Identification and characterization of a novel spore-associated subtilase from *Thermoactinomyces* sp. CDF

Guyue Cheng, Peiwei Zhao, Xiao-Feng Tang and Bing Tang

Correspondence
Bing Tang
tangb@whu.edu.cn

State Key Laboratory of Virology, College of Life Sciences, Wuhan University, Wuhan 430072, PR China

A gene encoding a spore-associated subtilase, designated protease CDF, was cloned from *Thermoactinomyces* sp. CDF and expressed in *Escherichia coli*. The enzyme gene is translated as a proform consisting of a 94 aa propeptide and a 283 aa mature protease domain. Phylogenetic analysis revealed that this enzyme belonged to the subtilisin family, but could not be grouped into any of its six known subfamilies. The mature protease CDF has an unusually high content of charged residues, which are mainly distributed on the enzyme surface. The recombinant proform of protease CDF formed inclusion bodies, but could be efficiently converted to the mature enzyme when the inclusion bodies were dissolved in alkaline buffers. The proform underwent a two-step maturation process, wherein the N-terminal part (85 residues) of the propeptide was autoprocessed intramolecularly, and the remaining 9-residue peptide was further processed intermolecularly. Protease CDF exhibited optimal proteolytic activity at 50–55 °C and pH 10.5–11.0. The enzyme was stable under high-pH conditions (pH 11.0–12.0), and NaCl could stabilize the enzyme at lower pH values. In addition, the enzyme was not dependent on calcium for either maturation or stability. By immunoblot analysis, protease CDF was found to be associated with spores, and could be extracted from the spores with 2 M KCl and alkaline buffers without damaging the coat layer, demonstrating that the protease CDF is located on the surface of the spore coat.

Received 30 May 2009
Revised 16 July 2009
Accepted 18 August 2009

INTRODUCTION

Thermoactinomyces species are thermophilic filamentous bacteria, closely related to the genus *Bacillus*, and flourish in decaying hay, compost and other high-temperature habitats. These bacteria produce highly resistant endospores at the tips of the hyphae, enabling them to survive in harsh environments over long periods, up to 9000 years (Nilsson & Renberg, 1990). Resembling those of *Bacillus* species, the spores of *Thermoactinomyces* species have highly ordered structures consisting of a core, cortex and coat, as well as an additional balloon-like exosporium surrounding the coat layer in some cases (Lacey, 1989; Yoon *et al.*, 2005). Although the spores are known as the dormant form of bacteria, many spore-associated enzyme activities have been identified, especially those of *Bacillus* species, where proteases are crucial for spore germination (Hilbert & Piggot, 2004). In addition, some spore-associated proteases contribute to the virulence of bacteria

(Ramarao & Lereclus, 2005). With regard to *Thermoactinomyces* species, esterase activity has been detected in the spores of *Thermoactinomyces vulgaris* by using fluorogenic substrates (Gazenko *et al.*, 1998). However, spore-associated proteases have not yet been identified and characterized in *Thermoactinomyces* species.

Similar to fungi and actinomycetes, *Thermoactinomyces* species can spread their spores into the air, and the airborne spores are recognized as important air pollutants. Inhalational exposure to the spores of *Thermoactinomyces*, as well as to those of fungi and actinomycetes, may cause respiratory disorders in humans and livestock. Farmer's lung disease (FLD) represents a classic example of lung hypersensitivity pneumonitis (Pepys *et al.*, 1963). The immunopathological reactions in the lungs may involve a specific hypersensitivity response to the spore antigens and a non-specific response to spore-associated biologically active components, including proteases, which have long been known as respiratory hazards (Pauwels *et al.*, 1978; Yike *et al.*, 2007). Protease activity has been described in antigen preparations of *Thermoactinomyces* species, and antibodies to two serine proteases from *Thermoactinomyces candidus* were detected in the blood of FLD patients, implying a possible relevance of the proteases in FLD (Roberts *et al.*, 1983).

Abbreviation: FLD, farmer's lung disease.

The GenBank/EMBL/DDBJ accession number for the gene encoding the protease CDF of *Thermoactinomyces* sp. CDF is EF108326.

Three supplementary figures are available with the online version of this paper.

Thermoactinomyces species are generally used as a source of thermostable proteases, among which thermitase from *T. vulgaris* is the best known (Teplyakov *et al.*, 1990). The extracellular serine proteases produced by *Thermoactinomyces* sp. HS682 (Tsuchiya *et al.*, 1992) and *Thermoactinomyces* sp. E79 (Lee *et al.*, 1996) are not only heat-resistant, but also stable under highly alkaline conditions. In addition, intracellular alkaline proteases have been described for *Thermoactinomyces* sp. HS682 (Tsuchiya *et al.*, 1997) and *T. vulgaris* (Schalinatus *et al.*, 1983b). Besides the proteases that are present in soluble forms, minor proteases have also been found to be linked to the cytoplasmic membrane of *T. vulgaris* (Schalinatus *et al.*, 1983a).

In this study, a novel subtilase (protease CDF) was identified in the spores of a newly isolated *Thermoactinomyces* sp. CDF. The gene encoding this enzyme was cloned and expressed in *Escherichia coli*. The biochemical properties, maturation process and localization of the enzyme were investigated. To our knowledge, protease CDF is the first spore-associated protease to be identified and characterized in *Thermoactinomyces* species.

METHODS

Reagents. Restriction enzymes and T4 DNA ligase were purchased from TaKaRa. Taq and *Pfu* DNA polymerases were purchased from BioStar. Proteinase K was from Amresco. Azocasein, azo-dye impregnated collagen (azocoll), keratin-azure, elastin-orcein and

oxidized insulin B chain were from Sigma. All other reagents were the purest ones commercially available.

Strains, growth conditions and preparation of spores. The bacterium *Thermoactinomyces* sp. CDF was isolated from the campus soil of Wuhan University, China. The bacteria were grown in 250 ml flasks containing 50 ml of Luria–Bertani (LB) medium at 55 °C for 2 days with constant shaking (150 r.p.m.), and were used for extraction of genomic DNA. Spores were prepared by culturing the bacterium on LB agar plates at 55 °C for 2 days. The lawn culture was gently scraped from the surface of the agar plates and washed three times with normal saline [0.9% (w/v) NaCl]. The pellet was treated with lysozyme (0.1 mg ml⁻¹) at 37 °C for 10 min in 50 mM Tris/HCl (pH 8.0) and then sonicated for ten 15 s bursts, each separated by 2 min cooling on ice. Thereafter, the suspension was subjected to a short low-speed centrifugation (4000 g, 5 min) to collect the spores. The spores were washed with normal saline by repeated centrifugation until the contaminated cell debris had been removed. When necessary, the suspension was resonicated before centrifugation. Microscopic examination was used to confirm that the collected spores were free of contaminants.

DNA manipulation, plasmid construction and mutagenesis. The genomic DNA of *Thermoactinomyces* sp. CDF was extracted according to the method of Orsini & Romano-Spica (2001), except that the cells were disrupted by grinding with a mortar and pestle, rather than microwave treatment. A fragment of the gene (*cdf*) encoding protease CDF was amplified from the genomic DNA with *Pfu* DNA polymerase by PCR employing two consensus-degenerate hybrid oligonucleotide primers (CODEHOPs), F and R (Table 1), designed on the basis of two highly conserved amino acid sequences around the catalytic residues His and Ser of subtilases (Wu *et al.*, 2004) (Supplementary Fig. S1a). The amplified DNA fragments were inserted into the *EcoRI* restriction site of pUC18 for DNA sequencing.

