Biochimica et Biophysica Acta, 1163 (1993) 149–157

BBAPRO 34471

Purification, characterization and molecular cloning of an acidic amino acid-specific proteinase from *Streptomyces fradiae* ATCC 14544

Kengo Kitadokoro, Etsuo Nakamura, Mikio Tamaki, Toshihiko Horii, Hiroyuki Okamoto, Masaru Shin, Tomohiro Sato, Takashi Fujiwara, Hiroshige Tsuzuki, Nobuo Yoshida and Hiroshi Teraoka

Shionogi Research Laboratories, Shionogi and Co., Ltd., Osaka (Japan)

(Received 30 October 1992)

Key words: Protein purification; Gene cloning; Amino acid sequence; Serine proteinase; Substrate specificity; (*S. fradiae*)

We have isolated a novel acidic amino-acid-specific proteinase from *Streptomyces fradiae* ATCC 14544, using benzyloxycarbonyl-L-Phe-L-Leu-L-Glu-*p*-nitroanilide (Z-Phe-Leu-Glu-pNA) as a substrate. A proteinase, which we propose to call SFase, was purified from the culture filtrate by salting out, repeated S-Sepharose chromatography, and affinity chromatography (CH-Sepharose-Phe-Leu-D-Glu-OMe). The purified enzyme showed a single band having an apparent molecular weight of 19 000 on sodium dodecyl sulfate polyacrylamide gel electrophoresis. When synthetic peptides were used as substrates, SFase showed high specificity for Z-Phe-Leu-Glu-pNA. Comparison with nitroanilides of glutamic acid and aspartic acid as substrates revealed that the reactivity was about 10-fold higher for a glutamyl bond than an aspartyl bond. SFase selectively hydrolyzed the -Glu-Ala- bond of two glutamyl bonds in the oxidized insulin B-chain within the initial reaction time until the starting material was completely digested. Diisopropylfluorophosphate and benzyloxycarbonyl-Phe-Leu-Glu chloromethylketone completely inhibited SFase, while metalloproteinase inhibitors, such as EDTA and *o*-phenanthrolin, did not inhibit the enzyme. The findings indicate that SFase can be classified as a serine proteinase, and is highly specific for a glutamyl bond in comparison with an aspartyl bond. To elucidate the complete primary structure and precursor of SFase, its gene was cloned from genomic DNA of the producing strain, and the nucleotide sequence was determined. Consideration of the N- and C-terminal amino-acid sequences of the mature protein of SFase indicates that it consists of 187 amino acids, which follows a prepropeptide of 170 residues. In comparison with the acidic amino-acid-specific proteinase from *Streptomyces griseus* (Svendsen, I., Jensen, M.R. and Breddam, K. (1991) FEBS Lett. 292, 165–167), SFase had 82% homology in the amino acid sequence. The processing site for maturation of SFase was a unique sequence (-Glu-Val-), so that the propeptide could be released by cleavage of the peptide bond between Glu and Val.

Introduction

Glutamic-acid-specific proteinase (V8 proteinase), which is produced by a pathogenic *Staphylococcus aureus* strain V8, has been widely used as a tool for analysis of protein sequence [1]. The V8 proteinase has been subjected to nucleotide sequencing [2], and also X-ray structural study [3]. Recently, we found a novel glutamic-acid-specific proteinase from *S. aureus* ATCC 12600, which we proposed to call SPase, and reported that it has a different protein structure from V8 proteinase [4].

Two kinds of acidic-amino acid-specific proteinase from nonpathogenic microorganisms, have been reported; one occurrs in a small amount in a commercially available pronase from *Streptomyces griseus* (Kaken Seiyaku, Tokyo) [5,6], and another is present in a small amount in proteinase type XVI from *Bacillus subtilis* (Sigma) [7]. Although the main cleavage point of these proteinases is the carboxyl side of the glutamic-acid residue, acidic amino acids such as cysteic acid and aspartic acid are also cleaved to a small extent with respect to the long oligo-peptide substrate. The primary structure of these proteinases has been reported, but gene cloning of them has not yet been achieved.

Recently, many attempts have been made to pro-

Correspondence: H. Teraoka, Shionogi Research Laboratories, Shionogi and Co., Ltd., 5-12-4 Sagisu, Fukushima-ku, Osaka 553, Japan.
Abbreviations: ATCC, American Type Culture Collection; BSA, bovine serum albumin; PAGE, polyacrylamide gel electrophoresis; PCR, polymerase chain reaction; SDS, sodium dodecyl sulfate.

duce bioactive peptides by gene technology using amino-acid-specific proteinases as tools for selective cleavage of fusion proteins which are expressed in a host strain. For this purpose, we screened various amino-acid-specific proteinases from nonpathogenic microorganisms and found a glutamic-acid-specific proteinase from *Bacillus licheniformis*, which we called BLase. We have reported on its enzymatic characterization and gene cloning [8,9]. Recently, a glutamic-acid-specific proteinase from Alcalase™, which has the same protein sequence and character as BLase, was reported by the Carlsberg group [10,11].

While searching for a glutamic-acid-specific proteinase, we isolated a novel proteinase from *Streptomyces fradiae* ATCC 14544, which we called SFase. In the present paper, we report on the purification, characterization and molecular cloning of SFase, and also discuss the processing mechanism of its maturation.

Materials and Methods

Materials

Oxidized bovine insulin B chain (oxd-insulin B chain) was purchased from Sigma; S-Sepharose Fast Flow from Pharmacia LKB Biotechnology; benzyloxycarbonyl-L-Phe-L-Leu-L-Glu-*p*-nitroanilide (Z-Phe-Leu-Glu-pNA) and EndoLysC from Boehringer-Mannheim Yamanouchi; t-butoxycarbonyl-Ala-Ala-Glu-pNA (Boc-Ala-Ala-Glu-pNA) and Boc-Ala-Ala-Asp-pNA from Bachem; and benzoyl-Tyr-pNA (Bz-Tyr-pNA), Bz-Arg-pNA, Bz-Trp-pNA, Z-Ala-Ala-Leu-pNA, succinyl-Ala-Ala-Ala-pNA (Suc-Ala-Ala-Ala-pNA), tosyl-Lys-chloromethylketone (TLCK) and tosyl-Phe-chloromethylketone (TPCK) from Peptide Institute (Japan). We synthesized Z-Gly-Ser-pNA and Z-Phe-Leu-Glu-chloromethylketone (Z-Phe-Leu-Glu-CK).

