Molecular cloning and nucleotide sequence of the gene encoding a calcium-dependent exoproteinase from *Bacillus megaterium* ATCC 14581

Sabine Kühn and Peter Fortnagel *

Universität Hamburg, Institut für Allgemeine Botanik, Abteilung Mikrobiologie, Ohnhorststraße 18, 2000 Hamburg 52, Germany

(Received 27 May 1992; revised 22 September 1992; accepted 25 September 1992)

The gene nprM encoding the calcium-dependent extracellular proteinase from *Bacillus megaterium* ATCC 14581 was cloned in the vector pBR322 and expressed in Escherichia coli HB101. The DNA sequence of the cloned 3.7 kb fragment revealed only one open reading frame consisting of 1686 bp with a coding capacity of 562 amino acid residues. A predicted Shine–Dalgarno (SD) sequence was observed 9 bp upstream from the presumptive translation start site (ATG). A possible promoter sequence (TAGACG for the –35 region and TATAAT for the –10 region) was found about 69 bp upstream of the ATG start site. The deduced amino acid sequence exhibited a 24 amino acid residue signal peptide and an additional polypeptide ‘pro’ sequence of 221 amino acids preceding the putative mature protein of 317 amino acid residues. Amino acid sequence comparison revealed 84.5% homology between the mature protein and that of a thermolabile neutral protease from *B. cereus*. It also shares 73% homology with the thermostable neutral proteases of *B. thermoproteolyticus* and *B. stearothermophilus*. The zinc-binding sites and the catalytic residues are completely conserved in all four proteases. NprM has a temperature optimum of 58°C, a pH optimum of between 6.4 and 7.2, and is stimulated by calcium ions and inhibited by EDTA. These results indicate that the enzyme is a neutral (metallo-) protease.

Introduction

Bacterial species of the genus *Bacillus* secrete a variety of enzymes, some of which are of commercial importance, particularly α-amylases and proteinases (Debabov, 1982). In the past, many protease genes from representatives of the genus have been cloned and characterized. For example, *Bacillus subtilis* secretes many different extracellular proteinases into the culture medium, especially at the end of the exponential growth phase (Stahl & Ferrari, 1984; Wong et al., 1984; Yang et al., 1984; Rufo et al., 1990; Sloma et al., 1990; Wu et al., 1990; Tran et al., 1991). The neutral proteases from *B. stearothermophilus* (Fuji et al., 1983; Takagi et al., 1985; Kubo et al., 1988; Kubo & Imanaka, 1988; Nishiya & Imanaka, 1990), *B. thermoproteolyticus* (thermolysin) (Latt et al., 1969; Stauffer, 1971; Titani et al., 1972; Matthews et al., 1972), *B. cereus* (Sidler et al., 1986; Pauptit et al., 1988), and *B. caldolyticus* (van den Burg et al., 1991) have been purified and the genes cloned and characterized. All of these proteases are synthesized in the form of ‘prepro’ enzymes.

Relatively little is known about proteases produced by *B. megaterium*. An intracellular protease initiates the extensive protein degradation that occurs during the early stages of spore germination (Loshon & Setlow, 1982). Another protease from *B. megaterium*, known as megateriopeptidase (Millet, 1968), is a hydrophobic aminopeptidase. It is the only known extracellular endopeptidase produced and excreted by *B. megaterium*. This protease is synthesized during both growth and sporulation. The formation of the enzyme during growth is repressed by free amino acids and glucose or acetate (Chaloupka, 1969; Millet & Aubert, 1969). Its synthesis during sporogenesis is rather insensitive to repression (Chaloupka et al., 1982).

Megateriopeptidase is most active at pH 7.2, is stabilized by calcium ions, and is inactivated by o-phenanthroline. These results classify it as a neutral (metallo-) proteinase. Its specificity for a series of dipeptide substrates is similar to that of thermolysin (Millet & Archer, 1969). Other similarities with thermo-
lysin, including a partial crossreaction with antithermolysin serum which causes loss of enzyme activity with the precipitin reaction, have already been described (Keay et al., 1971). In this paper, the cloning of the gene for a *B. megaterium* extracellular metalloproteinase is described, together with its characterization, and some properties of the gene product.

