Extremophiles (2002) 6:233–243
Digital Object Identifier (DOI) 10.1007/s00792-001-0248-1

ORIGINAL PAPER

Hyeung Jin Jang · Byoung Chan Kim · Yu Ryang Pyun
Yu Sam Kim

A novel subtilisin-like serine protease from *Thermoanaerobacter yonseiensis* KB-1: its cloning, expression, and biochemical properties

Received: March 12, 2001 / Accepted: September 4, 2001 / Published online: February 20, 2002

Abstract A gene, *tayI*, encoding a novel subtilisin-like protease, designated thermicin, from the extremely thermophilic bacterium *Thermoanaerobacter yonseiensis* KB-1 (DSM 13777) was cloned by using a sequence tag containing the consensus sequence of proteases. The gene consisted of 1,239 nucleotides, and the deduced amino acid sequence indicated that it is a preproenzyme with a 311-residue mature protein composed of canonical catalytic residues (Asp29, His64, and Ser252). Thermicin was overproduced in *E. coli* as a fusion protein with a histidine tag and purified by nickel nitrilotriacetic acid affinity chromatography. Thermicin from *E. coli* showed maximum proteolytic activity at 92.5°C and pH 9.0, and its half-life was 30 h at 80°C. In order to determine cleavage specificity, thermicin was incubated with insulin β chain, and the resulting peptides were analyzed by matrix assisted laser desorption/ionization-time of flight mass spectrometry. The carboxyl group side of the Val12, Leu15,17, Gly23, and Pro28 residues was cleaved. Thermicin is well known to hydrolyze Gly- and Pro-rich collagens. The K_m and k_{cat}/K_m values of thermicin for the hydrolysis of the synthetic substrate L-Gly-Pro- *p*-nitroaniline were 54.16 μM and 142.05 (10^5 s^{-1} M^{-1}), respectively, at 92.5°C and pH 9.0. Amino acid sequence comparison and phylogenetic analysis indicated that this enzyme belongs to a new subgroup with respect to its molecular evolution, when compared with previously characterized subtilisins. This result indicates that thermicin is a novel enzyme different from other thermostable proteases.

Communicated by G. Antranikian

H.J. Jang · Y.S. Kim (✉)
Program in Biomaterial Science and Engineering, Bioproducts Research Center, Department of Biochemistry, College of Science, Yonsei University, Seoul 120-749, Korea
Tel. +82-2-21233448; Fax +82-2-3923488
e-mail: yskim@yonsei.ac.kr

B.C. Kim · Y.R. Pyun
Program in Biomaterial Science and Engineering, Bioproducts Research Center, Department of Biotechnology, College of Engineering, Yonsei University, Seoul, Korea

Key words Subtilisin-like serine protease · Phylogenetic analysis · *Thermoanaerobacter yonseiensis* KB-1 · thermicin

Introduction

Several extremely thermostable proteases from thermophiles such as *Pyrococcus* sp. (Morikawa et al. 1994), *Desulfurococcus mucosus* (Cowan et al. 1987), *Aquifex pyrophilus* (Choi et al. 1999), *Aeropyrum pernix* K1 (Chavez Croocker et al. 1999), *Pyrococcus furiosus* (Blumentals et al. 1990), *Thermococcus stetteri* (Klingeberg et al. 1995), and *Pyrobaculum aerophilum* (Völkl et al. 1994) have been studied for their biotechnological applications as well as for their metabolic significance. Because of their high activity and stability at elevated temperatures, these enzymes are also good models for studying thermal stability of proteins (Rao et al. 1998). Moreover, the application of thermophilic microorganisms or their products in industrial processes has initiated a new era in biotechnology. Thermophiles are very stable at high temperatures, and this characteristic reduces risk of contamination by other microorganisms. This is one of the main reasons for characterizing and using these enzymes from thermophiles. Additional benefits during industrial processes include improved transfer rates and lower viscosity (Klingeberg et al. 1991). Due to the unusual properties of these enzymes, they are expected to fill the gap between biological and chemical processes (Koch et al. 1990; Ng and Kenealy 1986).

Recently, advances in analytical techniques have demonstrated that proteases, which catalyze the cleavage of peptide bonds, have highly specific and selective actions, such as activation of zymogenic forms of enzymes by limited proteolysis, blood clotting and lysis of fibrin clots, and processing and transport of secretory proteins across the membranes. In spite of the many proteases reported previously, however, there is not much information about enzymes that degrade collagens at elevated temperatures.

At present, more than 200 proteases have been assigned to the superfamily of subtilases (subtilisin-like serine pro-

teases), with representatives both among microorganisms (archaea, bacteria, fungi, and yeast) and among higher eukaryotes. All the enzymes belonging to this superfamily have in common a core structure, the catalytic domain, which is characterized by the presence of structurally conserved regions that correspond to common secondary structure elements. So far, most of the subtilisins that have been characterized are extracellular, and they have been subdivided into six families on the basis of sequence similarity. Thermostable serine proteases, which are produced by the archaea and by thermophilic bacteria, have also been identified; for example, thermitase has been found in some extremophiles and enzymes of the pyrolysin family have been found in thermophiles (Siezen and Leunissen 1997). These proteases in the subtilisin superfamily contain His, Asp, and Ser, the so-called catalytic triad, in their active core structure. Currently, there are six subgroups of the subtilisin superfamily, designated A to F. In this paper, we demonstrate the cloning and expression of a gene encoding a novel protease from the anaerobic bacterium *Thermoanaerobacter yonseiensis*. We show on the basis of a phylogenetic analysis that this thermostable enzyme belongs to a new class of the subtilisin family. Moreover, it has potential use in the treatment of collagen wastes.

Materials and methods

Chemicals

p-Nitroaniline (*p*NA) chromogenic substrates (Sigma, St. Louis, MO, USA) were used for substrate specificity: L-Pro-*p*NA, L-Gly-Pro-*p*NA, L-Phe-*p*NA, *N*-succinyl-Gly-Gly-Gly-*p*NA, *N*-succinyl-Gly-Phe-Gly-*p*NA, *N*-CBZ-Gly-Gly-Leu-*p*NA, PGly-Phe-Leu-*p*NA, *N*-succinyl-Ala-Ala-Pro-Leu-*p*NA, and Mca-Arg-Pro-Lys-Pro-Tyr-Ala-Nva-Trp-Met-Lys(Dnp)-NH_2 (Bachem, Heidelberg, Germany). Phenylglyoxal, phenylmethylsulfonyl fluoride (PMSF), and salicylaldehyde were also purchased from Sigma.

Bacterial strains and plasmids

Thermoanaerobacter yonseiensis KB-1 (Kim et al. 2001) was grown at 75°C in an enrichment culture medium (EM1 medium) containing the following components per liter of distilled water: 2.0 g; NaCl, 0.5 g; $MgSO_4 \cdot 7H_2O$, 0.8 g; $MgCl_2 \cdot 6H_2O$, 0.5 g; $CaCl_2 \cdot 2H_2O$, 0.1 g; $NaHCO_3$, 0.5 g; KH_2PO_4, 0.1 g; K_2HPO_4, 1.0 g; $(NH_4)_2SO_4$, 0.1 g; KCl, 0.1 g; KBr, 0.03 g; H_3BO_3, 0.03 g; $SrCl_2 \cdot 6H_2O$, 0.033 mg; Na_2WO_4, 0.026 mg; Na_2SeO_3, 1.0 g; yeast extract, 1.0 mg; resazurin, 10.0 g; sulfur powder, 10 ml; and trace elemental solution (1% stock solution, DSMZ-141, Deutsche Sammlung von Mikroorganismen und Zellkulturen, Braunschweig, Germany). The medium was reduced with $Na_2S \cdot 9H_2O$ to a final concentration of 0.5 g/l and then adjusted to pH 6.8 at room temperature. Fifty milliliters of the EM1 medium, contained in a serum bottle, was flushed with N_2 gas and sterilized by autoclaving at 105°C for 1 h. D-Xylose and vitamin solution DSMZ-141 were added from filter-sterilized stock solutions prior to inoculation. The final xylose concentration was 5 g/l, and cultures were incubated without shaking at 80°C. *E. coli* strains XL1-blue MRF′ and XLOLR were grown in Luria broth supplemented with tetracycline (20 μg/ml). The plasmids pBK-CMV and pBluescript II KS (Stratagene, La Jolla, CA, USA) were used for cloning. The plasmids pQE30 and *E. coli* M15 (pREP4) (Qiagen, Hilden, Germany) were used for expression.