Table 1. Oligonucleotide primers used in this study

Primer	Nucleotide sequence*
P1	5'-CGGGATCCAGAGTTTGATCCTGGCTCAGAACGAACGCT-3'
P2	5'-CGGGATCCTACGGCTACCTTGTACGACTTCACCCC-3'
F	5'-CGGAATTCGATGGCAACGGCCAYGGNACNCA-3'
R	5'-CGGAATTCGAGGTGTGGCCATNGANGTNCC-3'
TU1	5'-GGATTCGCTGACTTTGTCCATG-3'
TU2	5'-GATGCACCAAGTTGATCGCAC-3'
TU3	5'-CAACCCCCACAATCCCGGCCTC-3'
TD1	5'-GGCGGCAGGGAATCAAGG-3'
TD2	5'-GAAACCATCGCGGTGACATC-3'
TD3	5'-ACCGGATCCTGTCTGCCTGGCTG-3'
AD1	5'-NTCGASTWTSWGTT-3'
AD2	5'-NGTCGASWTGANAWGAA-3'
AD2	5'-WGTGNAGWANCANAGA-3'
AD4	5'-AGWGNAGWANCAWAGG-3'
PE1	5'-CGGAATTCATATGGAATACACCAATGGTCTGCTC-3'
PE2	5'-CGCCTCGAGATATCGGTAGATCTTGTGATAAAAC-3'
PN1	5'-CGGGATCCTTACGTTATCCGGATGTCGCTTTCC-3'
PN2	5'-CGGGATCCTTACAATGTTCCGACATAAGGTTCCG-3'
MTA	5'-GTTCCGGTGATCGCTATCCGGATGTC-3'
MLP	5'-CAGGAGTGTTCCGGTGTTCGACATAAG-3'
MSA	5'-GAGCGGGACTGCCATGGCGGTTC-3'

*Underlined sections represent restriction enzyme sites; Y=C or T; S=G or C; N=A, C, G or T; W=A or T. Bold nucleotides represent mutant sites. Italicized sequences represent initiation and termination codons.

The 5'- and 3'-ends of the *cdf* gene were amplified from genomic DNA by thermal asymmetrical interlaced PCR (TAIL-PCR) (Liu & Whittier, 1995). Three antisense primers for 5'-end TAIL PCR (TU1, TU2 and TU3) and three sense primers for 3'-end TAIL PCR (TD1, TD2 and TD3) were designed based on the known partial nucleotide sequence of the *cdf* gene (Table 1 and Supplementary Fig. S1a). The arbitrary degenerate primers used were AD1, AD2, AD3 and AD4 (Table 1). The amplified 5'- and 3'-end DNA fragments were ligated into the pMD18-T vector (TaKaRa), followed by DNA sequencing.

The *cdf* gene was amplified from genomic DNA with *Pfu* DNA polymerase by PCR employing primers PE1 and PE2 (Table 1), and inserted into the *NdeI*-*XhoI* sites of pET26b to construct the expression plasmid pCDFC for the proform of protease CDF fused with a His-tag at the C terminus. Meanwhile, the gene fragments encoding the N-terminal propeptides N1 [residues Met(-94) to Thr(-10)] and N2 [residues Met(-94) to Leu(-1)] (see Fig. 2) were also amplified from the genomic DNA with *Pfu* DNA polymerase by using primer pairs PE1/PN1 and PE1/PN2 (Table 1), respectively, and were inserted into the *NdeI*-*BamHI* restriction sites of pET26b to construct expression plasmids pCN1 and pCN2 for N1 and N2 without a His-tag, respectively.

The 'megaprimer' method of site-directed mutagenesis (Sarkar & Sommer, 1990) was employed to introduce point mutations into the *cdf* gene using pCDFC as template. The mutagenic primers were MTA for T(-10)A, MLP for L(-1)P and MSA for S223A (Table 1), respectively.

Expression and purification. *E. coli* BL21(DE3) cells containing recombinant plasmids were cultured in LB medium containing 30 µg kanamycin ml⁻¹ at 37 °C until the OD₆₀₀ reached 0.6. The target proteins were induced with 0.4 mM IPTG, and the cultivation was continued at 30 °C for 3 h. The cells were suspended in 50 mM Tris/HCl (pH 8.0) and disrupted by sonication. The insoluble fractions were collected by centrifugation at 13 000 g for 10 min at 4 °C and washed three times with inclusion body washing buffer [50 mM Tris/HCl, 1% (v/v) Triton X-100, pH 8.0]. The pellet was then subjected to three successive washes with 50 mM Tris/HCl (pH 8.0) to remove the contaminated Triton X-100 and yield purified inclusion bodies.

The inclusion bodies were solubilized in 50 mM Gly/NaOH (pH 11.0) and incubated at 37 °C for 4–6 h to activate the enzymes. Thereafter, the soluble fractions were dialysed against buffer A (50 mM Tris/HCl, 0.5 M NaCl, pH 7.5) containing 9 mM imidazole at 4 °C, and then loaded onto a Ni²⁺-charged Chelating Sepharose Fast Flow resin (Amersham Biosciences) column for purification. The eluted fractions were dialysed against 50 mM Gly/NaOH (pH 11.0) to yield purified enzymes.

Enzyme activity assay. The standard assay for assessing the proteolytic activity of the enzyme towards azocasein was conducted at 55 °C for 2 h in 500 µl of reaction mixture containing 4 µg enzyme ml⁻¹ and 0.5% (w/v) azocasein in 50 mM Gly/NaOH (pH 11.0). The reaction was terminated by the addition of 500 µl 40% (w/v) TCA. After standing at room temperature for 15 min, the mixture was centrifuged at 13 000 g for 10 min, and the absorbance of the supernatant was measured at 335 nm in a 1 cm cell. One unit (U) of activity was defined as the amount of enzyme required to increase the A₃₃₅ value by 0.01 unit per minute under the conditions described above.

The proteolytic activity of spores towards azocasein was carried out at 55 °C for 6 h in 500 µl of reaction mixture containing 4 mg (wet weight) spores and 0.5% (w/v) azocasein in 50 mM Tris/HCl (pH 8.0). The reaction was terminated and the A₃₃₅ value measured as described above. One unit (U) of activity was defined as the

amount of spores required to increase the A₃₃₅ value by 0.01 unit per minute under the conditions described above.

The proteolytic activity of the enzyme towards BSA and insoluble substrates (azocoll, keratin-azure and elastin-orcein) was measured as follows. The insoluble substrates were first washed three times with the same buffers as used in the reaction mixtures. The reaction mixture (600 µl) containing 10 µg enzyme ml⁻¹ and 3 mg each substrate in 50 mM Gly/NaOH (pH 11.0) or 50 mM Tris/HCl (pH 8.0) was incubated at 55 °C for 2 h or 0.5 h (azocoll). The reaction was terminated by the addition of 600 µl 40% TCA. After centrifugation, the absorbance of the supernatant was measured in a 1 cm cell at 280 nm for BSA, 540 nm for azocoll, 595 nm for keratin-azure or 570 nm for elastin-orcein. One unit (U) of activity was defined as the amount of enzyme required to increase the corresponding absorbance value by 0.01 unit per minute under the conditions described above.

Assay for hydrolysis of the oxidized insulin B chain by protease CDF. The oxidized insulin B chain (1 mg ml⁻¹) was mixed with protease CDF (4 µg ml⁻¹) in 50 mM Tris/HCl (pH 8.0). After incubation at 30 °C for 1 h and 12 h, 50 µl aliquots were removed, and the reaction was stopped by adding 50 µl 0.1% (v/v) trifluoroacetic acid. The hydrolysed products were subjected to peptide mass fingerprint analysis by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) using Applied Biosystems Voyager DE PRO.

SDS-PAGE and protein quantification. The SDS-PAGE was carried out with glycine-Tris (King & Laemmli, 1971) or Tricine-Tris buffer systems (Schagger & von Jagow, 1987). To prevent self-degradation of the protease during sample preparation (boiling) or electrophoresis, the sample was precipitated by 20% (w/v) TCA, and then washed with acetone before being subjected to SDS-PAGE. The gel was stained with Coomassie brilliant blue G-250. The amounts of target proteins were estimated by comparing band intensities with BSA standards using GeneTools gel imaging software (Syngene). The protein concentrations of the purified mature protease CDF and its intermediate were measured using the Bradford method (Bradford, 1976) with BSA as standard.

N-terminal sequence analysis. The proteins separated by SDS-PAGE were electroblotted onto a PVDF membrane. After staining with Coomassie brilliant blue R-250, the target protein bands were excised and subjected to N-terminal amino acid sequence analysis using a Procise 492 cLC peptide sequencer (Applied Biosystems).

