Cloning and Nucleotide Sequence of the Highly Thermostable Neutral Protease Gene from *Bacillus stearothermophilus*

By MOTOKI KUBO AND TADAYUKI IMANAKA*

Department of Fermentation Technology, Faculty of Engineering, Osaka University, Yamadaoka, Suita-shi, Osaka 565, Japan

(Received 4 November 1987; revised 23 February 1988)

The gene (nprM) for the highly thermostable neutral protease of *Bacillus stearothermophilus* MK232 was cloned in *Bacillus subtilis* using pTB53 as a vector. The nucleotide sequence of nprM and its flanking regions was determined. The DNA sequence revealed only one large open reading frame, composed of 1656 base pairs and 552 amino acid residues. A Shine–Dalgarno (SD) sequence was found 12 bases upstream from the translation start site (ATG). A possible promoter sequence (TTTTCC for the −35 region and TATTGT for the −10 region), which was nearly identical to the promoter for another thermostable neutral protease gene, nprT, was also found about 40 bases upstream of the SD sequence. The deduced amino acid sequence contained a signal sequence in its amino-terminal region. The sequence of the first five amino acids of purified extracellular protease completely matched residues 237–241 of the open reading frame. This suggests that the enzyme is translated as a large polypeptide containing a pre–pro structure as is known for other neutral proteases. The amino acid sequence of the extracellular form of this protease (316 amino acids, molecular mass 34266 Da) was identical to that of the thermostable neutral protease (thermolysin) from *Bacillus thermoproteolyticus* except for two amino acid substitutions (Asp37 to Asn37 and Gln119 to Gln119). The G + C content of the coding region of nprM was 42 mol%, while that of the third letter of the codons was lower (36 mol%). This extremely low content is an exceptional case for genes from thermophiles. When the protease genes, nprM and nprT, were cloned on pTB53 in *B. subtilis*, the expression of nprM was about 20 times higher than that of nprT. The reason for the difference between the two systems is discussed.

**INTRODUCTION**

Microbial metallo-(neutral)proteases (EC 3.4.24.4) from the genus *Bacillus* are widely distributed secretory enzymes, and the structural genes of several have been cloned and sequenced (Fujii et al., 1983; Takagi et al., 1985; Vasantha et al., 1984; Yang et al., 1984). These studies revealed the existence of a pre–pro structure for enzyme secretion. However, the proteases have not been well characterized enzymically. In contrast, the thermostable neutral protease from *Bacillus thermoproteolyticus* (thermolysin) has been intensively characterized (Endo, 1962; Kester & Matthews, 1977; Latt et al., 1969; Morihara & Tsuzuki, 1970; Stauffer, 1971). The primary and tertiary structures, the active site and substrate-binding site have been determined (Bigbee & Dahlquist, 1974; Matthews et al., 1972a, b; Titani et al., 1972). Until recently, thermolysin was the most thermostable protease known.

We have already cloned and sequenced the gene (nprT) for a thermostable neutral protease, NprT, from *Bacillus stearothermophilus* CU21 (Fujii et al., 1983; Takagi et al., 1985). The amino acid sequence is highly homologous (85%) with that of thermolysin. However, thermolysin is more thermostable than NprT. By site-directed mutagenesis, we succeeded in enhancing the thermostability of NprT (Imanaka et al., 1986). However, thermolysin was still more thermostable than the mutant form of NprT. We have independently isolated an industrially
useful strain, *B. stearothermophilus* MK232 (Kubo *et al.*, 1988), which produces a more thermostable neutral protease (NprM). When the enzymes are heated at 90 °C for 30 min, the remaining activity of NprM is about 45%, whereas that of thermolysin is about 30%. Moreover, the specific activity of NprM is 27 700 units (mg protein)^{-1}, which is higher than that [19 200 units (mg protein)^{-1}] of thermolysin by 40%.