Screening *T. yonseiensis* KB-1 genomic DNA library for a subtilisin-like protease

All DNA manipulations were carried out by using standard methods (Sambrook et al. 1989). *T. yonseiensis* genomic DNA was partially digested with *Sau*3AI, and fragments (2–12 kb in length) were ligated into a ZAP EXPRESS vector/*Bam*HI (Stratagene) that had been previously digested with *Bam*HI, and were packaged using Gigapack III Gold packaging extract (Stratagene). Based on the published sequences of subtilisin-like proteases (Choi et al. 1999), degenerate oligonucleotide primers were designed and synthesized to screen the genomic DNA library. A degenerate forward primer, 5′-AA(C/T) GG(G/A/T/C) CA(C/T) GG(G/A/T/C) AC(G/A/T/C) CA(C/T)-3′ was designed based on the NGHGTH sequence, which is conserved in the histidine region of the catalytic triad of subtilisin-like proteases. A reverse primer, 5′-(C/T)GG (C/T)GT (C/T)GC CAT (G/A/T/C)GA (G/A/T/C)GT (G/A/T/C)CC-3′, was derived from the sequence GTSMATP, which is conserved in the serine region of the catalytic triad of subtilisin-like proteases. Using the primers, a 585-bp polymerase chain reaction (PCR) product, designated pro II, was obtained and purified with a Gene Clean II kit (BIO101, Carlsbad, CA, USA). The pro II was labeled with ^{32}P by random priming with the Klenow fragment (Promega, Madison, WI, USA) and [α-^{32}P] dCTP (Amersham Pharmacia Biotech, Uppsala, Sweden) and used to screen the genomic DNA library.

Cloning and characterization of the protease gene

The library was plated on *E. coli* XL-1 blue and grown for 8 h at 37°C, and then plaques were transferred onto nitrocellulose membranes following the standard protocol (Sambrook et al. 1989). For hybridization, filters were prehybridized in hybridization solution [0.75 M sodium chloride; 0.075 M sodium citrate, pH 7.0; 0.5% sodium dodecyl sulfate (SDS); 100 μg/ml of salmon sperm DNA]; and 5 × Denhardt's solution at 65°C for 1 h. Hybridization was performed overnight at 65°C in the hybridization solution containing ^{32}P-labeled DNA probe as described above. Filters were washed with a high-stringency solution containing 15 mM sodium chloride, 1.5 mM sodium citrate (pH 7.0), and 0.1% SDS at 65°C, and then autoradiographed at –70°C with X-ray film. Positive plaques were identified after the film was developed and the autoradio-

graph was aligned to the plates. Clearly defined positive plaques were removed using a sterile Pasteur pipette and eluted in SM buffer (100 mM sodium chloride, 8 mM $MgSO_4$, 50 mM Tris–HCl (pH 7.5), and 2% gelatin). The plasmids were rescued according to the protocols of the ZAP EXPRESS predigested Gigapack cloning kit (Stratagene) into pBK-CMV phagemids with ExAssist helper phage. Plasmid DNA was then obtained and digested with *Apa*I and *Bam*HI to remove the insert and analyzed on 1% agarose gel in Tris-acetate-ethylenediaminetetraacetate (EDTA) (TAE) buffer. The clone having an open reading frame, *tayI*, was sequenced with an ABI 310 automatic sequencer with a terminator cycle sequencing ready reaction kit (PE Applied Biosystems, Warrington, UK). Sequence overlap was obtained by walking across the insert using synthetic oligonucleotide primers 18 or more bases long. Both strands were sequenced entirely at least once.

Computer analysis

Computer-assisted DNA- and protein-sequence analyses were performed using DNASIS version 7.0 and PROSIS software. Sequence similarity searches were also performed using by the BLAST program at the U.S. National Center for Biotechnology Information (NCBI) server, and the Clustal W multiple alignment program was used for sequence analysis (Thompson et al. 1994).

Construction of expression plasmid with *tayI*

Expression of thermicin in *E. coli* was carried out by using standard methods (Sambrook et al. 1989). The full-length sequence encoding the *tayI* proprotease was obtained from the genomic DNA of *T. yonseiensis* KB-1 by PCR amplification. Primers were designed to include a *Bam*HI site at the 5′ end and a *Sac*I site at the 3′ end of the tayI proprotease. The primers used were 5′-CTCAAAAGGATCCA AGGAAATCTATAC-3′ (sense primer) and 5′-ACAGAG CTCTTAAACTTTTTTTAAAGC-3′ (antisense primer). Purified PCR products were ligated into *Bam*HI–*Sac*I restriction sites of the pQE-30 expression vector containing the sequence encoding the 6 × His affinity tag (Qiagen). The ligation mixture was then subcloned into *E. coli* M15 strain (Qiagen) grown in Luria broth containing 50 µg each of kanamycin and ampicillin per ml. The resulting plasmid was designated pQEX1. The complete gene was sequenced.

Expression and purification of thermicin

For induction of pQEX1, isopropyl-β-D-thiogalactopyranoside (IPTG) was added to the culture broth ($O.D_{600}$, 0.6) to a final concentration of 1 mM. The cells were harvested every 3 h. Cells were broken by ultrasonic treatment as previously described (Sambrook et al. 1989). The cell debris was removed by centrifugation (14,000 *g*, 40 min, 4°C). The expression and purification of His-fused enzyme were carried out according to the manufacturer's instructions (Novagen, Madison, WI, USA). Thermicin was purified by nickel nitrilotriacetic acid (Ni-NTA) affinity chromatography in a column from an Ni-NTA kit (Novagen). Thermicin was eluted from the resin with 20 mM Tris–HCl, pH 7.9, containing 1 M imidazole/0.5 M NaCl as described in the manufacturer's instructions. The sample was diluted with 20 mM Tris–HCl, pH 7.0. The enzyme solution was then applied to an SP Sepharose ion-exchange column (Amersham Phamacia Biotech, Uppsala, Sweden), pre-equilibrated with the same Tris–HCl buffer, and the bound proteins were eluted with a linear gradient of 0 to 500 mM NaCl. Thermicin was eluted at a concentration of 0.5 M NaCl. This enzyme solution was dialyzed against 20 mM Tris–HCl, pH 7.0, containing 100 mM NaCl. The active fractions were pooled and concentrated. Protein concentrations were estimated by the dye-binding method (Bio-Rad, Richmond, CA, USA) by using bovine serum albumin as the standard.

Protease assay

Protease activities were measured at 75°C, with azocasein as the substrate by the Peek method previously described (Peek et al. 1992). One milliliter of 0.2% (w/v) azocasein in 100 mM hydroxyethylpiperazine ethanesulfonic acid (HEPES) buffer was preincubated for 10 min at 75°C, and the reaction was started by the addition of 10 µl of protease sample. After 10 min, the reaction was stopped by the addition of 0.5 ml of 15% (w/v) trichloroacetic acid. The reaction mixtures were cooled at room temperature for 15 min and centrifuged at 14,000 g for 5 min, and the absorbance of the supernatant was measured at 420 nm. One unit of protease activity was defined as giving an absorbance change of 1.0/min.