Preparation of antiserum against protease CDF. The recombinant pro-protease CDF expressed in *E. coli* was used as antigen to raise antibodies in rabbits. After SDS-PAGE, the protein bands on the gel were visualized with 0.25 M KCl and cut out. The gel slice with pro-protease CDF was washed thoroughly with distilled water, crushed into small pieces in normal saline, and used as immunogen. Two New Zealand White rabbits were immunized subcutaneously. Each injection contained 400 µg pro-protease CDF, and the rabbits were boosted twice in the same manner every 2 weeks. Ten days after the last dose, blood was drawn from the ear veins and placed at 4 °C overnight to clot. After centrifugation at 10 000 g for 10 min, the supernatant was recovered as antiserum.

Immunoblot analysis. Proteins separated by SDS-PAGE were electrophoretically transferred onto a nitrocellulose membrane. The blotted membrane was blocked with TBS [100 mM Tris/HCl, 0.9% (w/v) NaCl, pH 7.5] containing 5% (w/v) skim milk. The membrane was then incubated with the antiserum against protease CDF diluted 1:1000 in TBS containing 0.05% (v/v) Tween-20 (TBST) at room temperature for 1 h. After washing three times with TBST, the blotted

membrane was treated with a HRP (horseradish peroxidase)-conjugated goat anti-rabbit IgG (H+L) secondary antibody (Pierce) diluted 1:10 000 in TBST for 1 h at room temperature, and then washed three times with TBST. The immunoreactive proteins were identified using a 3,3'-diaminobenzidine tetrahydrochloride substrate kit (ZSGB-BIO).

Transmission electron microscopy. Ultrathin sections of the spores were prepared according to the method of Horsburgh *et al.* (2003). Specimens were viewed at a range of magnifications on an H-7000FA (HITACHI) transmission electron microscope operated at 75 kV. At least 30 spores of each sample were examined.

Nucleotide sequence accession number. The gene sequence of protease CDF reported in this paper has been deposited in the GenBank database under the accession number EF108326.

RESULTS

Proteolytic activity associated with spores

Thermoactinomyces sp. CDF displayed an optimum growth temperature of 55 °C. Its 16S rRNA gene showed 100% sequence identity to those of *T. vulgaris* KCTC 9076 (AF138739) and *T. candidus* KCTC 9557 (AF138732). Microscopic examination revealed that both hyphae and spores were present in the lawn culture scraped from agar plates (Fig. 1a), while only hyphae were observed in the submerged culture (data not shown).

The purified spores (Fig. 1b) were washed three times with 50 mM Tris/HCl (pH 8.0), and no proteolytic activity was detected in the final supernatant. In contrast, the washed spores displayed hydrolytic activity toward azocasein with a specific activity of 23.7 ± 1.1 mU mg⁻¹, and the activity could be partially inhibited by 10 mM PMSF or 10 mM EDTA (Fig. 1c), implying the existence of spore-associated serine protease(s). In addition, the spore-associated proteolytic activity was sensitive to SDS inhibition. Attempts to purify these protease(s) from the spores were not successful. Alternatively, we tried to clone the gene of the spore-associated serine protease and characterize the recombinant enzyme.

Cloning, sequencing and homology analysis

Two CODEHOP primers were used to perform PCR with the genomic DNA of *Thermoactinomyces* sp. CDF as template. The amplified DNA fragments (~500 bp) were ligated into pUC18 and then transferred into *E. coli* HB101. Sequence analysis of the inserts of the plasmids isolated from five randomly selected transformants revealed two kinds of DNA fragments. One was identical to a partial sequence for the extracellular E79 protease gene from *Thermoactinomyces* sp. E79 (Lee *et al.*, 1996). The other was found to be a part of an unknown subtilase gene (named as protease CDF), and the entire *cdf* gene was obtained by TAIL-PCR (Supplementary Fig. S1a). It is worth mentioning that the DNA sequence flanking the *cdf* gene was also obtained by subsequent TAIL-PCR and Adaptor-PCR, and at least seven genes involved in spore formation were found either upstream or downstream of the *cdf* gene (Supplementary Fig. S1b, GenBank accession number GQ202245).

Protease CDF comprises 377 amino acid residues, and no signal peptide could be detected using SignalP 3.0 (Bendtsen *et al.*, 2004). This enzyme is translated as a proform (pro-protease CDF), consisting of a 94 aa N-terminal propeptide and a 283 aa mature domain (Fig. 2). The mature protease CDF contains a large number of charged residues, which is not common in other subtilases (Table 2). Using Savinase (PDB ID: 1NDQ) as template, we constructed a structure model of protease CDF by automated homology modelling (Supplementary Fig. S2). According to this model, the charged residues were found to be distributed mainly on the enzyme surface.

The amino acid sequence of mature protease CDF shows highest identity (44%) with that of a putative peptidase (Mt peptidase) of *Moorella thermoacetica* ATCC 39073 (Fig. 2). In an unrooted phylogenetic tree of the subtilisin family, protease CDF and Mt peptidase are clearly located on a distinct branch from the members of the six known subfamilies of subtilisin (Supplementary Fig. S3), implying that they may belong to a novel subfamily.

Fig. 1. Microscopy (a, b) and proteolytic activity assay (c) of spores from *Thermoactinomyces* sp. CDF. (a) Lawn culture scraped from an agar plate. (b) Purified spores. Bars in (a) and (b), 1 μm. (c) Spores were pretreated with 10 mM PMSF, 10 mM EDTA or 0.5% SDS in 50 mM Tris/HCl (pH 8.0) at 37 °C for 30 min, followed by proteolytic activity assay using azocasein as substrate. The residual activity was calculated with the activity of untreated spores defined as 100%. The values are expressed as mean ± SD of three independent experiments.