The Protein Assay Kit was from Bio-Rad; protein molecular weight marker 'RAINBOW™' from Amersham; and the HPLC column 'Protein C_4' from Vydac.

Restriction endonucleases, M13 sequencing kits for dideoxy chain-termination method and the enzymes used for DNA manipulation were purchased from Takara Shuzo, the DNA amplification reagent (PCR) kit from Perkin Elmer Cetus, the nylon membrane (Zeta-probe) from Bio-Rad, and radiolabeled deoxyribonucleotide from Amersham. *Streptomyces fradiae* ATCC 14544 was used as a producer of SFase and *Escherichia coli* JM103 as the host for cloning the SFase gene. All other reagents were commercial products of the highest grade available.

Methods

Assay of enzyme activity. Enzyme activity was measured by Z-Phe-Leu-Glu-pNA as substrate, and the reaction was started by adding an appropriate amount of an enzyme solution composed of 0.2 mM Z-Phe-Leu-Glu-pNA, 2% DMF, 50 mM Tris-HCl and 2 mM $CaCl_2$ (pH 7.5) to make a final volume of 1.0 ml. After 10 min of reaction at 37°C, the absorbance of liberated *p*-nitroaniline was measured at 410 nm. When the substrate was Z-Phe-Leu-Glu-pNA, the specific activity of SFase was expressed as an amount (μmol) liberated of *p*-nitroaniline using 1.0 mg SFase in 1 min under the above conditions.

Protein concentration. The protein concentration was determined by a Protein Assay Kit using bovine serum albumin as the standard.

Purification of SFase. Streptomyces fradiae ATCC 14544 was cultivated by the methods described by Morihara et al. [12]. The culture was centrifuged at 4200 rpm for 30 min to remove the cells and the resultant supernatant was precipitated by addition of ammonium sulfate to 60% saturation. This precipitate was collected by centrifugation at 8000 rpm for 30 min, followed by dissolution in 10 mM Tris-HCl buffer containing 2 mM $CaCl_2$ (pH 7.5) and dialysis against the same buffer, with subsequent lyophylization. The resulting material was used as crude enzyme for the following preparation.

This crude enzyme (104.4 g) was dissolved in 1.2 l of 5 mM Tris-HCl buffer (pH 7.5) and dialyzed against the same buffer (20 l, 20 h). After centrifugation (8000 rpm, 20 min), the supernatant was combined with 1.0 l of S-Sepharose Fast Flow equilibrated with 5 mM Tris-HCl buffer (pH 7.5), and the mixture was slowly stirred 5 h at 4°C. After the supernatant was discarded by decantation, the S-Sepharose Fast Flow was poured into a glass column, which was subsequently washed with 3.0 l of 5 mM Tris-HCl buffer (pH 7.5), and eluted with 10 l of a linear salt gradient of 0–0.3 M NaCl in the same buffer. The fractions containing activity towards Z-Phe-Leu-Glu-pNA were collected to a volume of 2.5 l, dialyzed against 5 mM Tris-HCl buffer (pH 7.5, 20 l × 3, 20 h), and loaded onto an S-Sepharose Fast Flow column (5 × 25 cm) equilibrated with the same buffer. After the column had been washed with 2.5 l of the same buffer, elution was done with 5 l of a salt gradient of 0–0.1 M NaCl in the same buffer. The active fractions (1 l) were pooled and dialyzed against 5 mM Tris-HCl buffer (pH 7.5, 20 l × 2, 20 h), then applied to CH-Sepharose-Phe-Leu-D-Glu-OMe column (2.5 × 14 cm) equilibrated with the same buffer. The column was then washed with 500 ml of 5 mM Tris-HCl buffer (pH 7.5), and eluted with 1.0 l of a linear salt gradient of 0–0.1 M NaCl in the same buffer. The active fractions (200 ml) were pooled and dialyzed against 5 mM Tris-HCl buffer (pH 7.5, 20 l × 3, 20 h), then applied to the S-Sepharose Fast Flow column (3.2 × 40 cm) described above, equilibrated and eluted in the same way. The eluates with activity (200 ml) were dialyzed against 5 mM Tris-HCl buffer (pH 7.5,

20 l × 3, 20 h), and pooled. This final enzyme preparation showed a single protein band on sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE).

Hydrolysis of oxd-insulin B chain. To examine the cleavage points of the oxd-insulin B chain, the chain was dissolved in a buffer composed of 50 mM Tris-HCl (pH 7.5), to obtain a final concentration of 1.0 mg/ml. To this solution the enzyme was added (S/E = 139, weight by weight), and the reaction was allowed to proceed at 25°C. The reaction was terminated by addition of HCl to pH 3.0. The resultant solution was analyzed by HPLC using a Vydac Protein C_4 Column, 4.6 × 250 mm. After the column had been equilibrated with a mixture of water and acetonitrile (4:1) containing 0.1% trifluoroacetic acid, the above sample solution was loaded onto the column, which was then washed with the same mixture. Subsequently, the acetonitrile concentration was increased from 20% to 45% over 30 min. Each degradation product was collected from the effluent corresponding to each peak in the chromatogram and identified by amino-acid analysis.

SDS-PAGE. SDS-PAGE was performed on a 15% gel as described by Laemmli [13] using 'RAINBOW' (Amersham) as the molecular weight marker to estimate the molecular weight and the purity.

Amino-acid analysis. After hydrolysis in 4 M methanesulfonic acid containing 0.2% 3-(2-aminoethyl)indole at 110°C for 24 h, the amino-acid composition was determined with an amino-acid analyzer (Model 835 Hitachi).