Electrophoresis

Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) was carried out essentially as described by Laemmli (Laemmli 1970) using a separation gel of 12% polyacrylamide. Amersham Phamacia's LMW marker was used to estimate molecular mass. To prevent autolysis during sample preparation for electrophoresis, 1 mM of PMSF, a protease inhibitor, was added to each protease samples. After electrophoresis, protein bands were visualized by Coomassie blue staining. For determining the proteolytic activity, SDS gel electrophoresis was also carried out with PAGE plates, but containing 0.1% gelatin as described previously (Heussen and Dowdle 1980; Kelleher and Juliano 1984). After electrophoresis, the gels were incubated in glycine buffer (100 mM, pH 8.5) at 80°C for 30 min.

N-terminal amino acid analysis

The heat-activated recombinant proteases were electroblotted from SDS-polyacrylamide gels (12%) to a polyvinylidene difluoride membrane (Immobilon transfer,

0.45 μm; Millipore, Bedford, MA, USA). The protein bands were sequenced by placing the membrane pieces in the upper cartridge of a Precise protein sequencing system (PE Applied Biosystems) equipped with an on-line phenylthiohydantoin amino acid analyzer.

Effect of temperature, thermostability, and pH on thermicin activity

Residual protease activity was assayed at 90°C with Mca-Arg-Pro-Lys-Pro-Tyr-Ala-Nva-Trp-Met-Lys(Dpn)-NH_2. All thermostability studies were performed at protease concentrations of approximately 1 μg/μl and in 3-(cyclohexylamino)-1-propanesulfonic acid (CAPS) buffer. To assess the effect of calcium on thermostability, EDTA was added to a final concentration of 10 mM immediately prior to incubation. Experiments were performed in 1.5-ml plastic reaction tubes, from which samples were removed throughout the time course. These were left on ice until assayed for protease. To determine the effect of pH on enzyme activity, the protease was assayed at 90°C with 1 μM Mca-Arg-Pro-Lys-Pro-Tyr-Ala-Nva-Trp-Met-Lys(Dpn)-NH_2 in the following buffers: 100 mM HEPES (pH 6.5 to 8.0), 100 mM Tris–HCl (pH 8.0), and 100 mM CAPS (pH 9.0 to 11.0).

Effect of chemicals and metal ions on thermicin activity

The heat-activated thermicin was diluted with CAPS buffer (pH 9.0) to 100 ng/ml. Different concentrations of PMSF, phenylglyoxal, and salicylaldehyde were incubated with enzymes dissolved in 20 mM potassium phosphate buffer, pH 7.0 (for the inhibition by PMSF and phenylglyoxal), and in 100 mM CAPS buffer, pH 9.0 (for the inhibition by salicylaldehyde), at 25°C for 30 min. After incubation, the residual protease activity was assayed at 80°C by measuring fluorescence intensity at λ_{ex} = 325 nm and λ_{em} = 393 nm with Mca-Arg-Pro-Lys-Pro-Tyr-Ala-Nva-Trp-Met-Lys(Dpn)-NH_2 as substrate. Different concentrations of $CaCl_2$, $ZnCl_2$, $CuSO_4$, $MnCl_2$, and $MgCl_2$ were assayed with enzyme dissolved in 100 mM CAPS buffer, pH 9.0, at 80°C with Mca-Arg-Pro-Lys-Pro-Tyr-Ala-Nva-Trp-Met-Lys(Dpn)-NH_2.

Kinetics studies

Activity against Mca-Arg-Pro-Lys-Pro-Tyr-Ala-Nva-Trp-Met-Lys(Dpn)-NH_2 was determined at 92°C in a Kontron fluorescence spectrometer SFM-25 (Milan, Italy). Fluorescent assays were performed at λ_{ex} = 325 nm and λ_{em} = 393 nm, which should encounter no interference from Trp residues. The amount of substrate hydrolysis was calculated based on the fluorescence values of the Mca-Arg-Pro-Lys-Pro-Gln standard solution after subtraction of the reaction blank value (a stopping solution was added before the enzyme). The kinetic parameters, K_m and k_{cat}, were determined for Mca-Arg-Pro-Lys-Pro-Tyr-Ala-Nva-Trp-Met-Lys(Dpn)-NH_2 over a substrate concentration range of 2–100 μM and calculated by double-reciprocal plots (Nagase et al. 1994). The Michaelis-Menten kinetic parameters for activity of the thermicin against Mca-Arg-Pro-Lys-Pro-Tyr-Ala-Nva-Trp-Met-Lys(Dpn)-NH_2 in the temperature range 40°–100°C and pH 6–11 were determined by fitting the rate data measured at substrate concentrations between 1 μM and 100 μM.

Substrate specificities of thermicin

The protease activity of the thermicin was measured with an Emax, precision microplate reader (Molecular Devices, Sunnyvale, CA, USA) in a 96-well plate. Assays were performed in 100 mM CAPS buffer (pH 9.0) in a total volume of 100 μl. The reactions were initiated by the addition of the thermicin at a final concentration of 10 nM, and the formation of *p*NA was monitored continuously at 405 nm. The amount of substrate hydrolyzed was calculated from the absorbance at 405 nm by using a molar extinction coefficient of 9,650 M^{-1} cm^{-1} for free *p*NA. The kinetic parameters K_m and k_{cat} were determined by analysis of double-reciprocal plots of the initial velocity as a function of substrate concentration.

Activity of thermicin against oxidized insulin β chain

The cleavage sites of the oxidized insulin β chain hydrolyzed by the thermicin were determined essentially as described previously (Peek et al. 1992). Insulin β chain (1 mg/ml) was incubated with 100 ng of thermicin (1 : 10,000, w/w) in 200 μl of 20 mM ammonium bicarbonate buffer at 90°C and pH 8.0. Samples were separated by reverse-phase high-performance liquid chromatography (HP1100, Hewlett Packard, Palo Alto, CA, USA) and applied on a C_{18} column (250 mm × 4 mm, particle size 5 μm, Hewlett Packard) pre-equilibrated with the same buffer at a flow rate of 0.5 ml/min. The compounds on the column were eluted with a linear gradient (0% to 50% within 50 min) of acetonitrile with monitoring at 215 nm and then collected for amino acid analysis. The mass spectrometry analyses were performed by using a Voyager RP (PerSeptive Biosystems, Framingham, MA, USA) time-of-flight dual-stage reflector mass spectrometer. The accelerating voltage used was 20 kV and the flight path was 1.1 m. Mass to charge (m/z) ratios of digests of the oxidized insulin β chain were detected by using GRAMS 386 software (Galactic Industries, Salem, NH, USA), and they were compared to the predicted m/z ratios of the amino acid sequence of the oxidized insulin β chain. The thermicin was then compared with the peptide maps of other reported enzymes (Peek et al. 1993).

Digestion of collagen types I, II, and IV and fibronectin

Protease activities were measured at 75°C, with azocasein as the substrate by the Peek method previously described (Peek et al. 1993). The collagenolytic activity was measured according to the method of Jeon (Jeon and Kim 1999). Col-

lagen types I (Collaborative Biomedical Products), II, and IV and fibronectin (Sigma Chemical Co.) were reacted with thermicin (100 ng/ml) at 80°C in 100 mM CAPS buffer (pH 9.0) for 20 min and analyzed by SDS-PAGE.

Construction of a phylogenetic tree

An extensive search of the scientific literature (PubMed: http://www.ncbi.nlm.nih.gov/PubMed/) and databases (Blast and Swiss-Prot) was performed with the BLAST2 program to collect the amino acid sequences of enzymologically well-characterized subtilisins (Altschul et al. 1997). Amino acid sequences of the 21 enzymes chosen, along with that of thermicin, were aligned using the Clustal W multiple-alignment program (Thompson et al. 1994). Sequence segments around catalytic residues (Asp29, His64, and Ser252 in thermicin numbering) were picked up and aligned to modify the initial automatic alignment of all sequences, because the thermicin sequenced in this study showed very low similarity to other enzymes. Sites containing gaps were excluded from all analyses. A phylogenetic tree was constructed from the aligned sequences by the neighbor-joining method (Saitou and Nei 1987; Siezen and Leunissen 1997) based on p-distances using the Clustal W program. The tree presented in Fig. 1 is a bootstrap consensus after 1,000 repetitions and was displayed by the Tree View program (Page 1996).