**Methods**

**Bacterial strains and cloning vectors.** The *B. megaterium* type strain, ATCC 14581 (Skerman et al., 1980), was used as a source of the chromosomal DNA. *E. coli* HB101 (Boyer & Roulland-Dussoix, 1969) and *E. coli* JM101 (Messing, 1983) were used as hosts for cloning and sequencing experiments. The shuttle plasmid pJK302, consisting of the vectors pBR322 (4.36 kb) and pBC16-1 (2.84 kb) fused at the EcoRI restriction sites (Kreft & Hughes, 1982), and the plasmids pBR322 (Bolivar et al., 1977), pBI30, and pBI31 (IBI) were used as vectors.

**Media and growth conditions.** *E. coli* strains harbouring plasmids were cultivated in LB broth supplemented with ampicillin at 37 °C for 6 h. Cells from 200 ml of the culture were harvested by centrifugation (9800 g, 20 min at 4 °C) and were suspended in 10 ml 10 mM-Tris/HCl pH 7.4. The cell suspension was disrupted in a French press at 20000 p.s.i. at 4 °C and the cell debris removed by centrifugation (12000 g, 20 min at 4 °C). The protease activity of the supernatant was assayed with azocasein (10 mg ml⁻¹) (Sigma). Samples (0.5 ml) of the diluted enzyme extract and 5 mg azocasein were incubated for 20 min at 37 °C, 1.5 ml TCA (5%, v/v) was added and the mixture was incubated for 30 min at 37 °C. The denatured proteins were removed by centrifugation (16000 g, 2 min). The absorbance of the supernatant was measured at 405 nm. Cultures of *E. coli* containing pBR322 served as control.

**Assay of neutral protease activity.** *E. coli* transformants carrying the recombinant plasmid DNA were cultivated in LB broth supplemented with ampicillin at 37 °C for 6 h. Cells from 200 ml of the culture were harvested by centrifugation (9800 g, 10 min at 4 °C) and were suspended

**Results**

**Cloning of nprM in *E. coli* HB101**

Among several thousand ampicillin-resistant *E. coli* HB101 transformants, one was found to liberate protease activity into the medium, as detected by a clear zone around the colony on skim milk nutrient plates. This protease-producing transformant contained a 24 kb recombinant plasmid, consisting of the vector pJK302 and a 17 kb insert. By shotgun cloning this 17 kb insert was reduced to 3.7 kb (pSK44), during which treatment the component of pBC16-1 from the plasmid pJK302 was lost. Therefore pSK44 consisted of the vector pBR322 and an *EcoRI/EcoRV* insert coding for *nprM* (Fig.1).

**Characterization of NprM**

To confirm the cloning of a neutral protease gene, properties of the cloned protease were determined in crude cell extracts. The protease activity was inhibited

---

**Fig. 1.** Restriction map of the plasmid pSK44 consisting of the vector pBR322 (thin line) and the cloned *EcoRI/EcoRV* fragment from genomic DNA of *B. megaterium* ATCC 14581 (thick line). The arrows indicate the regions which are essential for expression of ampicillin resistance and protease activity.
Fig. 2. Calcium activation of NprM in crude cell extracts of *E. coli/pSK44*. Protease activity is in arbitrary units (OD₆₅₀/20 min). ○, Crude cell extract from *E. coli/pSK44*; ●, crude cell extract from *E. coli/pBR322* (control).

Fig. 3. Calcium-induced cell lysis during growth of *E. coli/pSK44*. Calcium chloride concentration (mM): ○, 0.05; ●, 0.5; ▼, 1.0; □, 2.0; and ■, 3.0.

Fig. 4. DNA-sequencing strategy. Subclones of the EcoRI/EcoRV fragment in pIBI30 or pIBI31 were used for sequencing. Thick line: ORF of the *nprM* gene. The arrows represent the directions and the extents of the sequences.

**B. megaterium protease gene**

The nucleotide sequence of *nprM* and its flanking regions was determined. The sequencing strategy is depicted in Fig. 4. The sequence revealed a single large ORF consisting of 1686 nucleotides, with a coding capacity for 562 amino acid residues (Fig. 5). A putative Shine–Dalgarno (SD) sequence (AGGGATAGGA) was found 9 bp upstream from a likely translation start site (ATG) (Moran et al., 1982). The calculated free energy of binding (ΔG) of this ribosome-binding site (Tinoco et al., 1973) with the 3′ end of the 16S rRNA from *B. subtilis* is −74.4 kJ mol⁻¹. This ΔG value as well as the distance between the initiation codon and the

by EDTA, had a pH optimum between 6.4 and 7.2, and a temperature optimum of 58°C. The protease activity was strongly dependent on the presence of calcium ions (Fig. 2). Control experiments using extracts of *E. coli/pBR322* showed negligible protease activity with azocasein as substrate. This background activity was significantly below the basic activity of *E. coli/pSK44* without activation by calcium addition. This basal level activation in extracts of *E. coli/pSK44* may be caused by traces of calcium always present in the culture.