This paper describes the cloning and expression in *Bacillus subtilis* of nprM, the gene for the highly thermostable neutral protease from *B. stearothermophilus* MK232, and compares the amino acid sequence deduced from the nucleotide sequence with that of thermolysin from the viewpoint of enzyme thermostability. The reason for the efficient expression of nprM in comparison with that of nprT in *B. subtilis* is also discussed.

**METHODS**

*Bacterial strains, plasmids, phages and culture conditions*. The *Bacillus* strains used were *B. steaothermophilus* MK232 [Cm^r nprM^+ (structural gene of neutral protease)] (Kubo *et al.*, 1988) and *B. subtilis* MT-2 (trpC2 leuC7 hsdR hsdM Npr^+*) (Fujii *et al.*, 1985). *Escherichia coli* K12 JM103 [Δ(lac-pro) thi strA endA sbcB hsdR] was the host strain for phages M13mp10 and M13mp11. Plasmids used were pTB53 (Km^Tc^) (Imanaka *et al.*, 1981a; Imanaka *et al.*, 1984) and pNP22 [Km^Tc^ nprT^+* (structural gene of neutral protease from *B. stearothermophilus* CU21)] (Takagi *et al.*, 1985). Bacteria were grown in L broth, on L agar or LC agar (L agar plus 1% casein) (Fujii *et al.*, 1983). The antibiotics used were kanamycin and tetracycline (each 5 mg ml^{-1}) for both *B. steaothermophilus* and *B. subtilis*.

*Isolation of plasmid and chromosomal DNA, restriction enzyme treatment, ligation of DNA and transformation*. Plasmid DNA was prepared by either the rapid alkaline extraction method or CsCl/ethidium bromide equilibrium density gradient centrifugation as described by Imanaka *et al.* (1982). Chromosomal DNA was prepared as described by Imanaka *et al.* (1981b). Treatment of DNA with restriction enzymes and ligation of DNA with T4 DNA ligase were done as recommended by the manufacturer. For transformation of *B. subtilis* with plasmid DNA, competent cells were prepared as described by Imanaka *et al.* (1981a).

*Gel electrophoresis for DNA analysis and isolation*. For the analysis of DNA, agarose or polyacrylamide gel electrophoresis was done under standard conditions (Maniatis *et al.*, 1982). Recovery of DNA from either low-melting point agarose or polyacrylamide gels was done by the standard method (Maniatis *et al.*, 1982).

*DNA sequencing analysis*. DNA sequencing was done by the dideoxy method with *E. coli* JM103 and M13 phages (Messing, 1983). Specific fragments were subcloned and a set of exonucleaseIII–mung bean nuclease–Klenow fragment generated deletions starting from HpaI and PvuII restriction sites were constructed using a deletion kit from Takara Shuzo. Both strands were sequenced and the sequence was generated across the restriction sites defining the ends of subcloned fragments.

*Detection of protease-producing colonies and assay of protease activity*. Protease-producing colonies on plates were detected, and protease activity was assayed for casein hydrolytic activity as described by Fujii *et al.* (1983). We defined 1 unit of protease as the quantity required to increase the absorbance at 275 nm by the equivalent of 1 μg tyrosine min^{-1} at 37 °C.

*Purification of extracellular protease*. The culture broth supernatant was treated by ultrafiltration [through UF-10ps and UF-100ps, filters (Tohos, Japan)] to remove smaller (molecular mass < 10 kDa) and larger (> 100 kDa) proteins, respectively; the resulting protein was precipitated by adding ethanol to a final concentration of 60% (v/v). After freeze-drying, the precipitate was dissolved in about 10 ml buffer solution (50 mM-Tris/HCl, pH 7.5, 5 mM-CaCl_2). The solution was subjected sequentially to Toyopearl HW-50 and TSK gel G-2000 SW HPLC column chromatography. The enzyme fractions for neutral protease were collected and immediately inactivated to minimize self-digestion by adding trichloroacetic acid to a final concentration of 2% (w/v). The protein was precipitated, isolated by centrifugation (15 000 g, 30 min) and dissolved in 0.8 ml 98% (v/v) formic acid. After desalting, the sample was subjected to TSK gel phenyl 5 PW RP HPLC column chromatography (Tohos); the peak fractions were collected and then freeze-dried. This sample was used for the determination of the amino acid sequence.