Nucleotide sequence accession number

The nucleotide sequence of the *tayI* gene of *T. yonseiensis* KB-1 has been deposited in the GenBank database under accession number AY028704.

Results

Cloning and sequencing of the *tayI* gene of *T. yonseiensis* KB-1

We constructed a genomic library of *T. yonseiensis*, screened it with a pro II sequence corresponding to that found in the subtilisin superfamily, and isolated a λ clone around 2.6 kb in size, which was designated as thermicin. A λ clone,

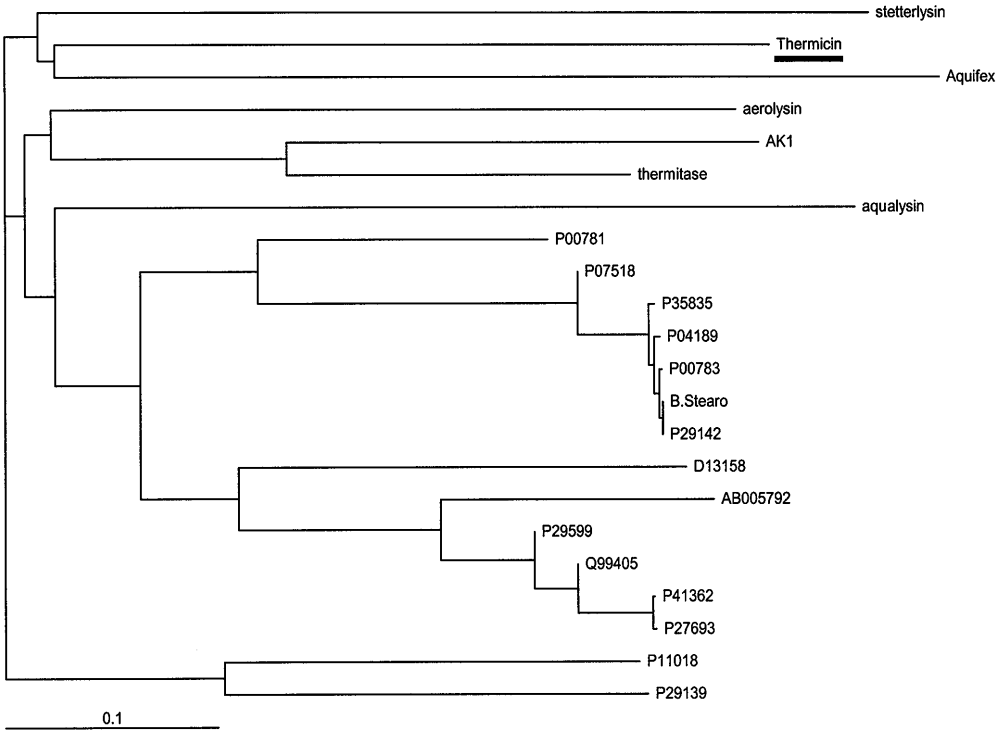

Fig. 1. Phylogenetic tree of subtilisins inferred from the amino acid sequence alignment of conserved regions around active sites. Conserved sequence segments around the catalytic triads (Asp29, His64, and Ser252 in thermicin numbering) of thermicin were aligned manually with those of other subtilisins of known sequence. *Bar* represents Knuc units. The values are based on a Kimura distance neighbor-joining analysis (1,000 replicates). Sources of sequences aligned: *hmhlys*, *Haloferax mediterranei* R4 (P71402); *bsak1*, *Bacillus* sp. strain AK1 (Q45670); *tstap*, *Thermoactinomyces* sp. E79 (Q56365); *tvther*, *Thermoactinomyces vulgaris* (P04072); *dnapv5*, *Dichelobacter nodosus* (Q46541); *dnbpr*, *Dichelobacter nodosus* (P42779); *dnapv2*, *Dichelobacter nodosus* (Q46540); *baalkp*, *Bacillus alcalophilus* PB92 (P27693); *bseyab*, *Bacillus subtilis* YaB (P20724); *taaqua*, *Thermus aquaticus* YT-1 (P08594); *trt41a*, *Thermus* rT41A (P80146); *alapr2*, *Alteromonas* sp. O-7 (Q53401); *vaproa*, *Vibrio alginolyticus* (P16588); *smserp*, *Serratia marcescens* IFO3046 (P09489); *smssp1*, *Serratia marcescens* (Q54483); *smssp2*, *Serratia marcescens* (Q54484); *llprtm*, *Lactococcus lactis* (cremoris) SK11 (P14308); *ldprtb*, *Lactobacillus delbrueckii* (*bulgaricus*) (Q9R6R1); *pfpyro*, *Pyrococcus furiosus* (P72186); *tsplst*, *Thermococcus stetteri* (O93635) (Saeki et al. 2000)

containing the longest sequence and around 2,605 bp in size, that shared a high sequence homology to the pro II sequence was used as a probe (Fig. 2.). The cloned sequence had an open reading frame of 1,200 bp that started with a putative translation initiation site at position 40 and could encode a polypeptide with 400 amino acid residues with a calculated molecular mass of 43 kDa, including a typical bacterial signal peptide sequence of 13 amino acids at the N-terminus and an 89-amino-acid proenzyme portion. The deduced amino acid sequence has 44% similarity and 37% identity to the sequences of other subtilisin-like proteases (data not shown). A putative Shine-Dalgarno sequence was found upstream of the GTG translation start codon. In addition, a potentially stable stem-loop terminator structure for the mRNA was found in the sequence directly downstream of the *tayI* gene.

Expression and purification of thermicin in *E. coli*

In the course of two purification steps (Ni-NTA affinity chromatography and SP Sepharose ion-exchange chromatography), a 595-fold purification was obtained (Table 1). For proper expression, we prepared a truncated construct of 400 amino acid residues without the N-terminal bacterial signal peptide but with a His tag. The recombinant protein was expressed at a relatively high level and could be purified by affinity chromatography. The recombinant protein migrated to a band of about 33 kDa in the SDS-PAGE analysis, which indicated that the recombinant protein was a mature form without a proenzyme portion. Conserved sequence segments around the catalytic triads (Asp29, His64, and Ser252 in thermicin numbering) of thermicin were aligned manually with those of other subtilisins of known sequence. The *tayI* gene showed about 37% sequence identity with subtilisin-like proteases. Amino acid sequences around catalytic triad residues are highly homologous with those of other serine proteases (Fig. 3). The 98-amino-acid sequence at the N-terminus was unique and did not show sequence similarity to the other proteases in the database (data not shown). Thus, consensus sequence segments of the 16 proteases, Ile23–Lys42, Asp60–Gly76, and Phe244–Leu264 (in thermicin numbering), were manually aligned, and the aligned segments combined with the pyrolysin group were used to construct a phylogenetic tree. Evolutionary distances among the proteases were computed, and a phylogenetic tree was constructed by a neighbor-joining algorithm, as shown in Fig. 1. Thermicin is quite distantly related to the other four clusters. Amino acid sequence comparison and phylogenetic analysis indicated that this enzyme can be classified into a new subgroup with respect to its molecular evolution, compared with other known subtilisins. The propeptide was not a part of the final folded active enzyme. Upon completion of the folding of the proenzyme, the propeptide was cleaved and removed from the active enzyme. The mechanism for the autoprocessing has been suggested to be intramolecular. In renaturation experiments, the autoprocessing of prosubtilisin E, as measured by an increase in subtilisin activity, increased almost linearly as the concentration of denatured prosubtilisin E increased (Ikemura and Inouye 1988). As a result of autocatalytic removal of the N-terminal prosequences, thermicin undergoes self-processing and activation. Thus, we performed N-terminal peptide sequencing to elucidate the accurate cleavage site of the recombinant enzyme. The N-terminal amino acid sequence of the 33-kDa protein was Leu-His-Ile-Ala-Thr-Gln-Glu-Ile, suggesting that the recombinant protein was processed to the mature form (Fig. 4). In spite of the presence of SDS in the gels, proteolytic activity was detectable (Fig. 4, lane 3). The constructed enzyme (45 kDa, protein band) was autolytically hydrolyzed to proteins 33 kDa and 12 kDa in size (Fig. 4, lane 2). The molecular mass of the small protein was identical to that of the propeptide, indicating that the small protein was the N-terminal part of the recombinant prothermicin including the His tag. A zymogram of the enzyme showed strong activity with the 33-kDa protein but no activity with the 12-kDa protein. However, we found some enzyme activity in the regions with a molecular mass of 45 kDa or higher. This result suggests that there was still some prothermicin left on the gel and that prothermicin also has some enzyme activity (Fig. 4, lane 3).