These results indicated that the main protease activity in the extracts of *E. coli/pSK44* was a neutral (metallo-) protease.

Growth of *E. coli/pSK44* without addition of extra calcium to the medium appeared normal. No protease was detectable in the culture fluid after removal of the bacterial cells and addition of 2 mM calcium. However, if calcium was added to a growing culture in concentrations higher than 0.5 mM, lysis of the bacterial cells occurred (Fig. 3). Microscopic inspection of the cells at this stage revealed spheroplast-like structures.

In order to confirm that the 3.7 kb insert of *pSK44* came from *B. megaterium*, Southern hybridization analysis of chromosomal DNA was performed using the dATP-7-biotin-labelled 0.47 kb AccI fragment of pSK44 as a probe. The labelled probe hybridized with a single 4 kb fragment in EcoRI digests (data not shown).

**Nucleotide sequence of nprM**

The nucleotide sequence of the *nprM* gene and its flanking regions was determined. The sequencing strategy is depicted in Fig. 4. The sequence revealed a single large ORF consisting of 1686 nucleotides, with a coding capacity for 562 amino acid residues (Fig. 5). A putative Shine–Dalgarno (SD) sequence (AGGGATAGGA) was found 9 bp upstream from a likely translation start site (ATG) (Moran et al., 1982). The calculated free energy of binding (ΔG) of this ribosome-binding site (Tinoco et al., 1973) with the 3′ end of the 16S rRNA from *B. subtilis* is −74.4 kJ mol⁻¹. This ΔG value as well as the distance between the initiation codon and the
Fig. 5. Nucleotide sequence and derived amino acid sequence of the neutral protease gene \textit{nprM} of \textit{Bacillus megaterium}. The nucleotide sequence is numbered from the first base of the translation start site. The predicted amino acid sequence is shown beneath the nucleotide sequence. Putative Shine-Dalgarno (SD) and promoter (-35 and -10) sequences are shown by solid lines below the nucleotide sequence. Horizontal arrows indicate terminator-like inverted-repeat sequences. The putative signal peptide cleavage site (open arrow) and the maturation site (filled arrow) of the predicted pre-pro-enzyme are indicated.

Putative ribosomal binding site are in good agreement with other reported ribosome-binding sites from several genes of different \textit{Bacillus} strains (Hager & Rabinowitz, 1985). A presumptive promoter sequence (−35 region: TAGACG; −10 region: TATAAT), was observed 69 bp upstream of the ATG start codon. The promoter shows good agreement with the consensus sequence for $\sigma^{70}$-promoters of \textit{E. coli} (Hawley & McClure, 1983) and the $\sigma^{54}$-promoters of \textit{B. subtilis} (Moran et al., 1982).

Two inverted-repeat sequences were found downstream of the termination codon (TAA) of the ORF (from nucleotides 1700–1730 and from nucleotides 1752–1795). The $\Delta G$ values of these stem-loop structures were calculated to be $-54.4$ kJ mol$^{-1}$ and $-80.1$ kJ mol$^{-1}$, respectively. These sequences might function as...
protease stearothermophilus protease thermolysin of centre of thermolysin are, however, fully conserved.

**Amino acid sequence of the extracellular neutral protease**

The ORF in the nucleotide sequence codes for a protein of 562 amino acid residues (Fig. 5). This predicted gene product (NprM) exhibits a typical ‘pre-pro’ structure common to many known extracellular Bacillus proteases. The putative ‘pre’ structure shows characteristics of a signal peptide (Perlman & Halverson, 1983; Kreil, 1981), with a positively charged NH₂-terminus. Amino acids 1–5 contain 4 positively charged amino acids (Lys) (n region) and are followed by an h region, containing a stretch of hydrophobic amino acids (amino acids 6–18) and a potential processing site consisting of 6 amino acids with the Ala-X-Ala recognition cleavage site for a signal peptidase.

The signal peptide (amino acids 1–24) is followed by a putative pro-sequence (amino acids 25–245).