CDF	-----MEYTNGLLPSS--AEKWNARPPOG-----	-----IRKIFYLRSGLSLDRCLKEMRYSGVRPIRYLSHLGMIIGE	-33
Mt peptidase	---MWPVLLAYTGVATAAGAMIAWSGNKKVG---	---KRKIIILFHQEKLPSCRALLQKRGQVLKELPLVHALVAR	-30
BCSP	<u>---MKNKIIVFLSVLSFIIGGFFNTNTSSAETS---</u>	---STDYVPNQLIVKFKQNASLSNVQSFHKSIVGANVLSKDDKLGFEVQV	-32
E79	<u>---MKRFLSVATALLLVLLAVPGTMFAASPAS---</u>	---TDDYVPGELIVKFKDGISAQSTQSIHAQYGAKSIEKSKYLGEFEVVK	-31
Thermitase			
Akl	<u>---MKFKAIVSLSLAVSMFLPFLVEAASNDGVESPCTVSEINVSHEKGAIVQGEVIVQFKEQVNAEEKAKALKEVGATAVPDNDRVKSKFNV---</u>		-32
Proteinase K	<u>---MRLSVLLSLPLALGAPAVEQRSEAPLIEA---</u>	---RGEMVANKYIVKFKEGSALSALDAAMEKISG---	-32
BPN'	<u>---MRGKKVWISLLFALALIFTMAFGSTSSAQAAG---</u>	---KSNGEKKYIVGFKQTMSTMSAAKKKDVISEKGGKVQKQFKYVDAA	-31
	↓ ↓		
	-10 -9 -1 +1		
CDF	YTRPSRSHKLEGIEYVESDIRITITEPYVGTLLNTP--VKQEN--LPWGIERIKAHKAWKWSKGRG--IRVAVIDTGIADHDHPAIRDNYRG		52
Mt peptidase	--LPAKGKVIIEELG-LHPDIRLIEDDFEVHTVELPAARIRQEKQIVPWGVERIGAPRAWQVAAGEK--VKVAVLDTGLDAGHPDLAANVRG		57
BCSP	FSKGTVKEKIKSYK--NNPDVEYAEPNYVHAFWTPNDP--YFNN--QYGLQKIQAPQAWDSQRSDPGVKVAIIDTGVQGSHPDLASKVIY		54
E79	F-DGSVEKMIEKYK--NNPNVEYVEPNHYVHIMWTPND--LTSR--QWGPQKQAPQAWDVTRSSSTVIAIVDTGVQTNHPDLQKQIVQ		53
Thermitase			
Akl	LKVGNEAVVKALN--NNPLVEYAEPNYLFNAAWTPNDTYQGY--QYGPQNTYTDYAWDVTKGSSGQCVIAVIDTGVQTNHPDLQKQIVQ		55
Proteinase K	FAATLDENMVRVLR-AHPDVEYIEQDAVVTINAAQTNAAPWGLA--RISSTSPGTSPTYYIDESAGQGSQCVIADTGVQTNHPDLQKQIVQ		55
BPN'	S-ATLNEKAVKELK-KDPSVAVVEEDHVAHAYAQSV-----PYGVSQIKAPALHSQGYTGSNVKVAVIDSGIDSSHPDLK--VAG		46
CDF	GINILSPYHT-PRDYNHGHTHVAGIIAGRTA-EAGIVGVAPETHIYGVKAFNRKGSANLSDLLSAINWCENK-----MHVINMSFGM		133
Mt peptidase	TQNIKFPQWR-AGDGNHGHTHVAGIIAALNN-SFGVGVAPRAEIIYGVKIFNRQGDGYISDIVAGLDWALKNK-----MQVVNMSFGT		138
BCSP	GHDYVDNDNT-SDDGNHGHTHCAGITGALTNNVSVGIAGVAPQTSIYAVRVLNDQSGSTLDAVAQGTREAADSG-----AKVISLSLGG		136
E79	GYSFVDNDSN-PQDGNHGHTHCAGIAAAVTNNGTIAGMAPNASIMPVRVLNNSGSGTMAAVANGIAYAAQNG-----ADVISLSLGG		135
Thermitase	GWDFVDNDST-PQNGNHGHTHCAGIAAAVTNNSTGIAGTAPKASILAVRVLNNSGSGTWTAVANGITYAADQG-----AKVISLSLGG		136
Akl	GYDFVDNDYD-PMDLNNHGHTHVAGIAAAETNNATGIAGMAPNTRILAVRALDRNGSGTSLDIADAIITYAADSG-----AEVINLSLGG		137
Proteinase K	VKTYIY--YS-SDRGNHGHTHCAGTVGSR-----YGVAKTKQLFVGVKVLDDNGSQYSTIIAGMDFVADSKNNRNCPKGVVASLSLGG		135
BPN'	GASMVPSETNPFQDNNSHGHTHVAGTVAALNN-SIGVLGVAPSASLYAVKVLGADGSGQYSWIINGIEWAIANN-----MDVINMSLGG		128
CDF	DKVSESLRHAIQIAHRKGIILMVAAGNQMSSRIDF--PARYPETIAVTSISKNGELSVFSNISKGVDAIAPGDRILSAWLHNSKREMS		220
Mt peptidase	SQPSQALEEAVRKCVQAGMVLVAAAGNEGRDSDVLY--PARYPGVIAVSAVDKKNLASFSSRGTEVTVTAPGVDILSTYPGGKYRTMS		225
BCSP	PNGGTALQQAQVQYAWNKGSVIVAAAGNAGNTKANY--PAYSEVIAVASTQSDRKSSFTYSGSWVDVAAPGSNIYSTYKGSYQSL		222
E79	TSGSSALQSAVQQAQVNSGAVVVAAGNSSSTPNY--PAYSOAIAVASTDQNDKSSFTYSGSVVDVAAPGSNIYSTYKGSYQSL		221
Thermitase	TVGNSGLQQAQVNYAWNKGSVVVAAAGNAGNTAPNY--PAYSNALIAVASTDQNDKSSFTYSGSVVDVAAPGSNIYSTYKGSYQSL		222
Akl	DCHTTTLENAVNYAWNKGSVVVAAAGNNGSSSTTFE--PASEYENLVIAVAVDQYDRLASFSNYGTWVDVAPGVDIVSTITGNRYAYMS		223
Proteinase K	EYSSSVNSAAARLQSSGVMAVAAAGNNDADARNYS--PASEYVCTVGSADRYDRSSFSNYGVSVIDFVAPGVDIVSTITGNRYAYMS		221
BPN'	PSGSAALKAADVKAASGVVVVAAAGNEGTSGSSTVGYPGKPSVIAVAVDSSNQASFSVSGPELDVMAPGVSIQSTLPGNKYGAYN		218
CDF	●GTSMAPPHVTGTISLLHLNLPYLNPQVRYIILTQSAEKHDHLKEIGALNAYEVLRFYHKIYRY-----		283
Mt peptidase	●GTSMACPHAAGVAALILAQDRRLSGRQVARIICRTAIKLPDLSPREQDGLVNATALAARFIFGTGTAGEESVG		301
BCSP	GTSMATPHVAGVAVLLANQGYNSNTQ--IRQIIESTTDKISGTGTWYKNGRVNAYKAVQYAKQLQENKAS-----		289
E79	GTSMATPHVAGLAALLAQGRSNTQ--IRAAIENTADKISGTGTWYFQHGRINAYKAVNY-----		288
Thermitase	GTSMATPHVAGVAGLLASQGRSASN--IRAAIENTADKISGTGTWYKNGRVNAYKAVQY-----		279
Akl	GTSMASPHVAGLAALLASQGRNNE--IRQAIETADKISGTGTWYFQYGRINSYNAVY-----		280
Proteinase K	GTSMATPHVAGLAAYLMTLGKTTAAS-ACRYIADTANKGDLNIPFGTVNLLAYNNYQA-----		280
BPN'	GTSMASPHVAGAAALILSKHPNWTNTQVRSSLENTTTKLG-DSFYFGKGLINVAQAQ-----		275

Fig. 2. Amino acid sequence alignment [CLUSTAL_X (Chenna *et al.*, 2003)] of protease CDF with Mt peptidase from *Moorella thermoacetica* ATCC 39073 (ABC20323), serine protease from *Bacillus cereus* (BCSP, Q7MOW3), E79 protease from *Thermoactinomyces* sp. E79 (AAB36499), thermitase from *T. vulgaris* (1105242A), Ak1 protease from *Bacillus* sp. Ak1 (Q45670), proteinase K from *Tritirachium album* Limber (P06873) and subtilisin BPN' from *B. amyloliquefaciens* (P00782). The known signal peptide sequences are underlined. White and black arrows indicate the first and second processing sites of the N-terminal propeptide of protease CDF, respectively. The numbers represent the positions of the amino acid residues starting from the N terminus of mature protease CDF. Amino acid residues of (pre)propeptides are coded with negative numbers. Filled circles mark the residues of the catalytic triad. Open circles indicate the putative metal-binding site in protease CDF. The boxed residues indicate one of the known metal-binding sites in thermitase (Ca-3) (Gros *et al.*, 1991), Ak1 protease (Na-1) (Smith *et al.*, 1999), proteinase K (Ca-1) (Betzel *et al.*, 1988) and subtilisin BPN' (Ca-2) (Bott *et al.*, 1988), respectively.

Expression and maturation of protease CDF

The recombinant pro-protease CDF was expressed as inclusion bodies in *E. coli*. The proform could be easily solubilized from the inclusion bodies with alkaline buffers, such as 50 mM Gly/NaOH (pH 11.0–12.0), and most host proteins could be removed due to their insolubility under such conditions. As shown in Fig. 3(b), the solubilization of the proform in alkaline buffer was accompanied by the appearance of two protein products with molecular masses of 32 kDa and 31 kDa, respectively. As the incubation time extended, the 32 kDa protein was finally converted to the 31 kDa protein (Fig. 3b). These results indicate that pro-protease CDF is converted to the mature form via an

intermediate (32 kDa) by processing the 94 aa propeptide in a stepwise manner. By Edman degradation sequencing of the 32 kDa and 31 kDa proteins, the first and second cleavage sites were found to be located at Thr(–10)-Ile(–9) and Leu(–1)-Asn1, respectively (Fig. 2). The proform of active-site mutant S223A (pro-S223A) could not convert to the mature form (Fig. 3a), suggesting that the active site is required for the autoprocessing of pro-protease CDF. When the second cleavage site [Leu(–1)-Asn1] was deleted by substitution of Leu(–1) with proline, the mutant pro-L(–1)P converted to the intermediate and was maintained in this form at pH 11.0 (Fig. 3a). Additionally, the conversion of pro-L(–1)P to the