Characteristics of thermicin

Thermicin was active at a broad range of temperatures (40° to 95°C) and pH values (pH 6 to 11). The highest activity of thermicin was detected at 92.5°C and pH 9.0. The stability of the protein produced was examined by measuring the decrease in activity after incubation at high temperature. The protein's half-life was about 30 h at 80°C, indicating that the recombinant protein is highly thermostable. PMSF inhibited enzyme activity, suggesting that serine residues

Table 1. Purification of thermicin

Purification step	Total protein (mg)	Total activity (units)	Specific activity (units/mg)	Yield (%)	Purification (fold)
Cell extract	712	102	0.14	100	1
Ni-NTA affinity chromatography	1.42	87	61.1	85.3	436
SP Sepharose	0.85	71	83.3	69.6	595

For experimental details, see the text
Ni-NTA, nickel nitrilotriacetic acid

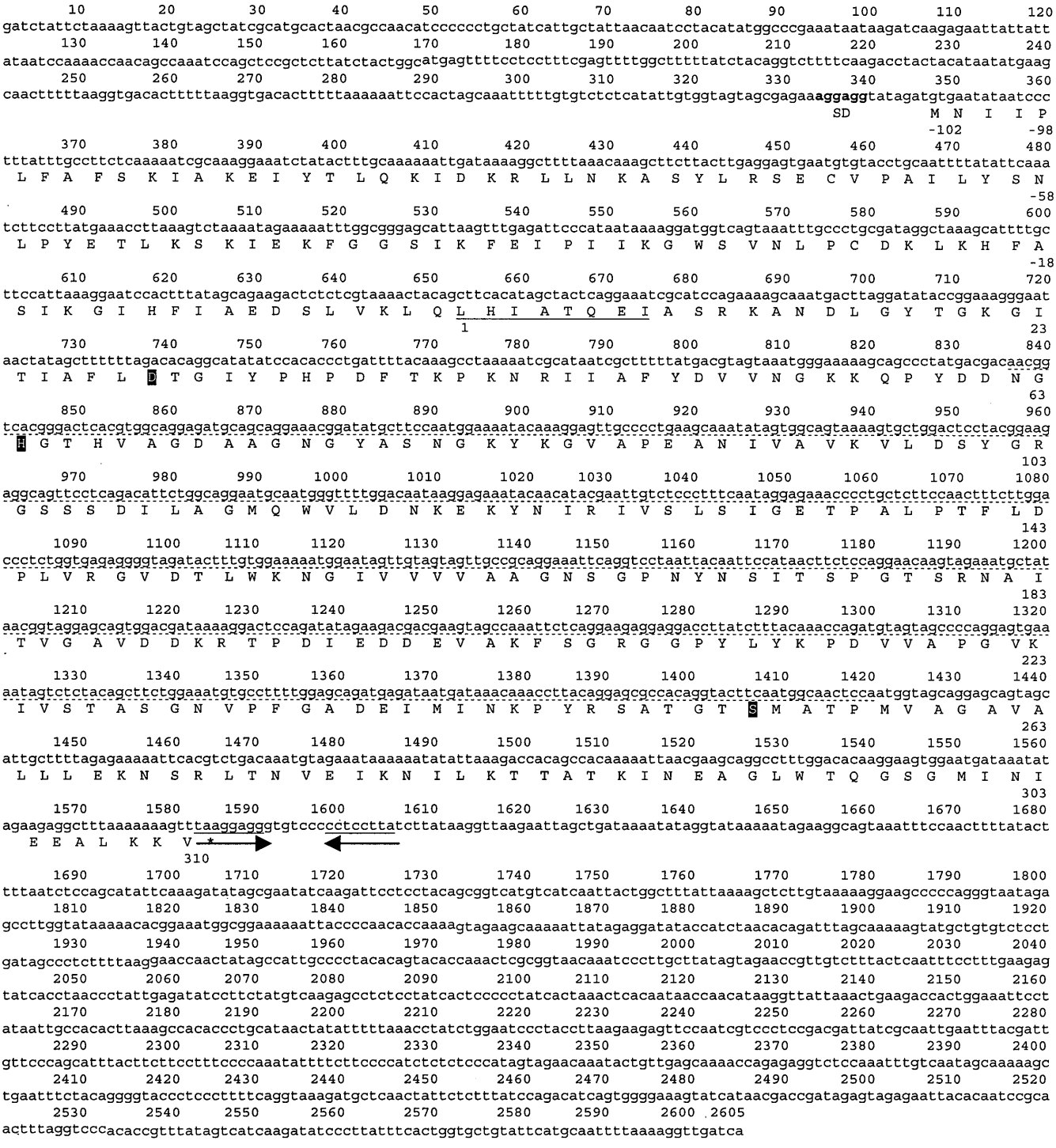

Fig. 2. Nucleotide and deduced amino acid sequences of the *tayI* gene. The deduced amino acid sequence of the gene product is indicated by the *single-letter codes* under the *nucleotide sequence*. A putative ribosome-binding site is indicated by *bold type*. The open reading frame (ORF) extends from Met (–102) to Val 310. The putative signal sequence is Met (–102) to Ala (–90). The N-terminal sequence of the thermicin mature form was determined by Edman degradation and is indicated by the *underlined amino acid sequence* (Leu1–Ile8). The probe sequence, pro II, is indicated by a *dotted line*. *Convergent arrows* show an inverted repeat downstream of the stop codon *taa* of the ORF. Amino acids in *black boxes* represent the catalytic residues commonly found in a serine-type protease

may be involved in enzyme catalysis. The enzyme activity was also slightly inhibited by the presence of 10 mM salicylaldehyde or 10 mM phenylglyoxal. Thermicin was also inhibited by transition-metal ions ($Cu^{2+} > Zn^{2+}$) (Table 2).