*Protein analysis by SDS-PAGE*. This was done as described by Matsumura *et al.* (1984).

*Amino-terminal amino acid sequence*. This was determined by Edman degradation as described by Takagi *et al.* (1985).

*Enzymes and chemicals*. Restriction endonucleases and M13 sequencing kit were purchased from Toyobo. T4 ligase, bacterial phosphatase and low-melting-point agarose were from Takara. Kanamycin and tetracycline were from Sigma. [α-32P]dCTP was from Amersham. All other chemicals were from Wako.

*Analysis of amino acid sequence homology*. Amino acid sequence homology was analysed with an NEC PC-9801 computer (Nippon Electric) and 'GENIAS' system (Mitsui).

---

*p*
Thermostable neutral protease gene

Results

Cloning of nprM in B. subtilis

Chromosomal DNA (about 5 μg) from B. stearothermophilus MK232 was digested with PstI and ligated with a PstI digest (about 2 μg) of pTB53 in a total volume of 50 μl. The ligation mixture was used to transform B. subtilis MT-2, and about 10⁵ transformants were obtained on LC agar containing 5 μg kanamycin ml⁻¹. Of these, eight colonies could form haloes. All were also resistant to tetracycline. Plasmid DNA prepared from each isolate (Km'Tc'Npr⁺) was used to transform strain MT-2. All the transformants exhibited the phenotype Km'Tc'Npr⁺.

To confirm the cloning of the thermostable neutral protease gene, extracellular protease was assayed in these transformants. The protease was inhibited by EDTA but not by phenylmethylsulphonyl fluoride (data not shown), indicating that the enzyme is a neutral-(metallo)protease. The pH optimum of the protease was between 6·5 and 7·5 (data not shown). Furthermore, protease activity was retained after heat treatment at 65 °C for 15 min, indicating that the protease was thermostable. These results imply that the thermostable neutral protease gene (nprM) from B. stearothermophilus MK232 was cloned in B. subtilis.

One of the recombinant plasmids carrying nprM was designated pTZ232. pTZ232 contained a 5·5 MDa insert of B. stearothermophilus MK232 DNA. The restriction map of pTZ232 is shown in Fig. 1. When the 1·2 MDa SalI–PstI fragment was deleted, protease activity was not lost. However, deletion of either the 2·0 MDa BamHI–PstI fragment or the 1·5 MDa HpaI–PstI fragment from pTZ232 resulted in loss of protease activity. It is concluded therefore that the nprM gene is located around the 2·0 MDa HindIII–PvuII fragment of pTZ232.

Nucleotide sequence of nprM

The nucleotide sequence of the nprM gene and its flanking regions was determined (Fig. 2). The sequence revealed only one large open reading frame, composed of 1656 nucleotides (552 amino acid residues). A Shine–Dalgarno (SD) sequence (GAAAAGG) was found 12 bases upstream from a likely translation start site (ATG) (Moran et al., 1982). A putative promoter
Thermostable neutral protease gene

Fig. 2. Nucleotide sequence and amino acid sequence of the neutral protease gene nprM. The nucleotide sequence is counted from the first base of the translation start site as +1. The amino acid sequence is shown beneath the nucleotide sequence. The amino-terminal amino acid sequence of extracellular protease, determined by Edman degradation, is specified below the amino acid sequence by arrows. The first amino acid of translation (Met) is counted as 1. The putative Shine-Dalgarno (SD) sequence (nucleotide -18 to -12) and promoter region (-35 and -10 regions) are shown by solid lines below the nucleotide sequence. Asterisks (***)) indicate a stop codon. A putative transcriptional terminator containing an inverted repeat is shown by arrows (++) below the nucleotide sequence. Amino acid substitutions in comparison with thermolysin (Titani et al., 1972) are indicated by broken lines.

