THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 269, No. 31, Issue of August 5, pp. 20167-20171, 1994
Printed in U.S.A.

Streptomyces griseus Protease C

A NOVEL ENZYME OF THE CHYMOTRYPSIN SUPERFAMILY*

(Received for publication, February 22, 1994, and in revised form, May 9, 1994)

Sachdev S. Sidhu‡, Gabriel B. Kalmar, Leslie G. Willis, and Thor J. Borgford§

From the Department of Chemistry and Institute of Molecular Biology and Biochemistry, Simon Fraser University, Burnaby, British Columbia V5A 1S6, Canada

In this report we describe a novel chymotrypsin-like serine protease produced by *Streptomyces griseus*. The enzyme has been tentatively named *S. griseus* protease C (SGPC). The gene encoding the enzyme (*sprC*) was identified and isolated on the basis of its homology to the previously characterized *S. griseus* protease B (SGPB). The *sprC* gene encodes a 457-amino acid pre-pro-mature protein of which only the 255 carboxyl-terminal amino acids are present in the mature enzyme. Mature SGPC contains two distinct domains connected by a 19-amino acid linker region rich in threonines and prolines. While the amino-terminal domain is homologous to *S. griseus* proteases A, B, and E and the α-lytic protease of *Lysobacter enzymogenes*, the carboxyl-terminal domain is not homologous with any known protease. However, the carboxyl-terminal domain shares extensive homology with chitin-binding domains of *Bacillus circulans* chitinases A1 and D, suggesting that the enzyme is specialized for the degradation of chitin-linked proteins. Recombinant expression and preliminary characterization of the catalytic properties of the enzyme are also reported. The primary specificity of SGPC is similar to that of SGPB; both enzymes preferentially cleave peptide bonds following large hydrophobic side chains.

The soil microorganism *Streptomyces griseus* secretes a variety of hydrolytic enzymes. Consequently, it is the source of a commercial, crude enzyme preparation known as Pronase. Several members of the chymotrypsin-like serine protease family have been purified from Pronase. The corresponding genes have been isolated using synthetic oligonucleotide probes based on the amino acid sequences of the enzymes. Past studies uncovered the enzymes *S. griseus* protease A (SGPA),[1] *S. griseus* protease B (SGPB) (1, 2), *S. griseus* protease E (SGPE) (3, 4), and *S. griseus* trypsin (SGT) (1, 5); however, the full breadth and divergence of enzymes secreted from *S. griseus* has not been determined. As part of an evolutionary study of the *S. griseus* proteases, we adopted a strategy for cloning protease genes whose products were either undetectable in, or absent from, the Pronase preparation. Initially, *S. griseus* genomic DNA was probed with a portion of the gene encoding the previously characterized protease SGPB. Our approach revealed six genomic fragments with significant homology to SGPB. Four of the fragments were found to contain previously known and characterized protease genes. The two remaining fragments, however, could not be correlated with any known proteases of *S. griseus*.

This report describes the cloning and sequencing of a novel serine protease gene contained in one of the two uncorrelated genomic fragments of *S. griseus*, as well as the expression and preliminary characterization of the gene product. We propose the genetic designation *sprC* for the unmapped gene and the name *S. griseus* protease C (SGPC) for the gene product.

MATERIALS AND METHODS

Enzymes and Reagents—T7 DNA polymerase was purchased from Pharmacia Biotech Inc., and calf intestinal alkaline phosphatase (CIP) was from Boehringer Mannheim. Vent DNA polymerase (New England Biolabs) was used for all polymerase chain reactions (PCR). All other enzymes for digesting or modifying DNA were purchased from New England Biolabs or Life Technologies, Inc.. Enzymes were used in accordance with the recommendations of the supplier. [α-^{32}P]dATP (~3,000 Ci/mmol) was from Amersham Corp. Ampicillin, kanamycin, activated charcoal, and *N*-succinyl-Ala-Ala-Pro-Phe-*p*-nitroanilide were from Sigma. Zeta-Probe was purchased from Bio-Rad and x-ray film from Eastman Kodak Co. All chemicals and reagents were of the highest grade commercially available. Bacterial strains and plasmids have been described previously (4), as have conditions for the growth and transformation of strains.