Digestion of oxidized insulin β chain

In order to determine substrate specificity and substrate preference with respect to amino acids, the insulin β chain

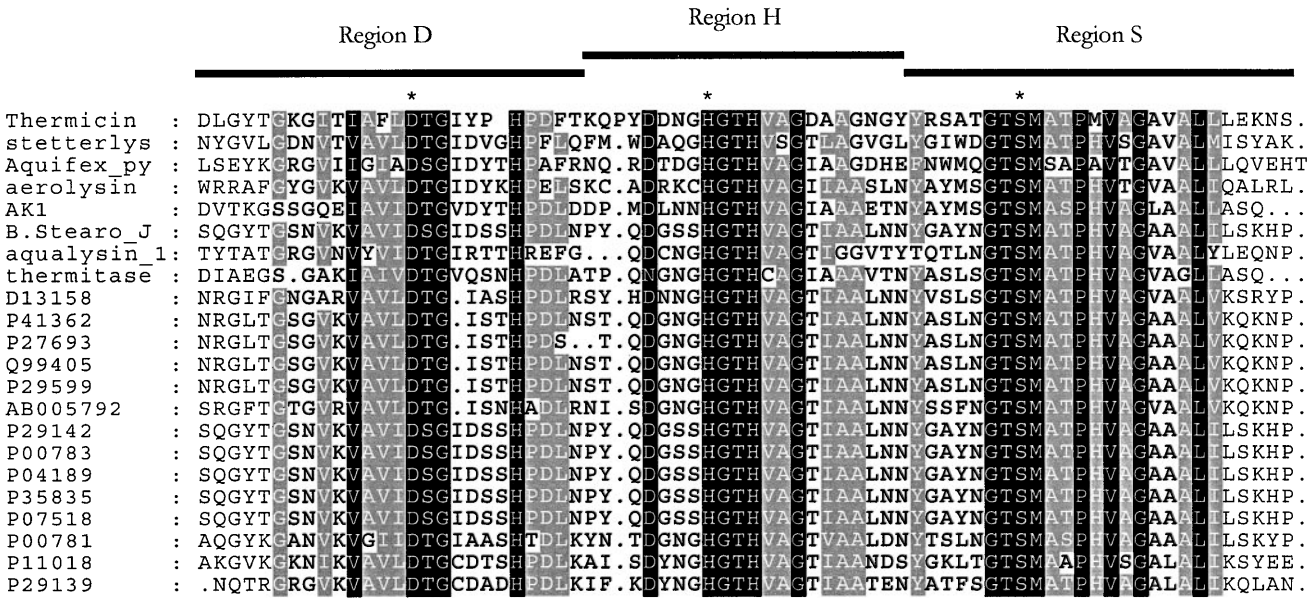

Fig. 3. Multiple sequence alignment among conserved sequence segments around the catalytic triads of thermicin and other subtilisin. Each amino acid sequence is indicated by *single-letter codes*. Sources of sequences aligned: *baalkp*, *Bacillus alcalophilus* PB92 (P27693); *bseyab*, *Bacillus subtilis* YaB (P20724); *dnapv5*, *Dichelobacter nodosus* (Q46541); *dnapv2*, *Dichelobacter nodosus* (Q46540); *dnbpr*, *Dichelobacter nodosus* (P42779); *trt41a*, *Thermus* rT41A (P80146); *alapr2*, *Alteromonas* sp. O-7 (Q53401); *vaproa*, *Vibrio alginolyticus* (P16588); *taaqua*, *Thermus aquaticus* YT-1 (P08594); *pfpyro*, Pyrococcus furiosus (P72186); *tsplst*, *Thermococcus stetteri* (O93635); *tstap*, *Thermoactinomyces* sp. E79 (Q56365); *tvther*, *Thermoactinomyces vulgaris* (P04072); *bsak1*, *Bacillus* sp. strain AK1 (Q45670); *hmhlys*, *Haloferax mediterranei* R4 (P71402). A common catalytic triad of the three amino acids: Asp (electrophile), *region D*; His (base), *region H*; and Ser (nucleophile), *region S* is shown by *asterisks*

Table 2. Effect of chemicals and metal ions on the activity of thermicin[a]

Compound	Residual thermicin activity (%)		
Concentration (mM)	0.1	1	10
Control	100	100	100
Phenylglyoxal	93.9	92.5	38.1
PMSF	101.8	80.0	9.5
Salicylaldehyde	91.6	59.7	16.6
$CaCl_2$		104.5	99.3
$ZnCl_2$		76.7	30.7
$CuSO_4$		55.5	26.4
$MnCl_2$		96.5	91.7
$MgCl_2$		101.0	113.5

PMSF, phenylmethanesulfonyl fluoride; CAPS, 3-(cyclohexylamino)-1-propanesulfonic acid

[a]Chemicals were added to solutions of thermicin (200 ng/ml in 100 mM CAPS buffer, pH 9.0) to give the final concentrations specified. After incubation at room temperature for 30 min, protease activity was determined by the methods described in the text

was treated with thermicin. The cleavage of the oxidized insulin β chain was detected after a 20-min incubation assay at 80°C. The products of thermicin digestion of the insulin β chain was analyzed by matrix assisted laser desorption/ionization-time of flight (MALDI-TOF) mass spectrometry (Fig. 5). The cleavage specificity that thermicin shows at Gly and Pro is unique. This characteristic has not been found in any of the reported proteases previously (Abraham et al. 1995; Peek et al. 1993).

Substrate specificities of thermicin

Thermicin showed high activity with the synthetic peptide L-Gly-Pro-*p*NA as substrate. However, it showed a significantly low activity (about 1/17th that of L-Pro-*p*NA), indicating that some amino acid in the P2 site might be also critical for catalysis. No detectable activity was observed with *N*-succinyl-Gly-Gly-Gly-*p*NA, *N*-succinyl-Gly-Phe-Gly-*p*NA, *N*-CBZ-Gly-Gly-Leu-*p*NA, *p*Gly-Phe-Leu-*p*NA, or *N*-succinyl-Ala-Ala-Pro-Leu-*p*NA (data not shown). L-Phe-*p*NA was hydrolyzed by thermicin at a rate that corresponded to an activity of about three-fifth that of thermicin against L-Gly-Pro-*p*NA (Table 3).

Digestion of collagen types I, II, and IV and fibronectin

Collagen types I, II, and IV and fibronectin were also treated with thermicin to determine substrate preference. As shown in Fig. 6, thermicin was able to hydrolyze collagen types I and II and fibronectin completely, whereas it was able to hydrolyze type IV only partially. This result suggests that collagen types I and II and fibronectin are better substrates than collagen type IV.

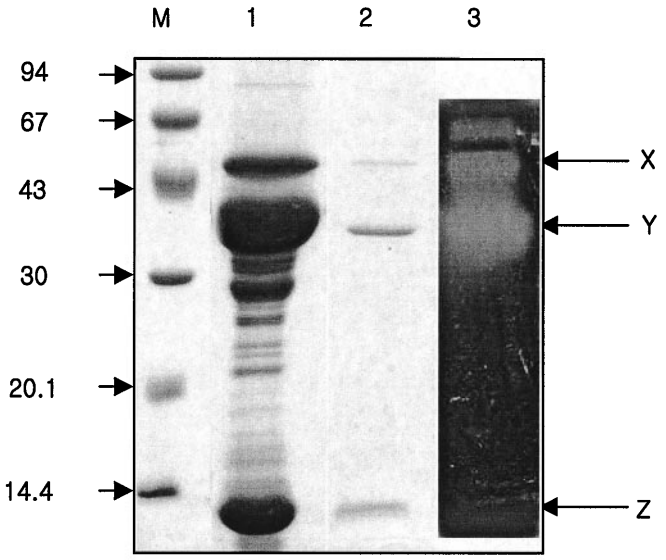

Fig. 4. Purification of thermicin expressed in *E. coli.* Expression and purification of the thermicin was analyzed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE): protein molecular mass markers (*M*), prothermicin purified by nickel nitrilotriacetic acid (Ni-NTA) affinity chromatography (*lane 1*), thermicin purified by ion-exchange column, SP column (*lane 2*) and active gel (*lane 3*). Arrows with *X*, *Y*, and *Z* indicate the positions of prothermicin, mature thermicin, and propeptide added His tags, respectively. Electrophoretic analysis of the purified protease (0.5 μg of protein) in SDS 12% (w/v) polyacrylamide gels. For the detection of proteolytic activity, the gels (*lane 3*) contained 0.1% gelatin, and incubation was performed at 80°C and pH 8.5 for 30 min. The *numbers on the left* represent the molecular masses of the following marker proteins in kilodaltons: phosphorylase *b*, 94; bovine serum albumin, 67; ovalbumin, 43; carbonic anhydrase, 30; trypsin inhibitor, 20.1; a-lactalbumin, 14.4