Table 2. Charged amino acid content in mature protease CDF and other subtilases

Subtilase	Organism	pH optimum	pI*	Number of residues							Reference
				Total	Negatively charged			Positively charged			
					D+E	D	E	K+R	K	R	
CDF	<i>Thermoactinomyces</i> sp. CDF	11.0	9.99	283	24	11	13	33	16	17	This study
Mt peptidase	<i>Moorella thermoacetica</i>	ND	9.12	301†	28	15	13	32	12	20	GenBank ABC20323
BCSP	<i>Bacillus cereus</i>	8.5	6.68	289	21	17	4	19	13	6	Moriyama <i>et al.</i> (1998)
YaB	<i>Bacillus</i> sp. YaB	10.5–12.0	10.19	268	9	6	3	13	4	9	Kaneko <i>et al.</i> (1989)
AprM	<i>Bacillus</i> sp. B18'	12.0–13.0	9.66	268	11	6	5	15	2	13	Masui <i>et al.</i> (1994)
Sendai	<i>Bacillus</i> sp. G-825-6	10.0	9.82	269	8	4	4	12	3	9	Yamagata <i>et al.</i> (1995)
E79	<i>Thermoactinomyces</i> sp. E79	11.0	5.96	278	14	13	1	10	4	6	Lee <i>et al.</i> (1996)
Thermitase	<i>Thermoactinomyces vulgaris</i>	7.5–9.5	7.63	279	15	13	2	15	10	5	Kleine (1982)
BPN'	<i>Bacillus amyloliquefaciens</i>	8.5–10.5	6.80	275	15	10	5	13	11	2	Markland & Smith (1967)
Carlsberg	<i>Bacillus licheniformis</i>	8.5–10.5	7.09	274	14	9	5	13	9	4	Jacobs <i>et al.</i> (1985)
E	<i>Bacillus subtilis</i>	8.5–10.5	6.80	275	14	9	5	12	8	4	Wong <i>et al.</i> (1984)

ND, Not determined.
*The pI values were calculated based on the amino acid sequences of the mature enzymes.
†The mature region of Mt peptidase was predicted according to the result of primary structure homology analysis (Fig. 2).

Fig. 3. SDS-PAGE analysis of the maturation of wild-type protease CDF and its mutants. (a) The inclusion bodies of wild-type (WT) and mutants [L(-1)P, T(-10)A and S223A] were solubilized in 50 mM Gly/NaOH (pH 11.0) at 20 µg ml⁻¹ and incubated at 37 °C. At the indicated time intervals, 100 µl aliquots were removed and subjected to SDS-PAGE. (b) SDS-PAGE analysis of the time-courses for maturation of wild-type (WT) and mutant L(-1)P at initial protein concentrations of 15, 75 and 150 µg ml⁻¹ at 37 °C in 50 mM Gly/NaOH (pH 11.0). At the indicated time intervals, 200, 40 and 20 µl aliquots were removed and subjected to SDS-PAGE. (c) Processing of the proform of S223A by active intermediate (I*) and mature protease (M*). The proforms of mutant S223A (150 µg ml⁻¹) in 50 mM Gly/NaOH (pH 11.0) were mixed with 15 µg ml⁻¹ purified I* [prepared from mutant L(-1)P] or M* and incubated at 37 °C. At the indicated time intervals, 100 µl aliquots were removed and subjected to Tricine-SDS-PAGE. The inclusion bodies of N-terminal propeptides N1 (residues -94 to -10) and N2 (residues -94 to -1) were solubilized in 50 mM Gly/NaOH (pH 11.0), and the soluble fractions were used as controls. Arrows indicate the positions of the proform (P), the intermediate (I), the mature enzyme (M), and the processed and truncated N-terminal propeptides (N2, N1) on the gel, respectively. Std, molecular mass standard.

intermediate form was independent of protein concentration (Fig. 3b), indicating that the auto-cleavage of peptide bond Thr(−10)–Ile(−9) occurs in an intramolecular manner.

The purified intermediate form of L(−1)P (I*) (Fig. 3c, lane I*) displayed a specific activity of 61.7 ± 2.6 U mg^{−1} or 74.1 ± 3.9 U mg^{−1} towards azocasein at pH 11.0 in the absence or presence of 0.5 M NaCl, respectively, similar to that of wild-type enzyme (Table 3). Therefore, its inability to process the first nine N-terminal residues [Ile(−9) to Pro(−1)] is due to the modification of the second cleavage site. Since no truncation of I* was observed under identical conditions where the intermediate remained active, the first nine N-terminal residues in the wild-type intermediate appear to be removed via a single peptide bond hydrolysis at the second cleavage site [Leu(−1)–Asn1], rather than a successive degradation process.

Unlike the case of pro-L(−1)P, pro-T(−10)A remained as the proform under the same conditions (Fig. 3a), suggesting that the deletion of the first cleavage site blocked the maturation process completely. Furthermore, the conversion of wild-type pro-protease CDF to the mature form was protein concentration dependent (Fig. 3b), implying that the peptide bond [Leu(−1)–Asn1] is hydrolysed intermolecularly. Supporting evidence came from the fact that pro-S223A could be processed to the mature form by active I* or mature protease CDF (M*), accompanied by the appearance of the intact N-terminal propeptide [N2, Met(−94) to Leu(−1)] and a truncated propeptide with nearly the same molecular mass as N1 [Met(−94) to Thr(−10)] (Fig. 3c). Because no accumulation of the intermediate was observed, the truncated propeptide was most likely the degraded product of N2.

Highly alkaline conditions (pH 11.0–12.0) were favourable for the maturation process (Fig. 4a). At pH values below 10.0 and in the absence of NaCl, the proform and intermediate of the wild-type or L(−1)P mutant suffered

autolysis. The presence of 0.5 M NaCl was beneficial for enzyme maturation by stabilizing the proform and intermediate at lower pH values (Fig. 4a). Additionally, pro-S223A was able to convert to the mature form at pH 8.0 in the presence of proteinase K (Fig. 4b), implying that exogenous protease can process pro-protease CDF at lower pH.

Biochemical properties of the recombinant protease CDF

The mature protease CDF with a His-tag at the C terminus was purified by affinity chromatography using a Ni²⁺-charged column (Fig. 3c, lane M*). The enzyme displayed pH and temperature optima of pH 10.5–11.0 and 50–55 °C (Fig. 5a, b), respectively, and was very stable at 37 °C (Fig. 5c, d). When the enzyme was incubated at 50 °C in the absence of NaCl, its half-lives at pH 11.0 and 8.0 were 120 min and 10 min, respectively. In the presence of 0.5 M NaCl, its half-lives at 50 °C increased markedly both at pH 11.0 (~460 min) and at pH 8.0 (~300 min) (Fig. 5c, d). It is worth mentioning that we have also constructed the recombinant protease CDF without the His-tag, which showed nearly the same properties as the enzyme with the His-tag, in terms of pH optimum, temperature optimum and stability (data not shown), implying that the His-tag does not affect enzyme properties. Unlike other subtilases that require calcium for stability, protease CDF did not show enhanced stability in the presence of 10 mM CaCl₂, and the enzyme stability was not affected by 10 mM EDTA. Among the metal ions tested, Fe²⁺ (1 mM) and Fe³⁺ (5 mM) inhibited the enzyme activity completely. Anionic detergent SDS (0.5 %) also caused a complete loss of the activity of protease CDF (data not shown).

Protease CDF was able to digest azocasein, BSA and azocoll, and could hydrolyse keratin-azure weakly, but it was inert towards elastin-orcin (Table 3). In most cases, the enzyme showed higher activities in the presence of 0.5 M NaCl than in the absence of NaCl, most likely due to the stabilizing effects of NaCl on the enzyme (Fig. 5c, d). However, the enzyme was inert towards BSA at pH 8.0 in the presence of 0.5 M NaCl, but could weakly hydrolyse this substrate in the absence of NaCl. The reason for this remains unclear. In the presence of 0.5 M NaCl, protease CDF showed slightly higher activities towards the insoluble substrates (azocoll and keratin-azure) at pH 8.0 than at pH 11.0, in contrast to its activity towards the soluble substrates (azocasein and BSA) (Table 3). This is probably because the soluble substrates are less stable under higher-pH conditions, and the unfolded polypeptides are more susceptible to proteolysis. A major cleavage site (Leu15–Tyr16) in oxidized insulin B chain was detected after a 1 h digestion by protease CDF, and two additional sites (Leu11–Val12 and Phe24–Phe25) were confirmed after a 12 h digestion (data not shown), implying that the enzyme prefers to hydrolyse peptide bonds with a hydrophobic amino acid residue at the P1 position.