N- and C-terminal sequence determination. Automated Edman degradation was performed with a Model 477A Applied Biosystems gas phase sequencer equipped with a 120A PTH amino-acid analyzer. After the SFase was inhibited by DFP (DIP-SFase), the N-terminal sequence was determined. The C-terminal sequence was determined by the digestion of DIP-SFase with EndoLysC proteinase (S/E = 40) at 24°C for 6 h in 25 mM Tris-HCl (pH 7.5), and the resulting digests were separated by HPLC using Protein C_4 Column (4.6 × 250 mm) under various conditions. Next, all fragments of peptide were examined using the amino-acid analyzer. We identified the peptide fragment, which had no basic amino acid on the carboxyl side, as the carboxyl-terminus peptide of SFase.

Synthesis of oligonucleotide primers. Oligonucleotide primers for PCR cloning of SFase gene were designed on the basis of the most frequent codon usage of *Streptomyces* genes [14], in which the third letter was almost restricted to G or C. Two kinds of single primers were synthesized on the basis of peptide sequences derived from the N-terminus of SFase and obtained by its digestion with EndoLysC proteinase by using Gene Assembler Plus DNA Synthesizer (Pharmacia LKB Biotechnology). Sense and antisense primers were 5′-GTCGCCGGCGGCGACGCCATCTA-3′ (23mer) and 5′-GTCCACGTTGGTGGTGGTGGTGTA-3′ (24mer), which were derived from possible DNA sequences corresponding to two parts of partial amino-acid sequences of SFase, Val-Ala-Gly-Gly-Asp-Ala-Ile-Tyr (N-terminal residues 1–8) and Tyr-Thr-Thr-Thr-Thr-Asn-Val-Asp (unknown position), respectively. The forward and the reverse primers specific for M13 sequencing were 5′-GTTTTCCCAGTCACGAC-3′ and 5′-CAGGAAACAGCTATGAC-3′, respectively.

Amplification of the DNA fragment encoding SFase. PCR was performed for 30 cycles on a DNA thermal cycler (Perkin Elmer Cetus) using a DNA amplification reagent kit. The thermal cycle of the PCR procedures included denaturation at 94°C for 30 s, annealing at 50°C for 1 min, and extension at 72°C for 1 min. The amplified DNA fragment, labeled by [α-^{32}P]dCTP with a random primer labeling system, was then used as the probe for hybridization.

Screening of the genomic and size-seleted DNA library. General techniques including transformation of *E. coli* cells, M13 DNA isolation and agarose gel electrophoresis were described by Maniatis et al. [15]. Genomic DNA of the *Streptomyces fradiae* ATCC 14544 was prepared as described by Marmur [16]. Southern hybridization between restriction enzyme-digested genomic DNA and ^{32}P-labeled DNA probe

TABLE I

Purification of SFase from Streptomyces fradiae ATCC 14544

Substrate: 0.2 mM Z-Phe-Leu-Glu-pNA in 2.0% DMF. Details are described in the text.

	Volume (ml)	Total A_{280} (nm)	Total activity (A_{410} (nm))	Specific activity (μmol/min per mg)	Activity recovery (%)
Crude enzyme	104.4 (g)	140800	76000	0.061	100.0
S-Sepharose (1st)	2300	13777	10488	0.087	14.1
S-Sepharose (2nd)	1000	4330	6510	0.171	8.6
Affinity chromatography	400	412	4216	1.163	5.5
S-Sepharose (3rd)	200	35	602	42.316	0.79

specific for SFase was carried out according to the method of Southern [17] with the following modification. The temperature for hybridization in 1 mM EDTA/0.5 M NaH_2PO_4 (pH 7.2)/7% SDS/1% BSA was 65°C, and the nylon membrane was washed with 15 mM NaCl/1.5 mM sodium citrate and 0.1% SDS at 50°C. A DNA fragment amplified by PCR was directly sequenced as described by Gibbs et al. [18]. Appropriate DNA fragments were inserted into M13mp10, and nucleotide sequencing was performed by the chain termination method described previously [19].

Nucleotide sequence accession number. The nucleotide sequence of SFase gene has been submitted to the GenBank™/EMBL/DDBJ Data Bank with the accession number D12470.

Results and Discussion

Purification of SFase

The purification procedures of SFase are summarized in Table I. The crude enzyme (104.4 g) from the culture broth of *Streptomyces fradiae* ATCC 14544 was used as starting material and then purified to homogeneity by sequential chromatographies. First, a rough separation was achieved by S-Sepharose chromatography (1st S-Sepharose) with the fraction containing both chymotrypsin- and trypsin-like proteinase activities. In this step, the yield of enzymatic activity toward Z-Phe-Leu-Glu-pNA was approx. 14%, indicating that the culture broth contains numerous proteinases [12] which non-specifically cleave the substrate and the specific activity of the culture broth is apparently increased. Next, the active fractions were pooled, and the linear gradient chromatography on S-Sepharose was performed (2nd S-Sepharose), by which SFase was eluted with around 0.05 M NaCl and high molecular weight contaminants were removed.

As shown in Table I, affinity chromatography of CH-Sepharose-Phe-Leu-D-Glu-OMe efficiently separated SFase from other proteinases, but a small amount of chymotrypsin- and trypsin-like proteinase activities was still found. This step increased the specific activity toward Z-Phe-Leu-Glu-pNA approx. 10-fold.

As the final step, linear gradient chromatography on S-Sepharose was performed again (3rd S-Sepharose). The chromatographic pattern and the results of SDS-PAGE are shown in Fig. 1. SFase showed a single band on SDS-PAGE, and the relative activity was increased 37-fold. The overall purification was approx. 700-fold and the yield was 0.79% (Table I).