Discussion

A gene, *tayI,* encoding the thermostable subtilisin-like protease thermicin was cloned from *Thermoanaerobacter yonseiensis* KB-1. The gene has characteristics similar to those of other genes in the subtilisin family, which is the largest class of serine-type proteases. In the sequenced region, there is only one open reading frame (ORF) that encodes more than 100 amino acids. This ORF contains the serine-type protease consensus sequence present in the sequence tag. GTG was identified as the translational initiation codon because it was located downstream of the putative ribosome-binding site. Many subtilisin family proteases are translated as the prepro form containing a secretion signal sequence. The signal sequence (presequence) is processed after translocation through the periplasmic membrane, and the following sequence (prosequence) is further processed to become the mature form (Choi et al. 1999). Thermicin was also produced as a 45-kDa precursor protein, comprising an N-terminal signal peptide (13 amino acid residues), an 89-residue N-terminal prosequence, and a 311-residue mature sequence. Prothermicin was produced in *E. coli* as a 45 kDa protease containing the N-terminal prosequence. But after the size of the produced protein was reduced by heat treatment at 80°C for 10 min, it showed full enzyme activity. This result indicated that thermicin produced in *E. coli* undergoes self-processing at an elevated temperature and that activation was caused by possibly autocatalytic removal of N-terminal prosequences. Autolysis is a problem when a protease is handled at high concentrations, but it can be prevented by inclusion of the serine protease inhibitor PMSF during preincubation at

Fig. 5. Cleavage of the oxidized insulin β chain (bovine) by thermicin

m/z	Amino acid sequence of Insulin β chain fragment
3495.9	F V N Q H L C [SO_3H] G S H L V E A L Y L V C [SO_3H] G E R G F F Y T P K A
2493.66	V N Q H L C [SO_3H] G S H L V E A L Y L V C [SO_3H] G E R G
2112.29	E A L Y L V C [SO_3H] G E R G F F Y T P K A
1798.96	Y L V C [SO_3H] G E R G F F Y T P K A
1714.85	F V N Q H L C [SO_3H] G S H L V E A L
1323.4	V C [SO_3H] G E R G F F Y T P

Table 3. Comparison of kinetic parameters for the hydrolysis of synthetic peptides and fluorogenic peptide substrates by thermicin[a]

Substrate	K_m (μM)	k_{cat} (s^{-1})	k_{cat}/K_m (10^5 s^{-1} M^{-1})
L-Pro-*p*NA	176.30 (±11)	149.27 (±3)	8.47
L-Gly-Pro-*p*NA	54.16 (±2)	769.33 (±2)	142.05
L-Phe-*p*NA	35.95 (±1)	312.54 (±4)	86.94
Mca-Arg-Pro-Lys-Pro-Tyr-Ala-Nva-Trp-Met-Lys(Dnp)-NH_2	1,315.79 (±71)	108.57 (±3)	0.83

*p*NA, *p*-nitroaniline

[a] Enzyme activity was assayed using acetone-precipitated sample by the method of Wells et al. (1987), modified slightly
L-Pro-*p*NA, L-Gly-Pro-*p*NA, L-Phe-*p*NA (Sigma) and Mca-Arg-Pro-Lys-Pro-Tyr-Ala-Nva-Trp-Met-Lys(Dnp)-NH_2 (Bachem) were used as the substrate

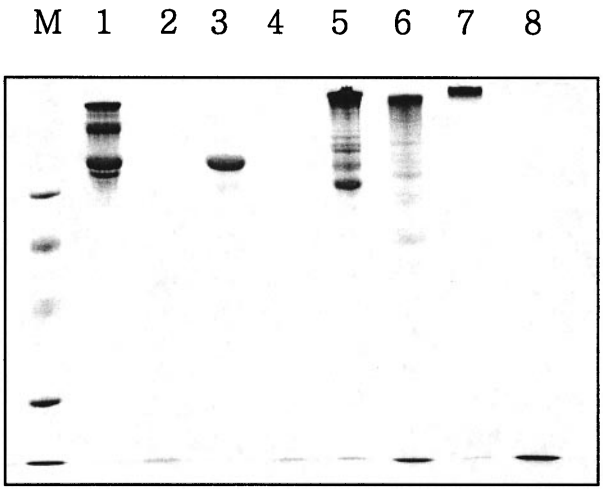

Fig. 6. Proteolytic digestion of collagen and fibronectin by thermicin. Collagen types I and II were dissolved in 0.1 M acetic acid and dialyzed against phosphate-buffered saline (PBS) at 4°C. Human type IV collagen and fibronectin were dissolved in 0.1 M acetic acid and dialyzed against PBS for 6 h at 4°C prior to use. Proteolytic digestions of various collagens and fibronectin were analyzed by SDS-PAGE. Various collagens and fibronectin were incubated without thermicin (*lanes 1, 3, 5, 7*) and with thermicin (1 μg) for 20 min at 80°C in 3-(cyclohexylamino)-1-propanesulfonic acid (CAPS) buffer (*lanes 2, 4, 6, 8*). Protein substrates were type I collagen (10 μg: without enzyme, *lane 1*; with enzyme, *lane 2*), type II collagen (10 μg: without enzyme, *lane 3*; with enzyme, *lane 4*), type IV (5 μg: without enzyme, *lane 5*; with enzyme, *lane 6*), and fibronectin (10 μg: without enzyme, *lane 7*; with enzyme, *lane 8*). *M*, protein molecular mass marker

80°C. However, PMSF does not appear to inhibit the conversion of the proenzyme to the mature protease (Li and Inouye 1996). This protein shows extreme stability at high temperatures and at a broad range of pH values. Maximum activity with Mca-Arg-Pro-Lys-Pro-Tyr-Ala-Nva-Trp-Met-Lys(Dpn)-NH_2 was observed at 92.5°C and pH 9.0. It maintains up to 50% of its activity even when held at 80°C for 30 h. Structural analysis has shown that among the major factors that stabilize proteins in hyperthermophiles are ionic bonds or salt bridges between charged amino acids (Siezen and Leunissen 1997). Disulfide bonds also contribute to the stability of a protease(Yip et al. 1995). However, the disulfide bond model for thermostability may not be applicable in the case of this enzyme because all cysteine residues were located in the prosequence. In addition, thermicin does not have a calcium-binding motif, according to the database analysis, suggesting that its thermostability is calcium-independent. Therefore, aside from those factors mentioned above, there may be other factors involved that account for the thermostability of thermicin. Thermicin was inhibited by salicylaldehyde and phenylglyoxal,. indicating that lysine and arginine residues may be involved in the catalysis by this enzyme. Although 23 lysine and ten arginine residues were found in mature thermicin, only one Lys96 residue has been found in the region conserved in various enzymes, indicating that Lys96 may play an essential role. Thermicin was also inhibited by $CuSO_4$, and $ZnCl_2$. A possible cause for the metal ion inactivation may be interactions with metal-sensitive amino acid residues leading to oxidation or complex formation (Basak et al. 1999).

The protease cleavage sites of the oxidized insulin β chain have been used to define specificity (Peek et al. 1993). The insulin β chain was cleaved at the Val, Leu, Phe, Gly, and Pro carboxyl groups. Similar cleavage patterns at Val, Leu, and Phe have also been reported for other proteases such as Rt41A proteinase, aqualysin I, and proteinase K (Peek et al. 1993). However, the cleavage at Gly and Pro by thermicin was unique. To investigate the P1 site, synthetic substrates were used for thermicin catalysis. As shown in Fig. 5, proline was recognized by thermicin, but synthetic substrates containing glycine at the P1 site were not digested, indicating that this protease recognizes not only the P1 site but also the P2 site. The triplet sequence Gly-X-Y is common within the triple helical domain of collagens, where X and Y are often proline and hydroxyproline (Harrington 1996). As expected, thermicin was able to hydrolyze collagen types I, II, and IV and fibronectin, which are components of the extracellular matrix.