sequence (−35 region: TTTTCC; −10 region: TATTGT), which was nearly identical to the known promoter (−35 region: TTTTCC; −10 region: TATTTT) for another thermostable neutral protease gene, nprT (Takagi et al., 1985), was also found about 40 bases upstream of the SD sequence. Amino-terminal amino acids deduced from the nucleotide sequence exhibited a typical signal sequence with basic amino acids (Lys, Arg), followed by a hydrophobic core (Holland et al., 1986; Perlman & Halvorson, 1983). A typical terminator sequence was also found immediately downstream from the open reading frame.
Amino acid sequence of the extracellular neutral protease

The extracellular form of the protease was purified from the culture broth of *B. stearothermophilus* MK232, and the amino-terminal amino acid sequence was determined through five cycles of Edman degradation using an automatic sequencer. The first five amino acids were Ile-Thr-Gly-Thr-Ser. The sequence completely matched that deduced from the nucleotide sequence starting at ATA (+709 to +711, Ile237 in Fig. 2, at position 1 as extracellular form). The corresponding amino acids are indicated with arrows in Fig. 2. The molecular mass of the extracellular protease calculated from the amino acid sequence (316 amino acids, mol. mass 34266 Da) was nearly the same as the known value for the enzyme (34 kDa) (Kubo et al., 1988). These results indicate that the open reading frame is translated as a large polypeptide containing a pre–pro structure, as is the case for other neutral proteases (Takagi et al., 1985; Vasantha et al., 1984).

The amino acid sequence of the polypeptide was compared with the published sequences of other thermostable neutral proteases (Takagi et al., 1985; Titani et al., 1972). The sequence of the extracellular form of this protease was identical to that of the thermostable neutral protease (thermolysin) from *B. thermoproteolyticus* except for the two amino acid substitutions [Asp37 to Asn37 (corresponding to Asn273 in Fig. 2), Glu119 to Gln119 (corresponding to Gln355 in Fig. 2)]. The extracellular proteases from *B. stearothermophilus* strains MK232 and CU21 show a very high degree of homology (85% of the amino acids are identical) whereas the pre–pro regions are less closely related (40% amino acid identity) (Fig. 3).

Expression of nprM and nprT in *B. subtilis*

The promoter sequences of nprT and nprM may be similar as mentioned above. Nevertheless, strain MK232 containing nprM produced much more protease than did strain CU21 containing nprT. To investigate the difference in enzyme productivity on an equal basis, the two protease genes were subcloned in *B. subtilis*, by using the low-copy-number vector plasmid pTB53 as shown schematically in Fig. 4. The recombinant plasmids containing nprM and nprT were
designated pTNM53 and pTNT53, respectively. *B. subtilis*(pTNM53) produced about 20 times more protease in 24 h than did the strain carrying pTNT53, although both strains exhibited the same growth pattern (Fig. 5).

**DISCUSSION**

The thermostability of protease NprM is higher than that of thermolysin by about 10%, and the specific activity of NprM is about 40% higher than that of thermolysin. To compare the amino acid sequences, we cloned the *nprM* gene in *B. subtilis* and determined the complete nucleotide sequence. The amino acid sequence of NprM deduced from the nucleotide sequence
Fig. 6. Three-dimensional structure of thermolysin (Matthews et al., 1972a). Open circles, α-carbon positions. The zinc atom is drawn stippled with its three protein ligands shown diagrammatically as broken lines. Four calcium atoms are shown as solid circles. Amino acid substitutions are shown by arrows.