The amino acid sequence of the predicted mature protein was compared with the published sequences of other neutral proteases of bacilli (Fig. 6). By homology comparisons we postulated the mature protein from amino acids 246–562.

**Comparison of the amino acid sequence of NprM with other neutral proteases**

The sequence of the predicted extracellular form of this protease shows the highest homology to the neutral protease of B. cereus (Sidler et al., 1986) with 84.5% homology. To the well analysed thermostable neutral protease thermolysin of B. thermoproteolyticus (Colman et al., 1972), NprM shows 72.9% homology; the amino acids with substrate-binding functions in the catalytic centre of thermolysin are, however, fully conserved. Furthermore, there are homologies of over 72% to the thermostable neutral proteases NPRM (Kubo & Imanaka, 1988) and NPRT (Takagi et al., 1985) from B. stearothermophilus and B. caldolyticus (van den Burg et al., 1991). Only limited homology (less than 50%) exists between NprM from B. megaterium and the neutral proteases from B. subtilis var. amylosacchariticus (Yoshimoto et al., 1990; Yang, 1984) and B. amylo liquefaciens (Levy et al., 1975; Vasantha et al., 1984).

**Discussion**

During the characterization of NprM from B. megaterium a strong dependence of the protease activity on calcium ions was found. If the calcium concentration in the culture was kept below 0.5 mM the bacteria grew normally. At higher concentrations, lysis of the bacterial cells occurred. The protease activity identified around colonies after growth on skim-milk nutrient agar was presumably the result of the release of significant amounts of protease from lysed E. coli cells in a high-calcium medium.

The enzyme activity in E. coli/pSK44 cell extracts without addition of calcium was marginal. It apparently arose from traces of calcium present in the culture media which were essential for E. coli to grow. The activity increased with elevated concentrations of calcium ions, up to a point of saturation. This activation followed first order kinetics, implying that there are distinctive interactions of the enzyme molecule with calcium ions. This may be due to either an essential requirement for calcium as cofactor for the protease NprM, or a requirement for calcium to facilitate the proper folding to form the active state of the enzyme. Although there are high sequence homologies to thermolysin, this direct calcium dependence of NprM was not described for thermolysin, where calcium has no effect on the activity but rather causes a reduction of its heat resistance (Feder et al., 1971).

Experiments with the NprM-expressing E. coli strain demonstrated the the neutral (metallo-) protease can neither be actively excreted nor passively pass through the membranes of intact cells. In Gram-negative and Gram-positive bacteria, different mechanisms exist for the release of exoenzymes into the periplasmic space or the culture fluid. In addition to the (pre-)pro-structure at the amino terminus in Gram-negative bacteria, many secreted proteins have a special secretion signal at the carboxy-terminal end (Nicaud et al., 1986; Pugsley, 1988), essential for translocating exoenzymes through the outer membrane. The protease of B. megaterium might be translocated through the inner membrane of E. coli if the Gram-positive pre-pro-structure is recognized and acts as a translocation signal. If so, it might not
penetrate the outer membrane, as it has no C-terminal secretion signal typical of Gram-negatives. At present it is not clear if the Bacillus protease expressed in E. coli remains periplasmic or cytoplasmic. One piece of evidence for the periplasmic location is that the calcium-activated enzyme disrupts the cell wall, causing formation of spheroplast-like structures. This could occur either by a direct interaction of the protease with the peptidoglycan or indirectly by inactivation of enzymes essential for cell wall synthesis.

Amino acid residues necessary for catalytic activity appear to be highly conserved in this class of Bacillus proteases. Zinc plays an essential role for the proteolytic activity of many neutral (metallo-) proteases and functions as the central atom of the neutral (metallo-) proteinase of B. megaterium (Keay et al., 1971) and of thermolysin of B. thermoproteolyticus. The zinc-binding amino acid residues of thermolysin are His-142, His-146 and Glu-164 (Colman et al., 1972). The same amino acids are found in identical positions in the amino acid sequence of NprM, as is true for the amino acid residues necessary for the formation of hydrogen bonds between the enzyme and the peptide substrate (Try-115, Ala-113, Asn-112, Arg-203) and the amino acid residues which are directly involved in catalysis (His-231, Glu-143) (Kester & Matthews, 1977 a,b).

References


with the homologous but more thermostable enzyme thermolysin. Journal of Molecular Biology 199, 525–537.