Characterization of SFase

The optimum pH of SFase was measured by using 0.2 mM Z-Phe-Leu-Glu-pNA as substrate in a series of 50 mM buffers which contain 10% DMF, acetate buffer (pH 3.2–6.1), phosphate buffer (pH 5.5–8.1), Tris-HCl buffer (pH 6.1–9.8) and glycinate buffer (pH 9.1–11.1). After the solution was allowed to react at 25°C for 10 min, the absorbance was measured at 410 nm. The effect of pH on the stability of SFase was examined with the above buffers after the mixture had been left standing for 1.5 h at 37°C. The pH of the solutions was adjusted to pH 7.5, then the activity was measured. The optimum pH of SFase was around 8.2 and it was

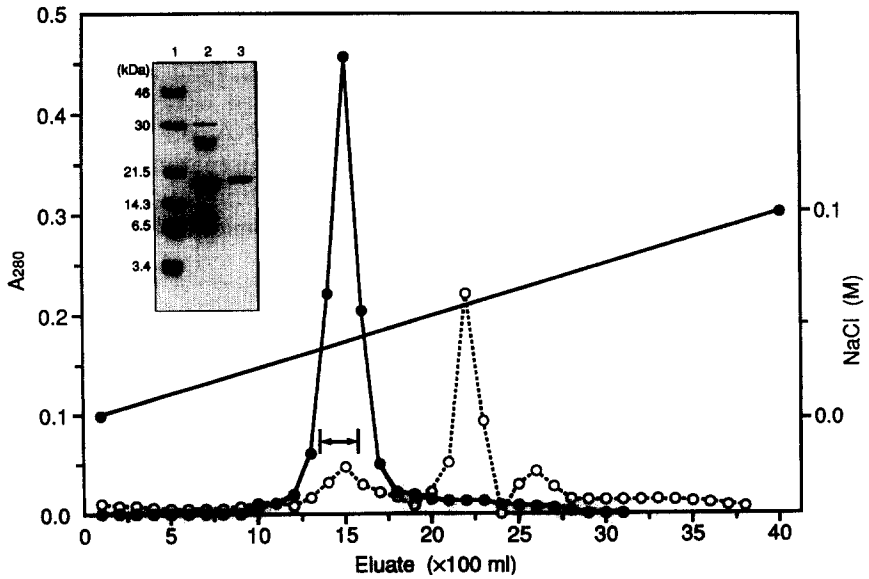

Fig. 1. S-Sepharose chromatography pattern and SDS-PAGE. The effluent profile of the SFase on the 3rd S-Sepharose Fast Flow column. Details are described in the text. The active fractions indicated by bars were collected. Absorbance at 280 nm, o — — — o; absorbance at 410 nm, which represents the activity degrading Z-Phe-Leu-Glu-pNA, •———•. The purified SFase was analyzed by SDS-PAGE: lane 1, molecular weight markers; lane 2, crude enzyme (starting material); lane 3, purified SFase.

stable at pH 4.5–9.0 (data not shown). These results indicate that SFase is typically an alkaline proteinase and shows the same property as the acidic amino-acid-specific proteinase from *S. griseus* [5].

The specificity of SFase for synthetic substrate was examined by using a series of *p*-nitroanilide substrates, in which the P1 position [20] is Glu, Asp, Leu, Trp, Tyr, Ser, Ala and Arg, respectively. As is evident from Table II, SFase showed high reactivity for Z-Phe-Leu-Glu-pNA, but only weakly cleaved the Asp substrate; i.e., the activity toward Boc-Ala-Ala-Asp-pNA is approx. 1/10 that of Boc-Ala-Ala-Glu-pNA. The other six substrates could not be cleaved (data not shown), indicating that SFase is an acidic amino-acid-specific proteinase.

The substrate specificity of SFase for a long oligo-peptide substrate was examined with oxd-insulin B chain which consists of 30 amino-acid residues. When the SFase was incubated with oxd-insulin B chain, the starting material was completely digested in 5 min, which allowed the production of two peaks on the HPLC. The amino-acid analysis of each peak resulted in identification of the first half peak as the peptide residue number 1–13, and of the latter half peak as 14–30. Although the oxd-insulin B chain contains two glutamic-acid residues, SFase cleaved the -Glu^{13}-Ala^{14}- bond, but not the -Glu^{21}-Arg^{22}- bond in the initial reaction. Furthermore, although oxd-insulin B chain contains two cysteinic acids as acidic amino acids, SFase did not cleave these residues. It was reported that in the degradation of oxd-insulin A chain, the acidic amino-acid-specific proteinase from *S. griseus* slowly hydrolyzed cysteinic acid [5]. The result on SFase seems to be inconsistent with the results on the acidic amino-acid-specific proteinase from *S. griseus*, which Breddam et al. [11] proposed to call SG-GSE and examined with synthetic hexapeptides as substrates. They proposed that SG-GSE prefer the Arg residue than the Ala residue at the P1′ position in the substrates. This may arise from the difference in the species between *S. fradiae* and *S. griseus*, or due to the difference between the long oligo-peptide substrate and the synthetic peptide substrate. Taken together, we speculate that SFase has different specificity from SG-GSE at the S1′ position. Thus, these results suggested that the enzymatic properties of SFase may be somewhat different from those of the other acidic amino acid-specific proteinases.

TABLE II

Hydrolysis of glutamyl and aspartyl p-nitroanilide substrates with SFase

Conditions: 50 mM Tris-HCl, 2 mM $CaCl_2$ (pH7.5), 25°C; Enzyme, 4.2 μg/ml; substrates, 0.2 mM in 2.0% DMF. Details are described in the text.

