Note

Nucleotide Sequence of the Subtilisin NAT Gene, aprN, of Bacillus subtilis (natto)

Takashi Nakamura, Youhei Yamagata, and Eiji Ichishima*

Laboratory of Molecular Enzymology, Department of Applied Biological Chemistry, Faculty of Agriculture, Tohoku University, 1-1, Tsutsumidori-Amamiyamachi, Aobaku, Sendai 981, Japan

Received March 30, 1992

A serine proteinase from Bacillus subtilis (natto) (formerly designated Bacillus natto), subtilisin NAT (formerly designated BSF), is probably the most important enzyme for the characteristic taste and flavor of natto. It hydrolyzes raw materials during natto fermentation. We have purified extracellular subtilisin NAT from B. subtilis (natto) and characterized the mode of action of the enzyme. The molecular weight and isoelectric point of the enzyme are 30,000 and 8.7, respectively. Ouchterlony double immunodiffusion showed no cross-reaction between subtilisin NAT and the subtilisins BPN' and Carlsberg.

This nucleotide sequence of the subtilisin NAT gene, aprN, of B. subtilis (natto) NC 2—1 ts 25-23 was isolated from a commercial starter from Naruse Fermentation Chemical Institute, Nerima. The amino acid sequence of the enzyme was compared to the sequences of other subtilisins.

Shotgun cloning experiments to obtain the aprN were first done with an Escherichia coli host vector system. Plasmid DNA was isolated by alkaline extraction. Chromosomal DNA of B. subtilis (natto) NC 2—1 ts 25-23 prepared by the method of Saito and Miura was partially digested with Sau3AI and ligated to BamHI-digested pUC19. E. coli HB101 was transformed with the ligated mixture and plated on agar containing 1% skim milk. After 3 to 4 days of cultivation in Luria-Bertani medium at 37°C, three colonies of 30,000 ampicillin-resistant transformants (A') formed halos. Recombinant plasmids containing 9.7, 5.2, or 4.1 kb of inserted DNAs were recovered, purified from the halo-forming transformants, and designated pTNT70, pTNT80, and pTNT83, respectively.

To get the aprN gene from B. subtilis (natto), we used a synthetic nucleotide oligomer, 5'-ATGGAATTCGT/CAGATATC/ wATCTAATTCCATGGATCG-3', as a probe. The octadecamer was synthesized on the basis of the amino acid sequence of the subtilisin NAT that was in the reverse order of neighboring Met119-Met134 and that was common to four subtilisins: subtilisin E,4,5 Amylosaccharitic,46 BPN',7,8 and Carlsberg.9

Southern blotting with DNA–DNA hybridization of HindIII-digested chromosomal DNA of B. subtilis (natto) with the 32P-labeled probe gave a positive band at 1.2 kb, as expected from the restriction pattern of pTNT70, pTNT80, and pTNT83. E. coli cells harboring the cloned HindIII fragment in the pUC19 vector had protease activity in the extracellular fraction but did not induce isoazophyto-β-galactoside.

The proteolytic activities found in the 24-h culture of Luria-Bertani broth were analyzed with a fluorogenic substrate of subtilisin NAT, succinyl-l-leucyl-l-leucyl-l-valyl-l-tyrosyl-4-methylcoumarin-7-amide as reported previously. All culture broths of transformed strains of E. coli HB101 carrying pTNT plasmids gave stained zones on SDS-PAGE at the position of the molecular weight of 30,000 by Western blotting. The washed-cell fraction had no proteolytic activity and no positive bands.

The complete nucleotide sequence of the aprN gene is shown in Fig. 1. As reported for subtilisin genes E4,5 and Amylosaccharitic,6 the aprN gene started with GTG as its initiation codon, followed by an open reading frame of 114 nucleotides, and terminating at three stop codons, TAA TAG TAA. The putative transcription terminator sequence TAAAAAAAGGAGGT TCTTCCATACCTGGCTTCTTTTTA was 7 bp downstream from the C-terminus of the mature protein region (the arrows indicate the terminators).

The amino acid sequence deduced from the DNA sequence is shown in Figs. 1 and 2. By comparison with other subtilisin genes,4–6 the sequence of the protein-coding region of subtilisin NAT was found to code for 29 residues of a signal peptide and 77 residues of a propeptide that precedes the 275 residues with the molecular weight of 27,700 of mature subtilisin NAT.