Table 3. Specific activities of protease CDF towards various protein substrates

The values are expressed as the mean $\pm$ SD of three independent experiments.

Substrate	Specific activity (U mg ^{−1})			
	pH 8.0		pH 11.0	
	+0.5 M NaCl	−NaCl	+0.5 M NaCl	−NaCl
Azocasein	53.3 ± 0.6	20.1 ± 1.0	67.6 ± 2.0	57.1 ± 3.4
BSA	0	4.2 ± 1.9	22.7 ± 1.9	15.7 ± 1.1
Azocoll	48.4 ± 5.5	10.6 ± 3.8	40.8 ± 5.0	27.7 ± 0.6
Keratin-azure	3.8 ± 1.5	0	0.8 ± 0.5	1.1 ± 0.3
Elastin-orcin	0	0	0	0

Fig. 4. Effects of pH and NaCl (a), and proteinase K (b) on the processing of pro-protease CDF and its mutants. (a) Inclusion bodies of the proforms of wild-type (WT) and mutant L(-1)P were solubilized in 50 mM Gly/NaOH (pH 11.0) at a protein concentration of 300 $\mu\text{g ml}^{-1}$. The solutions were immediately diluted 10-fold in 50 mM Tris/HCl (pH 7.0–9.0) or 50 mM Gly/NaOH (pH 10.0–12.0), respectively, in the presence or absence of 0.5 M NaCl. After incubation at 37 °C for 2 h, samples were subjected to SDS-PAGE analysis. (b) The proform of mutant S223A was mixed with proteinase K (K) in 50 mM Tris/HCl (pH 8.0) at a final concentration of 30 $\mu\text{g ml}^{-1}$ or 3 $\mu\text{g ml}^{-1}$, and was incubated at 37 °C. At the indicated time intervals, aliquots were removed and subjected to SDS-PAGE analysis. Arrows indicate the positions of the proform (P), the intermediate (I) and the mature enzyme (M) on the gel, respectively. Std, molecular mass standard.

Detection of protease CDF in *Thermoactinomyces* sp. CDF

By immunoblot analysis, the proform and the mature form of protease CDF were detected in the lawn culture (Fig. 6a,

lane L), and small amounts of both forms were found in the intracellular soluble fraction (Fig. 6a, lane N). The appearance of two protein bands with molecular masses larger than that of the pro-protease CDF might result from cross-reaction between the antibodies and other proteins.

Fig. 5. Effects of pH and temperature on recombinant protease CDF activity (a, b) and stability (c, d). (a) Using azocasein as a substrate, the proteolytic activity of the purified enzyme (4 $\mu\text{g ml}^{-1}$) was measured at 55 °C in the following 50 mM buffers: Tris/HCl (pH 7.0–9.0), Gly/NaOH (pH 9.0–13.0). (b) Enzyme activity was measured in 50 mM Gly/NaOH (pH 11.0) at the temperatures indicated. Relative activity was calculated with the highest level of activity observed, at pH 11.0 or 55 °C, defined as 100%. (c, d) The enzyme (4 $\mu\text{g ml}^{-1}$) in 50 mM Gly/NaOH (pH 11.0) (c) or in 50 mM Tris/HCl (pH 8.0) (d) was incubated at 37 °C or 50 °C in the presence or absence of 0.5 M NaCl. At the indicated time intervals, aliquots were removed and subjected to activity assays. The residual activity was expressed as a percentage of the original activity. The reported data are the mean of three independent experiments; and standard deviations for the datasets were less than 10% of the means.

Fig. 6. Location of protease CDF in *Thermoactinomyces* sp. CDF, examined by immunoblot analysis (a, b) and electron micrographs (c). (a) Bacterial lawns were scraped from LB agar plates and washed with normal saline. The washed bacterial lawns (lane L) were suspended in 50 mM Tris/HCl (pH 8.0) and sonicated. After centrifugation, the supernatant was used as the intracellular soluble fraction (lane N). The purified spores (lane S) were suspended in 2 M KCl, 1 % (w/v) SDS and normal saline, respectively, and incubated at 37 °C for 1 h. After centrifugation, the supernatants (↑) and precipitates (↓) were subjected to SDS-PAGE and immunoblot analysis. The control sample (lanes C and C') was generated by solubilizing the inclusion bodies of pro-protease CDF in 50 mM Gly/NaOH (pH 11.0) containing 0.5 M NaCl and incubating at 37 °C for 30 min. The amount of the control sample loaded for SDS-PAGE analysis (lane C) was three times that for the immunoblot (lane C'). (b) The purified spores were suspended in 50 mM sodium phosphate (pH 6.0), 50 mM Tris/HCl (pH 8.0) and 50 mM Gly/NaOH (pH 10.0–12.0), respectively. After incubation at 37 °C or 50 °C for 1 h, the supernatants (↑) and precipitates (↓) were subjected to immunoblot analysis. The horizontal arrows in (a) and (b) indicate the positions of the proform (P), the intermediate (I) and the mature enzyme (M) on the gel or nitrocellulose membrane, respectively. Std, molecular mass standard. (c) Ultrathin-section electron microscopy of the spores recovered from the lawn culture (i), purified spores (ii), and the spores extracted with 2 M KCl (iii) or 50 mM Gly/NaOH (pH 11.0) (iv). Magnification, $\times 40\,000$. Bars, 100 nm. cx, cortex; ct, coat; oc, outer coat; ilc, inner lamellar coat.

Notably, the proform and the mature form of protease CDF were detected in the purified spores (Fig. 6a, lane S). However, the intermediate was hardly detected, probably due to its instability at physiological pH, which is unlikely to exceed pH 11.0, where the intermediate could be stabilized (Fig. 4). Meanwhile, we can not rule out the possibility that the intermediate had rapidly converted to the mature form with the aid of other protease(s) of the host.

The protease CDF could not be removed from the spores by washing with normal saline (Fig. 6a) or buffers with pH values below 10 (Fig. 6b). However, the enzyme could be extracted from the spores with 2 M KCl, while most other spore proteins could not (Fig. 6a). When the spores were

suspended in 1 % SDS, the spore-associated protease CDF was released into the solution and suffered autolysis. Similar to the inclusion bodies of pro-protease CDF, most of the spore-associated proforms could be solubilized from the spores at pH values above 11.0, and were able to convert to the intermediate and mature forms (Fig. 6b). It was noticed that, in the absence of NaCl, no autolytic fragments of pro-protease CDF were detected in the spore samples suspended in buffers with pH values below 10 at 37 °C or 50 °C (Fig. 6b), unlike the cases of recombinant proform, which suffered autolysis under the same conditions (Fig. 4). This result indicates that protease CDF can be stabilized by binding to the spore. Ultrathin-section electron microscopy of the spores demonstrated that the coat structure remained intact after being washed with 2 M

KCl or 50 mM Gly/NaOH (pH 11.0) (Fig. 6c), by which treatment spore-associated protease CDF could be extracted from the spores, suggesting that the enzyme is located on the spore surface.

DISCUSSION

Enzymic properties

Protease CDF has a pH optimum of pH 10.5–11.0, and is more stable at higher pH values. In common with other high-alkaline proteases, protease CDF possesses a high content of Arg residues (17 Arg), which has been proposed to be very important for alkaline stability of the enzyme (Makhatadze *et al.*, 2003; Masui *et al.*, 1994). Interestingly, protease CDF also has a high content of Lys residues (16 Lys) and negatively charged residues (11 Asp and 13 Glu). These charged residues, constituting about 20% of the total 283 residues of this enzyme, are mainly distributed on the enzyme surface. The highly charged surface probably contributes to the pH-dependence of the enzyme with respect to solubility and stability. In consideration of the fact that protease CDF is associated with the spore, the highly charged surface may play an important role in the binding of the enzyme to the spore. Supporting evidence comes from the observation that this enzyme could be extracted from spores with 2 M KCl, which is expected to weaken the electrostatic interactions between protease CDF and other spore components by increasing the buffer's ionic strength. Likewise, the previously reported 2 M KCl-extractable serine protease from the spores of *Bacillus cereus* (Boschwitz *et al.*, 1985; Moriyama *et al.*, 1998; Tesone & Torriani, 1975) also has a high content of charged residues (Table 2). Moreover, our results indicate that the spore-bound protease CDF is more stable than the soluble one, resembling the cases of immobilized enzymes that can be stabilized by carriers. Based on the evidence that the stability of soluble protease CDF increases at higher pH values or in the presence of NaCl, it is rational that the electrostatic interaction between protease CDF and the spore is involved in the stabilization of the enzyme.