Substrate	Specific activity (μmol/mg per min)
Z-Phe-Leu-Glu-pNA	36.26
Boc-Ala-Ala-Glu-pNA	14.34
Boc-Ala-Ala-Asp-pNA	1.62

TABLE III

Amino-acid compositions of acidic amino-acid-specific proteinases

	SFase		SPGE [6]	BLase [8,9]	V8 protei-nase [2]	SPase [4]
	obs.	DNA				
Asp	19.4	18	15	18	65	80
Thr	21.5	21	19	28	19	19
Ser	22.9	23	25	32	10	10
Glu	5.9	5	4	11	24	23
Pro	2.9	3	4	10	25	32
Gly	33.5	31	34	29	22	22
Ala	20.5	19	22	12	17	17
Cys/2	4.4	4	4	4	0	0
Val	18.9	20	20	13	19	19
Met	1.1	1	1	2	3	3
Ile	6.1	7	7	13	15	15
Leu	5.9	5	5	5	8	8
Tyr	10.5	10	9	17	7	7
Phe	4.3	4	4	5	9	9
Lys	4.0	4	5	6	13	13
His	3.8	4	4	4	8	8
Arg	7.1	7	5	9	2	2
Trp	0.9	1	1	4	2	2
Residues		187	188	222	268	289
Mol. weight		18702	18336	23567	29006	31286

The effects of various inhibitors on the activities of SFase were examined. SFase was completely inactivated by 2.0 mM DFP, whereas metalloproteinase inhibitors, such as EDTA and *o*-phenanthrolin, did not inhibit the proteinase activity. TLCK and TPCK, which are typical inhibitors for trypsin- and chymotrypsin-like proteinase, did not inhibit the SFase activity, but Z-Phe-Leu-Glu-CK inactivated SFase. Thus, SFase was classified into the serine proteinase, with the specificity for glutamic acid being very high, but that for aspartic acid being low.

