH optimum for degrading casein (pH 5.6 vs pH 6.0-6.3) and a lower sensitivity to Cu²⁺ (28% vs 100% inhibition at 1 mM-Cu²⁺) (Næs et al., 1991; Monnet et al., 1987). The *Lactobacillus* proteinase showed a higher degree of sequence similarity to the Wg2 and the 763 proteinases than to the SK11 proteinase (Kiwaki et al., 1989). According to the sequence data, the Lp151 proteinase appears to be more closely related to the PI proteinases (e.g. Wg2), hydrolysing primarily β-casein, than to the PIII proteinases (e.g. SK11) hydrolysing x-, β- and κ-caseins. The *Lb. paracasei* NCDO 151 proteinase differs from the Wg2 proteinase in two positions regarded as important for substrate specificity and binding, namely at position 142, where Ala was found as in the SK11 proteinase, and at position 747, where Glu was found. Also the Lp151 proteinase has retained Tyr31 and Tyr112 as in the subtilisins Carlsberg and DY (see Kok, 1990). The influence of these findings on the substrate binding and specificity is at present not known. One would perhaps expect to find differences in the digestion pattern of β-caseins between the Lp151 and Wg2 proteinases on the one hand and the 763 proteinase on the other. Even though the Wg2 and the 763 proteinases are 99% similar, the 763 proteinase carries the Arg747 and Lys748 of the SK11 proteinase which are important for substrate binding and specificity (KiwaI et al., 1989; Vos et al., 1991). The amino acid differences between the *Lactobacillus* and the lactococcal proteinases appear not to be randomly scattered along the polypeptide chain (Fig. 5). Seemingly, there are long sequences where little change has occurred. The few changes found in these regions are predominantly conservative amino acid changes. The importance of these regions remains obscure. The signal sequences and the prosequences also seem to be highly conserved.

The nascent polypeptide of the *Lactobacillus* proteinase is cleaved at position 187. This would give a polypeptide of *M* = 180659.

Less is known about the PrtM protein. It is a membrane-bound lipoprotein (Haandrikmann et al., 1991a) with a typical leader sequence of bacterial lipoproteins (von Heijne, 1989). It is somehow involved in the maturation of the proteinase (Haandrikmann et al., 1991b; Vos et al., 1989b). No obvious homology to other proteins was found when searching the Swiss protein database (release 19.0) with the PrtM amino acid sequence. The known lactococcal *prtM* genes are virtually identical, and the amino acid changes between the *Lactobacillus* and the lactococcal *prtM* proteins appear to be randomly spread along the polypeptide
chain. Downstream of the prtP gene, outside the region of homology to the lactococcal genes, a palindromic sequence followed by a stretch of U residues was found. This may act as a transcription-terminating signal (Rosenberg & Court, 1979). This would give a prtP mRNA of about 5860 nucleotides, which corresponds well with that observed by Northern blotting (not shown). The region upstream of the prtP gene contained putative non-overlapping promoter sequences for both the prtP and the prtM genes typical of E. coli vegetative promoters (Rosenberg & Court, 1979). For both the prtP and prtM genes the putative –35 sequences are located within palindromic sequences, which may thus be involved in regulation of transcription. The promoter region of the prtP and prtM genes carried a deletion of 30 and 34 bp relative to the Wg2 and the SK11 promoter region, respectively. The lactococcal prtM gene possesses two proposed promoter sites. The sequence corresponding to the more upstream –10 sequence of the lactococcal prtM gene (Kok et al., 1988) is absent in the Lactobacillus promoter. The observed lack of proteolytic activity in Lc. lactis MG1363 may be due to the loss of the upstream prtM promoter.

We wish to thank Birgitta Baardsen for excellent technical assistance. This work was supported by Gilde Norge Ans.

References