Calcium binding has been shown to be essential for structural stability of members of the subtilase superfamily (Siezen & Leunissen, 1997). However, the stability of protease CDF is calcium-independent, but depends on NaCl. By sequence comparison analysis, only one metal-binding site could be predicted in protease CDF (Fig. 2), which corresponds to the weak calcium-binding site of subtilisin BPN' (Ca-2) (Bott *et al.*, 1988), Carlsberg (Ca-2) (McPhalen & James, 1988) and thermitase (Ca-3) (Gros *et al.*, 1991), as well as the sodium-binding site of Ak.1 protease (Na-1) (Smith *et al.*, 1999). In the absence of added Ca^{2+} , the weak site in subtilisin BPN', Carlsberg or thermitase is occupied by a monovalent cation, such as Na^+ or K^+ , which will be displaced by Ca^{2+} at high calcium concentrations (Bott *et al.*, 1988; Gros *et al.*, 1991; McPhalen & James, 1988). Here, we tentatively postulate

that the corresponding site in protease CDF is involved in the binding of Na^+ , and contributes to the enzyme stability.

Maturation

In subtilases, the N-terminal propeptide usually acts as an intramolecular chaperone to facilitate the correct folding of the mature domain, and as a potent inhibitor of the mature enzyme (Kojima *et al.*, 1997). The maturation process of subtilases has been extensively studied for subtilisin E and BPN', of which the peptide bond between the N-terminal propeptide and the mature domain is hydrolysed in an autocatalytic way, mediated by the active site of the mature enzyme (Gallagher *et al.*, 1995; Yabuta *et al.*, 2001). Unlike those of subtilisin E and BPN', pro-protease CDF converts to the 31 kDa mature enzyme via a 32 kDa intermediate, and its N-terminal propeptide is autoprocessed in a stepwise manner. Recently, it has been reported that the N-terminal region of the mature domain in the proform of Tk-subtilisin is disordered when the calcium ion binds to the Ca-7 site, which is required to promote the autoprocessing reaction. After autoprocessing of the N-terminal propeptide, the nine residues of the disordered new N terminus of mature Tk-subtilisin are truncated (Tanaka *et al.*, 2007). In another study, it was shown that the seventeen-residue flexible linker part between the propeptide and mature domain of kumamolisin-As is further truncated by exogenous protease after the intramolecular autocleavage of the peptide bond between the propeptide and the linker under acidic conditions (Comellas-Bigler *et al.*, 2004). In the case of protease CDF, the second processing site in the proform or the intermediate is susceptible to intermolecular proteolysis, suggesting that the region around this site is exposed and flexible, resembling the cases of Tk-subtilisin and kumamolisin-As. However, protease CDF differs in that its maturation is calcium independent, and alkaline conditions can speed up the maturation process.

Localization

The spore of *Thermoactinomyces* sp. CDF has the same ordered structure as the endospore of *Bacillus* species, in which the coat forms the outermost layer, according to the results of ultrathin-section electron microscopy (Fig. 6c). Protease CDF is associated with the spore coat, and is also present in the intracellular soluble fraction of the cell. It appears that the enzyme is assembled onto the spore surface after its synthesis in the mother cell. The spore-associated protease CDF is extractable with 2 M KCl. A KCl-extractable ectoprotease has been described for the dormant spores of *B. cereus*. The ectoprotease of *B. cereus* was proposed to be unnecessary for the germination process, based on the observation that the removal or inactivation of the ectoprotease did not affect the spore germination (Moriyama *et al.*, 1998). Our preliminary results (unpublished data) show that the spores of

Thermoactinomyces sp. CDF can germinate normally after treatment with 0.5% (w/v) SDS, which can completely inactivate soluble protease CDF. However, we can not exclude the possibility that the spore-bound protease CDF is SDS-resistant to some extent. The physiological role of this enzyme remains to be elucidated. On the other hand, considering that the airborne spores of *Thermoactinomyces* species are recognized as important respiratory hazards, the evidence that protease CDF possesses collagenolytic activity is particularly interesting, because collagens are the major component in the animal extracellular matrix and are present throughout the body, including the lung. Moreover, protease CDF was not sensitive to EDTA, whereas the spore-associated proteolytic activity could be partially inhibited by 10 mM EDTA (Fig. 1), implying that the spores of *Thermoactinomyces* sp. CDF contain other surface-associated metalloprotease(s) and/or serine protease(s). Further investigation of these proteases and their potential relevance to pathogenesis is warranted.

ACKNOWLEDGEMENTS

We thank Dr X. Gao for helpful discussion and critical reading of the manuscript. This work was supported in part by the National Grand Fundamental Research Program (973) (2004CG719606), the National Natural Science Foundation (30370018 and 30470019) and the Specialized Research Fund for the Doctoral Program of Higher Education (20070486050) of China.

REFERENCES

- Bendtsen, J. D., Nielsen, H., von Heijne, G. & Brunak, S. (2004). Improved prediction of signal peptides: SignalP 3.0. *J Mol Biol* **340**, 783–795.
- Betz, C., Pal, G. P. & Saenger, W. (1988). Three-dimensional structure of proteinase K at 0.15-nm resolution. *Eur J Biochem* **178**, 155–171.
- Boschwitz, H., Halvorson, H. O., Keynan, A. & Milner, Y. (1985). Trypsinlike enzymes from dormant and germinated spores of *Bacillus cereus* T and their possible involvement in germination. *J Bacteriol* **164**, 302–309.
- Bott, R., Ultsch, M., Kossiakoff, A., Graycar, T., Katz, B. & Power, S. (1988). The three-dimensional structure of *Bacillus amyloliquefaciens* subtilisin at 1.8 Å and an analysis of the structural consequences of peroxide inactivation. *J Biol Chem* **263**, 7895–7906.
- Bradford, M. M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem* **72**, 248–254.
- Chenna, R., Sugawara, H., Koike, T., Lopez, R., Gibson, T. J., Higgins, D. G. & Thompson, J. D. (2003). Multiple sequence alignment with the CLUSTAL series of programs. *Nucleic Acids Res* **31**, 3497–3500.
- Comellas-Bigler, M., Maskos, K., Huber, R., Oyama, H., Oda, K. & Bode, W. (2004). 1.2 Å crystal structure of the serine carboxyl proteinase pro-kumamolisin; structure of an intact pro-subtilase. *Structure* **12**, 1313–1323.
- Gallagher, T., Gilliland, G., Wang, L. & Bryan, P. (1995). The prosegment-subtilisin BPN' complex: crystal structure of a specific 'foldase'. *Structure* **3**, 907–914.
- Gazenko, S. V., Reponen, T. A., Grinshpun, S. A. & Willeke, K. (1998). Analysis of airborne actinomycete spores with fluorogenic substrates. *Appl Environ Microbiol* **64**, 4410–4415.
- Gros, P., Kalk, K. H. & Hol, W. G. (1991). Calcium binding to thermitase. Crystallographic studies of thermitase at 0, 5, and 100 mM calcium. *J Biol Chem* **266**, 2953–2961.
- Hilbert, D. W. & Piggot, P. J. (2004). Compartmentalization of gene expression during *Bacillus subtilis* spore formation. *Microbiol Mol Biol Rev* **68**, 234–262.
- Horsburgh, G. J., Atrih, A. & Foster, S. J. (2003). Characterization of LytH, a differentiation-associated peptidoglycan hydrolase of *Bacillus subtilis* involved in endospore cortex maturation. *J Bacteriol* **185**, 3813–3820.
- Jacobs, M., Eliasson, M., Uhlen, M. & Flock, J. I. (1985). Cloning, sequencing and expression of subtilisin Carlsberg from *Bacillus licheniformis*. *Nucleic Acids Res* **13**, 8913–8926.
- Kaneko, R., Koyama, N., Tsai, Y. C., Juang, R. Y., Yoda, K. & Yamasaki, M. (1989). Molecular cloning of the structural gene for alkaline elastase YaB, a new subtilisin produced by an alkalophilic *Bacillus* strain. *J Bacteriol* **171**, 5232–5236.
- King, J. & Laemmli, U. K. (1971). Polypeptides of the tail fibres of bacteriophage T4. *J Mol Biol* **62**, 465–477.
- Kleine, R. (1982). Properties of thermitase, a thermostable serine protease from *Thermoactinomyces vulgaris*. *Acta Biol Med Ger* **41**, 89–102.
- Kojima, S., Minagawa, T. & Miura, K. (1997). The propeptide of subtilisin BPN' as a temporary inhibitor and effect of an amino acid replacement on its inhibitory activity. *FEBS Lett* **411**, 128–132.
- Lacey, J. (1989). *Thermoactinomyces*. In *Bergey's Manual of Systematic Bacteriology*, vol. 4, pp. 2573–2585. Edited by S. T. Williams, M. E. Sharpe & J. G. Holt. Baltimore, MD: Lippincott, Williams & Wilkins.
- Lee, J. K., Kim, Y. O., Kim, H. K., Park, Y. S. & Oh, T. K. (1996). Purification and characterization of a thermostable alkaline protease from *Thermoactinomyces* sp. E79 and the DNA sequence of the encoding gene. *Biosci Biotechnol Biochem* **60**, 840–846.
- Liu, Y. G. & Whittier, R. F. (1995). Thermal asymmetric intercalated PCR: automatable amplification and sequencing of insert end fragments from P1 and YAC clones for chromosome walking. *Genomics* **25**, 674–681.
- Makhatadze, G. I., Loladze, V. V., Ermolenko, D. N., Chen, X. & Thomas, S. T. (2003). Contribution of surface salt bridges to protein stability: guidelines for protein engineering. *J Mol Biol* **327**, 1135–1148.
- Markland, F. S. & Smith, E. L. (1967). Subtilisin BPN'. VII. Isolation of cyanogen bromide peptides and the complete amino acid sequence. *J Biol Chem* **242**, 5198–5211.
- Masui, A., Fujiwara, N. & Imanaka, T. (1994). Stabilization and rational design of serine protease AprM under highly alkaline and high-temperature conditions. *Appl Environ Microbiol* **60**, 3579–3584.
- McPhalen, C. A. & James, M. N. (1988). Structural comparison of two serine proteinase-protein inhibitor complexes: eglin-c-subtilisin Carlsberg and CI-2-subtilisin Novo. *Biochemistry* **27**, 6582–6598.
- Moriyama, R., Sugimoto, K., Zhang, H., Inoue, T. & Makino, S. (1998). A cysteine-dependent serine protease associated with the dormant spores of *Bacillus cereus*: purification of the protein and cloning of the corresponding gene. *Biosci Biotechnol Biochem* **62**, 268–274.
- Nilsson, M. & Renberg, I. (1990). Viable endospores of *Thermoactinomyces vulgaris* in lake sediments as indicators of agricultural history. *Appl Environ Microbiol* **56**, 2025–2028.