The molecular weight of proteinase was estimated to be 19000 by SDS-PAGE, as shown in Fig. 1. This value agreed well with that of 18702 calculated from the amino acid sequence deduced from DNA, which is described below. The amino-acid composition of SFase was analyzed and compared with those deduced from the nucleotide sequence (Table III). The amino-acid composition obtained and that deduced from DNA showed good agreement. Table III also shows the amino-acid composition of other acidic amino-acid-specific proteinases reported thus far. SFase is very similar to SGPE (Ref. 6, identical with recently renamed SG-GSE) in both amino-acid composition and molecular weight, but not to other proteinases. The amino-terminal sequence of DIP-SFase was determined up to the 31th residue, and the sequence Val-

```
   1  GT CGA CCG GCG CCG TGG TCG GCC AGC AGC CTT CGG CGG CGG CCG CGT CCG GCA CCA ACG ACA AGG CCG GCG CCC CGC AGA ACC    83
  84 TGA TGC GCT GGA CGC TGA CCC GCG CCA TCA AGG AGA CGC TGG TCC CGC CGA CCG GGT ACG GCT ACC CGC ACA TGC TGA CGC CCG   167
 168 TGC AGC TGA CGC CCG CCC CGG CCC GTA CGC ACG GGC CGG CGG GCC GAC GCC CCG CCC CGC CGC ACC GCG GCC GGG CTG CGG CCC   251
 252 GCG AGG CCG CCC CCT GAC GCC GCC CCC TTC CGG CCC CCA CGC CGG GAG GGG GCG GTC TCC ATG TCC GGG TCG TCG CGC CGG AGG   335
 336 CCG GGC GCG GCG GAA ACG CGC CCT TGC GGT CAG TGA AGC GTT CAT GTC AAG TTG CTC CCT CGG TAC GTC CCC ACC CGT CAC CGA   419
        SD
   1                        Met Arg Arg Thr Thr Arg Ala Arg Thr Gly Leu Ser Ala Leu Leu Leu Ala Ala Ser Leu Gly Leu Gly    23
 420 CAA GGA GAC CCC GGC ATG AGA CGC ACC ACC CGC GCG CGC ACC GGT CTG TCC GCC CTG CTC CTC GCC GCC TCC CTC GGC CTC GGC   503

  24 Ala Ala Pro Ala Gly Ala Asp Ala Pro Gln Arg Pro Ala Pro Thr Pro Ala Ser Asp Ser Ala Ala Ala Leu His Ala Leu Asp    51
 504 GCC GCC CCG GCC GGA GCC GAC GCC CCC CAG AGG CCC GCC CCC ACC CCG GCC TCC GAC TCC GCC GCC GCC CTC CAC GCC CTC GAC   587

  52 Ala Ala Val Glu Arg Thr Leu Gly Asp Asp Ser Ala Gly Thr Tyr Val Asp Ala Gly Thr Gly Glu Leu Val Val Thr Val Thr    79
 588 GCC GCC GTC GAG CGG ACC CTC GGC GAC GAC AGC GCC GGC ACC TAC GTG GAC GCC GGC ACC GGC GAA CTC GTC GTC ACC GTC ACC   671

  80 Thr Glu Ala Ala Ala Ala Lys Val Arg Ala Ala Gly Ala Thr Pro Arg Arg Val Gln Arg Gly Ala Ala Glu Leu Asp Ala Ala   107
 672 ACC GAG GCC GCC GCC GCC AAG GTC CGC GCG GCG GGC GCC ACG CCC CGG CGG GTC CAG CGC GGC GCC GCC GAG CTC GAC GCG GCC   755

 108 Met Ala Ala Leu Glu Ala Arg Ala Lys Ile Pro Gly Thr Ser Trp Gly Leu Asp Pro Arg Thr Asn Arg Ile Ala Val Glu Ala   135
 756 ATG GCC GCC CTG GAG GCA CGG GCC AAG ATC CCC GGC ACC TCG TGG GGG CTG GAC CCG CGC ACC AAC CGG ATC GCC GTG GAG GCC   839

 136 Asp Ser Ser Val Ser Ala Arg Asp Leu Ala Arg Leu Arg Lys Val Ala Ala Ser Leu Asp Gly Ala Val Ser Val Thr Arg Val   163
 840 GAC TCC TCC GTC TCC GCC CGC GAC CTG GCC CGG CTC CGC AAG GTC GCC GCC TCC CTC GAC GGC GCC GTC AGC GTC ACC CGC GTC   923
                                ▼••• ••• ••• ••• ••• ••• ••• •••
 164 Pro Gly Val Phe Gln Arg Glu Val Ala Gly Gly Asp Ala Ile Tyr Gly Gly Gly Ser Arg Cys Ser Ala Ala Phe Asn Val Thr   191
 924 CCC GGC GTG TTC CAG CGC GAG GTG GCC GGC GGC GAC GCC ATC TAC GGC GGC GGC TCG CGC TGC TCG GCG GCC TTC AAC GTC ACC  1007

 192 Lys Asn Gly Val Arg Tyr Phe Leu Thr Ala Gly His Cys Thr Asn Leu Ser Ser Thr Trp Ser Ser Thr Ser Gly Gly Thr Ser   219
1008 AAG AAC GGC GTT CGG TAC TTC CTG ACC GCC GGG CAC TGC ACC AAC CTC TCG TCC ACC TGG TCG TCC ACC TCC GGC GGC ACG TCC  1091
                                                                     ••• ••• ••• ••• ••• ••• ••• •••
 220 Ile Gly Val Arg Glu Gly Thr Ser Phe Pro Thr Asn Asp Tyr Gly Ile Val Arg Tyr Thr Thr Thr Thr Asn Val Asp Gly Arg   247
1092 ATC GGC GTC CGC GAG GGC ACC AGC TTC CCG ACC AAC GAC TAC GGC ATC GTC CGC TAC ACG ACC ACG ACC AAC GTG GAC GGC CGG  1175

 248 Val Asn Leu Tyr Asn Gly Gly Tyr Gln Asp Ile Ala Ser Ala Ala Asp Ala Val Val Gly Gln Ala Ile Lys Lys Ser Gly Ser   275
1176 GTC AAC CTG TAC AAC GGC GGC TAC CAG GAC ATC GCC TCC GCG GCC GAC GCC GTC GTG GGC CAG GCC ATC AAG AAG AGC GGC TCC  1259

 276 Thr Thr Lys Val Thr Ser Gly Thr Val Ser Ala Val Asn Val Thr Val Asn Tyr Ser Asp Gly Pro Val Tyr Gly Met Val Arg   303
1260 ACG ACC AAG GTC ACC AGC GGC ACC GTC AGC GCC GTC AAC GTC ACC GTC AAC TAC AGC GAC GGC CCC GTC TAC GGC ATG GTC CGC  1343

 304 Thr Thr Ala Cys Ser Ala Gly Gly Asp Ser Gly Gly Ala His Phe Ala Gly Ser Val Ala Leu Gly Ile His Ser Gly Ser Ser   331
1344 ACG ACC GCC TGC TCC GCC GGC GGC GAC AGC GGC GGC GCC CAC TTC GCC GGT TCC GTC GCC CTC GGC ATC CAC TCG GGC AGC TCC  1427

 332 Gly Cys Thr Gly Thr Asn Gly Ser Ala Ile His Gln Pro Val Arg Glu Ala Leu Ser Ala Tyr Gly Val Asn Val Tyr ***       358
1428 GGC TGC ACG GGC ACC AAC GGC TCC GCC ATC CAC CAG CCG GTC CGC GAG GCC CTG AGC GCC TAT GGC GTG AAC GTG TAC TGA CCC  1511

1512 GCC CCG CCC GGC CGG GCG GAG TGC CAC AGT GGG CGG TAT GAG TGA TCG TCT CCA TAC CGC CCA CAC CTA CGA GCT GGC ACC GGC  1595

1596 CGT GCT CGC CTC GAT CCG CGC CCT CCT CCA GGA CGC CTT CGA AGG CGA CCT GAG CGA GGA GGA CTG GGA GCA CGG CCT CGG CGG  1679

1680 CGT CCA CGC GTA CGC CGA GGA CCG CGG CGG GCG GAT CGT GGC GCA CGG CAG CGT CGT CAT GCG CCG CGT GGT CCA CGC CGG GCG  1763

1764 CGC CCA CCG CGT CGG GTA CGT GGA GGC CGT GGG CGT ACG GGC GGA CCG CCG CCG CCA GGG GCT GGG CAC GCG GGT GAT GGC GGC  1847

1848 GCT GGA GGA GGT CCT CGA CGG GGC GTA CGA CTT CGG CGC GCT CTC CGC CTC CGA CGC GGG CGC CCT GCT CTA GCT CGC GCG GCT  1931

1932 GGC ACC TGT GGA ACG GCC GCA TCG AGG CGT ACG GCC CGC ACG GGG TGG TCC GGA TGG CGG ACG AGG AGG ACT GCA CCT TCC TCG  2015

2016 GCC CTC GGC GGG CCG CCC GCT GCC CGC GCC GGA GGG CGG TCC CCT GCA G                                                  2064
```

Ala-Gly-Gly-Asp-Ala-Ile-Tyr-Gly-Gly-Gly-Ser-Arg-Xaa-Ser-Ala-Ala-Phe-Asn-Val-Thr-Lys-Asn-Gly-Val-Arg-Tyr-Phe-Leu-Thr-Ala- was obtained. In order to examine the C-terminal sequence, the DIP-SFase was digested by EndoLysC proteinase and each degradation product was separated by HPLC. After analysis of the amino-acid sequence of all peaks, a peptide fragment which does not have a basic amino acid was obtained; it was Glu-Ala-Leu-Ser-Ala-Tyr-Gly-Val-Asn-Val-Tyr. In addition, one of the sequence with a basic amino acid Tyr-Thr-Thr-Thr-Thr-Asn-Val-Asp-Gly-Arg was obtained.