was identical to that of thermolysin except for two amino acid substitutions (Asp37 to Asn37 and Glu119 to Gln119, for the extracellular form). Both substitutions are the addition of an amino group to an acidic amino acid, resulting in an uncharged polar amino acid. The two substitutions are located in Domain A (Ile1–Gly136) and not in Domain B (Ile156–Lys316) of the enzyme (Fig. 6), and might contribute to the effective electrostatic balance in Domain A without any significant change of three-dimensional protein structure. In other words, Asn37 and Gln119 might reduce the repulsion with negatively charged amino acids in the three-dimensional vicinity, Asp32 and Asp72, respectively (Fig. 6). Moreover, the substituted amino acids, Asn37 and Gln119, might also increase the hydrogen bonding with neighbouring amino acids. Thus, Asn37 is surrounded by Asn19, Asn21, Asn33, Thr34, Arg35, Asn96 and Asn97 (Fig. 6), all of which are potentially able to form hydrogen bonds. Likewise, Gln119 is surrounded by Arg101, Ser102, Ser103, Asn116, Ser118 and Met120 (Fig. 6). According to the statistical data on enzyme
thermostability mentioned by Argos et al. (1979), the amino acid substitution Asp to Asn could enhance thermostability. This is one example where amino acid substitutions which might increase hydrogen bonding or give a good electrostatic balance were generated in nature and result in enhanced thermostability and specific activity of the enzyme.

We have also succeeded in enhancing the thermostability of the neutral protease NprT by replacing Gly144 with Ala144 (Imanaka et al., 1986). The amino acid substitution was made to increase the internal hydrophobicity and stabilize an α-helix for strong packing. Therefore, many factors such as the hydrophobicity and charge of amino acids and their hydrogen bonding, in addition to changes in secondary and tertiary structures of an enzyme, should be considered in the enhancement of thermostability.

The G + C content of therophilic bacterial genes is in general higher than that of mesophilic genes (Kagawa et al., 1984; Nakajima et al., 1985; Takagi et al., 1985; Winter et al., 1983). In our previous work, the G + C content of the coding region for the nprT gene was 58 mol%, while that of the third letter of the codons was higher (72 mol%) (Takagi et al., 1985). Surprisingly, however, the G + C content of the coding region for the nprM gene was 42 mol%, while that of the third letter of the codon was lower (36 mol%). These values are very similar to or a little lower than those for the neutral proteases from B. amyloliquefaciens (46 and 49 mol%, respectively: Vasantha et al., 1984) and B. subtilis (44 and 42 mol%, respectively: Yang et al., 1984). Although the amino acid sequence homology between the two extracellular proteases from B. stearothermophilus is very high (Fig. 3), the G + C contents of nprM and nprT are quite different from each other. The difference between the two genes suggests that they may possibly have different evolutionary origins.

There exists a direct perfect repeat sequence (at positions −23 to −7, and −3 to 14) in the translation start site region. Therefore, the direct repeat might play an important role in the regulation of gene expression. Since a typical SD sequence is located at positions −18 to −12, the same sequence (at positions 3 to 9) might be used as another SD sequence. In this case, it is most likely that Met7 in Fig. 2 is used as a translation start codon. Even if the two sets of translation start sites functioned in reality, no difference would be found in the extracellular protease.

We compared the expression of the nprM gene with that of the nprT gene in B. subtilis. In spite of being cloned in the same low-copy-number vector plasmid (pTB53), the productivity of NprM was about 20 times higher than that of NprT. The difference in the enzyme productivity might be explained by the following arguments. (1) Since the signal sequence and the following sequences and the corresponding initiation codons are found, the translation start sites of NprT as mentioned above (see Fig. 3), the secretion efficiencies of the two enzymes might be different. (2) Since the G + C contents of both genes are quite different, the codon usage might affect the gene expression in B. subtilis. (3) Since the two tandem SD sequences and the corresponding initiation codons are found, translation efficiency might be increased. Mechanisms for the high productivity of NprM are now under investigation.

We thank K. Seto, K. Murayama, Tosoh Corporation and M. Takagi for helpful discussion. We are grateful to F. Sakiyama, Institute for Protein Research, Osaka University, for the determination of amino-terminal amino acid sequence of neutral protease. We also thank H. Shima for technical assistance.

REFERENCES


and characterization of antibiotic resistance plasmids from thermophilic bacilli and construction of deletion plasmids. *Journal of Bacteriology* 146, 1091–1097.