The 21 proteases analyzed phylogenetically were unequivocally grouped into four clusters on the basis of the amino acid sequence analysis: the pyrolysin, thermitase, proteinase K, and subtilisin groups. This result is consistent with the phylogenetic tree reported by Siezen and Leunissen (1997). Multiple alignments also allowed a phylogenetic tree to be constructed on the basis of a comparison of the thermicin precursor with the proteases (data not shown). Positions containing gaps were ignored. Interestingly, thermicin was quite distantly related to the other four clusters. However, thermicin was considered to be more closely related to the thermostable proteases from the hyperthermophilic archaea and thermophilic bacteria.

In conclusion, a novel thermostable protease that shows unique cleavage specificity at Gly and Pro was discovered in *T. yonseiensis* KB-1 isolated from hot springs. This enzyme is categorized as a new class of subtilisin, distinctly different from the pyrolysin, thermitase, and proteinase K groups, and from other subtilisin groups. This enzyme will be industrially applicable for the degradation of collagen at elevated temperatures without contamination.

Acknowledgments This work was supported by a grant from the Korea Science and Engineering Foundation through the Bioproducts Research Center at Yonsei University (1999G0102).

References

Abraham LD, Chow DT, Breuil C (1995) Characterization of the cleavage specificity of a subtilisin-like proteinase from *Ophiostoma piceae* by liquid chromatography/mass spectrometry and tandem MS. FEBS Lett 374:208–210

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389–3402

Basak A, Toure BB, Lazure C, Mbikay M, Chretien M, Seidah NG (1999) Enzymic characterization in vitro of recombinant proprotein convertase PC4. Biochem J 343:29–37

Blumentals II, Robinson AS, Kelly RM (1990) Characterization of sodium dodecyl sulfate-resistant proteolytic activity in the hyperthermophilic archaebacterium *Pyrococcus furiosus*. Appl Environ Microbiol 56:1992–1998

Chavez Croocker P, Sako Y, Uchida A (1999) Purification and characterization of an intracellular heat-stable proteinase (pernilase) from the marine hyperthermophilic archaeon *Aeropyrum pernix* K1. Extremophiles 3:3–9

Choi IG, Bang WG, Kim SH, Yu YG (1999) Extremely thermostable serine-type protease from *Aquifex pyrophilus*. Molecular cloning, expression, and characterization. J Biol Chem 274:881–888

Cowan DA, Smolenski KA, Daniel RM, Morgan HW (1987) An extremely thermostable extracellular proteinase from a strain of the archaebacterium *Desulfurococcus* growing at 88°C. Biochem J 247:121–133

Harrington DJ (1996) Bacterial collagenases and collagen-degrading enzymes and their potential role in human disease. Infect Immun 64:1885–1891

Heussen C, Dowdle EB (1980) Electrophoretic analysis of plasminogen activators in polyacrylamide gels containing sodium dodecyl sulfate and copolymerized substrates. Anal Biochem 102:196–202

Ikemura H, Inouye M (1988) In vitro processing of pro-subtilisin produced in *Escherichia coli*. J Biol Chem 263:12959–12963

Jeon OH, Kim DS (1999) Molecular cloning and functional characterization of a snake venom metalloprotease. Eur J Biochem 263: 526–533

Kelleher PJ, Juliano RL (1984) Detection of proteases in polyacrylamide gels containing covalently bound substrates. Anal Biochem 136:470–475

Kim BC, Grote R, Lee DW, Antranikian G, Pyun YR (2001) *Thermoanaerobacter yonseiensis* sp. nov., a novel extreme thermophilic, xylose-utilizing bacterium that grows at up to 85°C. Int J Syst Evol Microbiol 51:1539–1548

Klingeberg M, Hashwa F, Antranikian G (1991) Properties of extremely thermostable proteases from anaerobic hyperthermophilic bacteria. Appl Microbiol Biotechnol 34:715–719

Klingeberg M, Galunsky B, Sjoholm C, Kasche V, Antranikian G (1995) Purification and properties of a highly thermostable, sodium dodecyl sulfate-resistant and stereospecific proteinase from the extremely thermophilic archaeon *Thermococcus stetteri*. Appl Environ Microbiol 61:3098–3104

Koch R, Zablowsky P, Spreinat A, Antranikian G (1990) Extremely thermostable amylolytic enzyme from the archaebacterium *Pyrococcus furiosus*. FEMS Microbiol Lett 71:21–26

Laemmli UK (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227:680–685

Li Y, Inouye M (1996) The mechanism of autoprocessing of the propeptide of prosubtilisin E: intramolecular or intermolecular event? J Mol Biol 262:591–594

Morikawa M, Izawa Y, Rashid N, Hoaki T, Imanaka T (1994) Purification and characterization of a thermostable thiol protease from a newly isolated hyperthermophilic *Pyrococcus* sp. Appl Environ Microbiol 60:4559–4566

Nagase H, Fields CG, Fields GB (1994) Design and characterization of a fluorogenic substrate selectively hydrolyzed by stromelysin I (matrix metalloproteinase-3). J Biol Chem 269:20952–20957

Ng T K, Kenealy WR (1986) Industrial application of thermostable enzymes. In: Brock TD (ed) Thermophiles, general, molecular, and applied microbiology. Wiley, New York, pp 197–215

Page RDM (1996) Tree View: an application to display phylogenetic trees on personal computers. Comput Appl Biosci 12:357–358

Peek K, Daniel RM, Monk C, Parker L, Coolbear T (1992) Purification and characterization of a thermostable proteinase isolated from *Thermus* sp. strain Rt41A. Eur J Biochem 207:1035–1044

Peek K, Veitch DP, Prescott M, Daniel RM, MacIver B, Bergquist PL (1993) Some characteristics of a proteinase from a thermophilic *Bacillus* sp. expressed in *Escherichia coli*: comparison with the native enzyme and its processing in *E. coli* and in vitro. Appl Environ Microbiol 59:1168–1175

Rao MB, Tanksale AM, Ghatge MS, Deshpande VV (1998) Molecular and biotechnological aspects of microbial proteases. Microbiol Mol Biol Rev 62:597–635

Saeki K, Okuda M, Hatada Y, Kobayashi T, Ito S, Takami H, Horikoshi K (2000) Novel oxidatively stable subtilisin-like serine proteases from alkaliphilic *Bacillus* spp.: enzymatic properties, sequences, and evolutionary relationships. Biochem Biophys Res Commun 279: 313–319

Saitou N, Nei M (1987) A neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 44: 406–425

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Laboratory Press, Cold Spring Harbor, NY, USA

Siezen RJ, Leunissen JA (1997) Subtilases: the superfamily of subtilisin-like serine proteases. Protein Sci 6:501–523

Thompson JD, Higgins DG, Gibson TJ (1994) Clustal W: improving the sensitivity of progressive multiple sequence alignments through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673–4680

Völkl P, Markiewicz P, Stetter KO, Miller JH (1994) The sequence of a subtilisin-type protease (aerolysin) from the hyperthermophilic archaeum *Pyrobaculum aerophilum* reveals sites important to thermostability. Protein Sci 3:1329–1340

Wells JA, Powers DB, Bott RR, Graycar TP, Estel DA (1987) Designing substrate specificity by protein engineering of electrostatic interactions. Proc Natl Acad Sci U S A 84:1219–1223

Yip KS, Stillman TJ, Britton KL, Artymiuk PJ, Baker PJ, Sedelnikova SE, Engel PC, Pasquo A, Chiaraluce R, Consalvi V (1995) The structure of *Pyrococcus furiosus* glutamate dehydrogenase reveals a key role for ion-pair networks in maintaining enzyme stability at extreme temperatures. Structure 3:1147–1158