Cloning of the SFase gene

For cloning of the SFase gene, PCR methodology [21] was adopted. Upstream and downstream primers were deduced from the two known amino-acid sequences at positions 1 to 8 of the N-terminal sequence and a peptide fragment, Tyr-Thr-Thr-Thr-Thr-Asn-Val-Asp (unknown position), which was isolated from EndoLysC proteinase digests of DIP-SFase. A DNA fragment of about 220 bp was amplified by PCR with the sense primer and antisense primer as described in Materials and Methods. The direct DNA sequence analysis of 220-bp PCR product revealed that the deduced amino-acid sequence contained all the sequences up to the 31th determined by the N-terminal amino-acid sequencing of SFase (Fig. 2), demonstrating that the amplified DNA fragment is derived from SFase gene. The 220-bp fragment thus obtained was then used as a probe for Southern hybridization of the digest of *S. fradiae* genomic DNA with appropriate restriction enzymes. When digested with *Sal*I/*Pst*I, two relatively small DNA fragments were detected; a strongly hybridizing 2-kb and a weakly hybridizing 4-kb, which suggested the occurrence of two related genes. As a more specific probe for the SFase gene, a single oligonucleotide (5′-TGCTCCACCTGGTCGTCCAC-CTCC-3′) corresponding to residues 38–45 of SFase was synthesized and used as the probe for Southern hybridization of the *Sal*I/*Pst*I-digested genomic DNA described above. By changing the probe, the 2-kb fragment alone was detected in 2nd Southern hybridization (data not shown). The 2-kb *Sal*I/*Pst*I fragment was subsequently eluted from a gel slice and ligated with *Sal*I/*Pst*I-digested M13mp10. The resulting recombinant phages were transfected into *E. coli* JM103. The transformants which did not turn blue on an X-gal plate were screened by plaque hybridization with the SFase gene-derived 220-bp probe. Several positive clones were isolated and found to carry a gene coding for SFase.

Nucleotide sequence of the SFase gene

On the basis of the results described above, we determined the nucleotide sequence of SFase gene involved in the 2-kb *Sal*I/*Pst*I fragment. The whole nucleotide sequence was determined by the chain termination method [19] in both orientations. The nucleotide sequence of SFase gene and its flanking regions together with the amino-acid sequence derived therefrom are shown in Fig. 2. Starting from the initiation codon ATG at nucleotide position 435 and terminating in the TGA codon at nucleotide position 1506, a single open reading frame (ORF) was found: it was composed of 1071 nucleotides (357 amino-acid residues), including the C-terminal amino-acid sequence (Fig. 2). The translational initiation codon (ATG) was preceded by a putative ribosome-binding site (AAGGAG). Further upstream was the promoter-like consensus sequence (−35 and −10 regions) shown by wavy lines in Fig. 2. The first approx. 30 residues from the N-terminus resemble a typical signal peptide sequence, with a short sequence containing two positively charged residues followed by a long hydrophobic sequence and a potential signal peptide cleavage site (Ala-X-Ala). Two transcription terminator structure-like inverted sequences were also found downstream from the termination codon (Fig. 2).

SFase is one of the extracellular proteinases synthesized and secreted by *S. fradiae*. Its nucleotide sequence analysis suggests that the gene product is initially synthesized as a 357-residue precursor protein, followed by processing to a mature enzyme consisting of 187 amino-acid residues. Since the N-terminal sequence consisting of around 30 residues seems to be the signal peptide, the length of the pro-sequence is as long as 140 amino-acid residues. In fact, all other known extracellular proteinases produced by *Bacillus subtilis* are synthesized as pre-pro-enzyme, in which the length of the pro-sequence goes up to approx. 200 amino-acid residues in the case of neutral proteinase

Fig. 2. Nucleotide sequence and deduced amino-acid sequence of SFase. The amino-acid sequence is shown above the nucleotide sequence. The first amino acid of translation (Met) is numbered 1 and the cleavage site giving the mature enzyme is indicated by the arrowhead at position 170. The putative ribosome-binding site (SD) is boxed and the consensus sequence (−35 and −10 regions) of the promoter is indicated by wavy lines. Two pairs of inverted repeat sequences upstream of the coding region are shown by opposing arrows. Asterisks indicate the termination codon. The probable transcription terminator in the 3′-flanking region is indicated by inverted arrows. The underlined amino acids represent the sequence determined by protein sequencing of SFase. Sense and antisense primers used for PCR cloning were synthesized based on the amino-acid sequence marked with a dotted line.

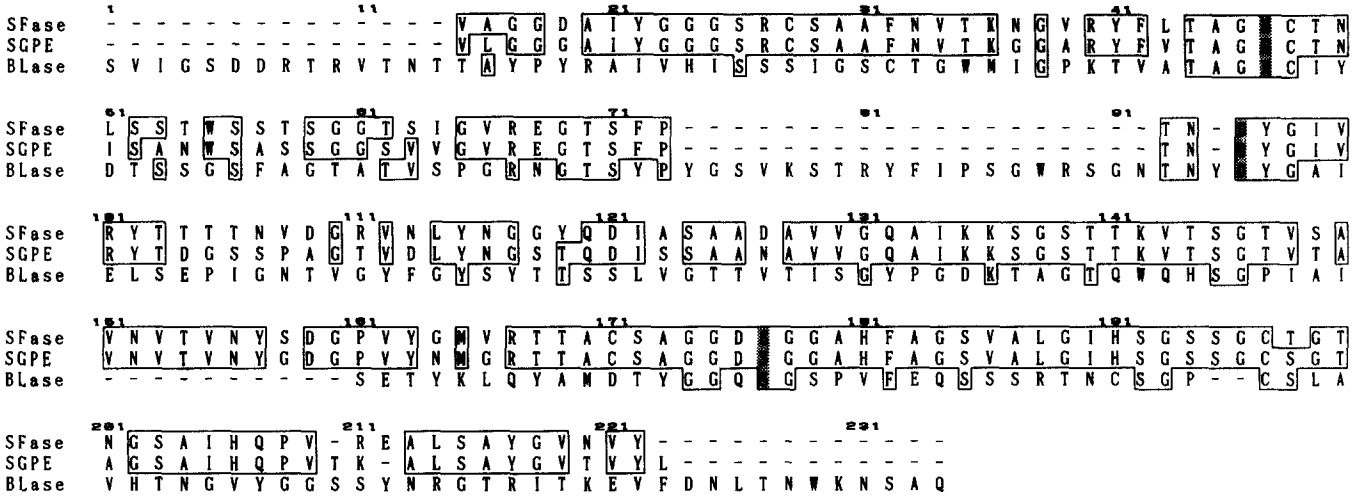

Fig. 3. Alignment of the amino-acid sequences for SFase, SGPE and BLase. Identical residues of three proteinases are boxed, and the catalytically essential triad residues His, Asp and Ser are shown by dashes.