- Orsini, M. & Romano-Spica, V. (2001). A microwave-based method for nucleic acid isolation from environmental samples. *Lett Appl Microbiol* **33**, 17–20.
- Pauwels, R., Devos, M., Callens, L. & van der Straeten, M. (1978). Respiratory hazards from proteolytic enzymes. *Lancet* **1**, 669.
- Pepys, J., Jenkins, P. A., Festenstein, G. N., Gregory, P. H., Lacey, M. E. & Skinner, F. A. (1963). Farmer's lung: thermophilic actinomycetes as a source of "farmer's lung hay" antigen. *Lancet* **2**, 607–611.
- Ramarao, N. & Lereclus, D. (2005). The InhA1 metalloprotease allows spores of the *B. cereus* group to escape macrophages. *Cell Microbiol* **7**, 1357–1364.
- Roberts, R. C., Nelles, L. P., Treuhaft, M. W. & Marx, J. J. (1983). Isolation and possible relevance of *Thermoactinomyces candidus* proteinases in farmer's lung disease. *Infect Immun* **40**, 553–562.
- Sarkar, G. & Sommer, S. S. (1990). The "megaprimer" method of site-directed mutagenesis. *Biotechniques* **8**, 404–407.
- Schagger, H. & von Jagow, G. (1987). Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa. *Anal Biochem* **166**, 368–379.
- Schalinatus, E., Behnke, U. & Ruttloff, H. (1983a). Intracellular proteases of *Thermoactinomyces vulgaris*. 1. Occurrence of the proteases in the cell and dynamics of their formation during cultivation. *Nahrung* **27**, 371–377.
- Schalinatus, E., Ruttloff, H. & Behnke, U. (1983b). Intracellular proteases of *Thermoactinomyces vulgaris*. 2. Biochemical characterization of proteases in cell extracts. *Nahrung* **27**, 379–386.
- Siezen, R. J. & Leunissen, J. A. (1997). Subtilases: the superfamily of subtilisin-like serine proteases. *Protein Sci* **6**, 501–523.
- Smith, C. A., Toogood, H. S., Baker, H. M., Daniel, R. M. & Baker, E. N. (1999). Calcium-mediated thermostability in the subtilisin superfamily: the crystal structure of *Bacillus* Ak.1 protease at 1.8 Å resolution. *J Mol Biol* **294**, 1027–1040.
- Tanaka, S.-I., Matsumura, H., Koga, Y., Takano, K. & Kanaya, S. (2007). Four new crystal structures of Tk-subtilisin in unautoprocessed, autoprocessed and mature forms: insight into structural changes during maturation. *J Mol Biol* **372**, 1055–1069.
- Teplyakov, A. V., Kuranova, I. P., Harutyunyan, E. H., Vainshtein, B. K., Frommel, C., Hohne, W. E. & Wilson, K. S. (1990). Crystal structure of thermitase at 1.4 Å resolution. *J Mol Biol* **214**, 261–279.
- Tesone, C. & Torriani, A. (1975). Protease associated with spores of *Bacillus cereus*. *J Bacteriol* **124**, 593–594.
- Tsuchiya, K., Nakamura, Y., Sakashita, H. & Kimura, T. (1992). Purification and characterization of a thermostable alkaline protease from alkalophilic *Thermoactinomyces* sp. HS682. *Biosci Biotechnol Biochem* **56**, 246–250.
- Tsuchiya, K., Ikeda, I., Tsuchiya, T. & Kimura, T. (1997). Cloning and expression of an intracellular alkaline protease gene from alkalophilic *Thermoactinomyces* sp. HS682. *Biosci Biotechnol Biochem* **61**, 298–303.
- Wong, S. L., Price, C. W., Goldfarb, D. S. & Doi, R. H. (1984). The subtilisin E gene of *Bacillus subtilis* is transcribed from a sigma 37 promoter in vivo. *Proc Natl Acad Sci U S A* **81**, 1184–1188.
- Wu, J., Bian, Y., Tang, B., Chen, X., Shen, P. & Peng, Z. (2004). Cloning and analysis of WF146 protease, a novel thermophilic subtilisin-like protease with four inserted surface loops. *FEMS Microbiol Lett* **230**, 251–258.
- Yabuta, Y., Takagi, H., Inouye, M. & Shinde, U. (2001). Folding pathway mediated by an intramolecular chaperone: propeptide release modulates activation precision of pro-subtilisin. *J Biol Chem* **276**, 44427–44434.
- Yamagata, Y., Isshiki, K. & Ichishima, E. (1995). Subtilisin Sendai from alkalophilic *Bacillus* sp.: molecular and enzymatic properties of the enzyme and molecular cloning and characterization of the gene, *aprS*. *Enzyme Microb Technol* **17**, 653–663.
- Yike, I., Rand, T. & Dearborn, D. G. (2007). The role of fungal proteinases in pathophysiology of *Stachybotrys chartarum*. *Mycopathologia* **164**, 171–181.
- Yoon, J.-H., Kim, I.-G., Shin, Y.-K. & Park, Y.-H. (2005). Proposal of the genus *Thermoactinomyces sensu stricto* and three new genera, *Laceyella*, *Thermoflavimicrobium* and *Seinonella*, on the basis of phenotypic, phylogenetic and chemotaxonomic analyses. *Int J Syst Evol Microbiol* **55**, 395–400.

Edited by: J. M. van Dijk