The nucleotide sequence of subtilisin NAT gene was homologous to that of subtilisin E4,5 in B. subtilis 1168, with discrepancies at 13 nucleotides of the 1473 nucleotides we sequenced. Twelve of these nucleotides were found in the mature subtilisin coding sequence, so there were changes in two amino acids between the NAT and E genes. The nucleotide sequence was also homologous to that of subtilisin Amylosaccharitic46 in B. subtilis var. amylosaccharitic, with discrepancies at 27 nucleotides of the 1473 sequenced. Twenty of these were found in the mature subtilisin coding sequence, so there were changes in four amino acids between the NAT and Amylosaccharitic genes. The primary structure of the mature region of the subtilisin NAT had 99.5 and 99.3% homology with the primary structure of mature subtilisins E4,5 and Amylosaccharitic.6 The homology of total amino acid residues between subtilisin NAT and subtilisin BPN'7,8 from Bacillus amyloliquefaciens or Carlsberg9 from Bacillus licheniformis was 86% or 72%, and sequences were conserved around the three essential amino acids (serine-221, histidine-64, and aspartic acid-32) in the catalytic center.

References


* To whom correspondence should be addressed.
Fig. 1. Nucleotide Sequence of Subtilisin NAT Gene and Deduced Amino Acid Sequence of the Enzyme.

The putative SD sequence is overlined and the transcription terminator is indicated by arrows. Discrepancies in the nucleotide sequence between subtilisins NAT and E are indicated by asterisks. The deduced amino acid sequences of the N terminus and three peptides are underlined.
Subtilisin NAT  AQSVPYGJISIKAPALHSGQYTGNSNVKAV
Subtilisin E   AQSVPYGJISIKAPALHSGQYTGNSNVKAV
Subtilisin Amylosacchariticus AQSVPYGJISIKAPALHSGQYTGNSNVKAV
Subtilisin BPN' AQSVPYGVJISIKAPALHSGQYTGNSNVKAV
Subtilisin Carlsberg AQTVPYGJILKADKVAGQFNGHNSNVKAV

* 40 50 60 * 70 80 90
IDSIGIDSPDLNVGGASFVPSZTNPQDGGSHGTVAGTIAALHSQGVGSLPSAL
IDSIGIDSPDLNVGGASFVPSZTNPQDGGSHGTVAGTIAALHSQGVGSLPSAL
IDSIGIDSPDLNVGGASFVPSZTNPQDGGSHGTVAGTIAALHSQGVGSLPSAL
IDSIGIDSPDLNVGGASFVPSZTNPQDGGSHGTVAGTIAALHSQGVGSLPSAL
LDTGQAEBSPDLNVGGASFVAGFYA-N-TDNGHSGTVAGTVAALHSQGVGSLPSAL

100 110 120 130 140 150
YAVKVLDSTSGQYTVQSWIINGIEWAINNNMDVINLNGSATGSLT TKCTVVDKAVSGIVV
YAVKVLDSTSGQYTVQSWIINGIEWAINNNMDVINLNGSATGSLT TKCTVVDKAVSGIVV
YAVKVLDSTSGQYTVQSWIINGIEWAINNNMDVINLNGSATGSLT TKCTVVDKAVSGIVV
YAVKVLDSTSGQYTVQSWIINGIEWAINNNMDVINLNGSATGSLT TKCTVVDKAVSGIVV
YAVKVLDSTSGQYTVQSWIINGIEWAINNNMDVINLNGSATGSLT TKCTVVDKAVSGIVV

160 170 180 190 200 210
AAAGNEGSGGSGSTSTGVYPAKYPSTIAVGAANNSSQRAFSVSAGSELDMAPGVSQSTLP
AAAGNEGSGGSGSTSTGVYPAKYPSTIAVGAANNSSQRAFSVSAGSELDMAPGVSQSTLP
AAAGNEGSGGSGSTSTGVYPAKYPSTIAVGAANNSSQRAFSVSAGSELDMAPGVSQSTLP
AAAGNEGSGGSGSTSTGVYPAKYPSTIAVGAANNSSQRAFSVSAGSELDMAPGVSQSTLP
AAAGNEGSGGSGSTSTGVYPAKYPSTIAVGAANNSSQRAFSVSAGSELDMAPGVSQSTLP

220* 230 240 250 260 270
GTYGAYNGTSNATPHVAGAAALILSHKPTWTNAQVRDRLESTATLGLNSFYGYKGLINV
GTYGAYNGTSNATPHVAGAAALILSHKPTWTNAQVRDRLESTATLGLNSFYGYKGLINV
GTYGAYNGTSNATPHVAGAAALILSHKPTWTNAQVRDRLESTATLGLNSFYGYKGLINV
GTYGAYNGTSNATPHVAGAAALILSHKPTWTNAQVRDRLESTATLGLNSFYGYKGLINV
GTYGAYNGTSNATPHVAGAAALILSHKPTWTNAQVRDRLESTATLGLNSFYGYKGLINV
GTYGAYNGTSNATPHVAGAAALILSHKPTWTNAQVRDRLESTATLGLNSFYGYKGLINV

Fig. 2. Amino Acid Sequences of Subtilisin NAT and Related Bacterial Serine Proteases.
Asterisks identify the aspartic acid, histidine, and serine residues corresponding to active sites in the subtilisins E, Amylosacchariticus, BPN', and Carlsberg.