[22]. The pro-peptide sequence is suggested to mediate the folding of the proteinase [23]. Two pairs of inverted repeat sequence located upstream from the consensus sequence for the −10 and −35 regions of SFase gene promoter have been reported to be characteristic in the *Streptomyces* promoter sequence [24], suggesting involvement of gene regulation in the some way. An interesting feature in the amino-acid sequence of SFase is the processing site for its maturation. The cleavage point is located on the C-terminal side of Glu residue in the -Glu-Val- sequence as indicated by the arrowhead in Fig. 2. Since SFase cleaves the carboxyl side of the Glu residue, it is speculated that the matured SFase is released into the medium after cleavage of the propeptide by an intramolecular autoproteolytic mechanism. Another mechanism is suggested that the participation of a metalloproteinase which can cleave the peptide bond on the N-terminal side of the Val residue, although other possibilities are not excluded. Thus, the processing mechanism of SFase remains to be elucidated using the technique of site-directed mutagenesis.

Fig. 3 shows the alignment of the primary sequence of various acidic amino-acid-specific proteinases. SFase and SGPE produced by *Streptomyces* have 82% homology in the primary sequence, whereas a low degree of homology (below 25%) was found with the glutamic acid-specific proteinase from *Bacillus licheniformis*. However, the catalytically essential triad residues His-57, Asp-102 and Ser-195 (chymotrypsinogen numbering) for serine proteinase are conserved at the positions corresponding to His-47, Asp-96 and Ser-177, and the regions of catalytic site topographies of the three proteinases show highly homologies. Also, SFase has four cysteine residues (Cys-28, Cys-48, Cys-171, Cys-197) in positions identical to those of SGPE. Since two disulfide bridges in SGPE are formed between Cys-28 and Cys-48 and between Cys-171 and Cys-197 [6], the disulfide bridges of SFase may form in a similar manner to those of SGPE.

The acidic amino acid-specific proteinases from various sources have been reported, but the three-dimensional structure of these proteinases as well as the hydrolysis mechanism of the glutamyl bond have not been elucidated. To clarify the correlation of the three-dimensional structure and function of the acidic amino-acid-specific proteinase, attempts at crystallization of SFase are in progress in our laboratory.

References

1 Drapeau, G.R., Boily, Y. and Houmard, J. (1972) J. Biol. Chem. 247, 6720–6726.
2 Carmona, C. and Gray, G.L. (1987) Nucleic Acids Res. 15, 6757.
3 Gehrig, L.M.B., Delbaere, L.T.J. and Drapeau, G.R. (1985) J. Mol. Biol. 185, 651.
4 Yoshikawa, K., Tsuzuki, H., Fujiwara, T., Nakamura, E., Iwamoto, H., Matsumoto, K., Shin, M., Yoshida, N. and Teraoka, H. (1992) Biochim. Biophys. Acta 1121, 221–228.
5 Yoshida, N., Tsuruyama, S., Nagata, K., Hirayama, K., Noda, K. and Makisumi, S. (1988) J. Biochem. 104, 451–456.
6 Svendsen, I., Jensen, M.R. and Breddam, K. (1991) FEBS Lett. 292, 165–167.
7 Niidome, T., Yoshida, N., Ogata, F., Ito, A. and Noda, K. (1990) J. Biochem. 108, 965–970.
8 Teraoka, H., Tamaki, M., Nakamura, E., Shin, M., Yoshida, N., Tsuzuki, H., Fujiwara, T. and Matsumoto, K. (1991) European Patent Application 91309737.4.
9 Kakudo, S., Kikuchi, N., Kitadokoro, K., Fujiwara, T., Nakamura, E., Okamoto, H., Shin, M., Tamaki, M., Teraoka, H., Tsuzuki, H. and Yoshida, N. (1992) J. Biol. Chem. 267, 23782–23788.
10 Svendsen, I. and Breddam, K. (1992) Eur. J. Biochem. 204, 165–171.
11 Breddam, K. and Meldal, M. (1992) Eur. J. Biochem. 206, 103–107.
12 Morihara, K. and Tsuzuki, H. (1967) Biochim. Biophys. Acta 139, 382–397.
13 Laemmli, U.K. (1970) Nature (London) 227, 680–685.
14 Wright, F. and Bibb, M.J. (1992) Gene 113, 55–65.

15 Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular Cloning, Cold Spring Harbor Laboratory, Cold Spring Harbor.
16 Marmur, J. (1961) J. Mol. Biol. 3, 208–218.
17 Southern, E.M. (1975) J. Mol. Biol. 98, 503–517.
18 Gibbs, R.A., Nguyen, P., McBride, L.J., Koepf, S.M. and Caskey, C.T. (1989) Proc. Natl. Acad. Sci. USA 86, 1919–1923.
19 Sanger, F. (1981) Science 214, 1205–1210.
20 Schechter, T. and Berger, A. (1967) Biochem. Biophys. Res. Commun. 27, 157–162.
21 Saiki, R.K., Gelfand, D.H., Stoffel, S., Schare, S.J, Higuchi, R., Horn, G.T., Mullis, K.B. and Erlich, H.A. (1988) Science 239, 487–491.
22 Yang, M.Y., Ferrari, E. and Henner, D.J. (1984) J. Bacteriol. 160, 15–21.
23 Ikemura, H., Takagi, H. and Inouye, M. (1987) J. Mol. Biol. 262, 7859–7864.
24 Daza, A., Martin, J.F., Vigal, T. and Gil, J.A. (1991) Gene 108, 63–71.