Expression of Recombinant SGPC—The expression of SGPC in *B. subtilis* required preparation of a special medium (YTC broth) as follows. Activated charcoal (30 g/L) was added to a solution containing 35 g/liter tryptone, 2 g/liter yeast extract, and 5 g/liter NaCl. This mixture was stirred for 15 min and then centrifuged for 20 min at 10,000 × *g*. The supernatant was decanted and filtered through Whatman No. 1 filter paper using a Buchner funnel. The charcoal-treated medium was then autoclaved, and kanamycin (20 μg/ml) and $CaCl_2$ (10 mM) were added from sterile stock solutions prior to inoculation. Cultures were grown with vigorous shaking at 30 °C.

Construction of Genomic Library—The isolation of *S. griseus* chromosomal DNA and hybridization with radiolabeled probe were described previously (4). *S. griseus* IMRU3499 chromosomal DNA was digested to completion with *Bam*HI and size fractionated on a 1.0% agarose gel. DNA fragments ranging in size from 3.6 to 5.2 kilobase pairs (kbp) were isolated and cloned into pUC18 (6) which had been linearized with *Bam*HI and treated with calf intestinal alkaline phosphatase. The *S. griseus Bam*HI fragments (0.2 μg) and linearized pUC18 (0.2 μg) were ligated in a final volume of 10 μl, and the ligation mixture was used to transform *E. coli* DH5α/P3.

Cloning of the Gene Encoding SGPC—A DNA fragment encoding amino acids 9–185 of mature SGPB (B-mat) (2) was radiolabeled and used to select genomic clones as described previously (4). A plasmid containing a 4.8-kbp insert was found to hybridize strongly with the probe; this plasmid was designated pDS-C.

DNA Sequencing—pDS-C restriction fragments were selected for sequencing on the basis of Southern blot hybridization with the B-mat probe. Sequencing strategies were as described previously (4).

* This work was supported in part by a grant from the Natural Sciences and Engineering Research Council of Canada. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EMBL Data Bank with accession number(s) L29018.

‡ Recipient of a Medical Research Council of Canada studentship.

§ Recipient of a British Columbia Health Research Foundation scholarship. To whom correspondence should be addressed. Tel./Fax: 604-291-3571; E-mail: thor_borgford@sfu.ca.

[1] The abbreviations used are: SGPA, SGPB, SGPC, and SGPE refer to *S. griseus* proteases A, B, C, and E, respectively; SGT, *S. griseus* trypsin; CIP, calf intestinal phosphatase; PCR, polymerase chain reaction; YT, yeast extract/tryptone; YTC, yeast extract/tryptone/$CaCl_2$; kbp, kilobase pair(s); PAGE, polyacrylamide gel electrophoresis.

Expression of SGPC and SGPB in B. subtilis—Both SGPC and SGPB were expressed in *B. subtilis* using a previously described secretion expression vector (pEB-11) (4). To amplify DNA fragments encoding the pro-mature portions of SGPC and SGPB, the following oligonucleotides were used as PCR primers.

CF1 (5′ oligonucleotide, *sprC*): 5′-AGTACTGACCCCGCGGCCACC-3′
CR1 (3′ oligonucleotide, *sprC*): 5′-GAATTCGGGACCGGGGCGG-3′
BF1 (5′ oligonucleotide, *sprB*): 5′-GGTACCTGCGCAACCCCCCGGACG-3′
UMR (universal M13 reverse primer): 5′-AACAGCTATGACCATG-3′

CF1 and CR1 were used as PCR primers with the plasmid pDS-C as template; the amplified product was digested with *Eco*RI and ligated into pUC18 that had been digested with *Eco*RI and *Sma*I and treated with CIP. BF1 and UMR were used as PCR primers with pNC-B (a pUC19 derivative containing *sprB*) as template; the amplified product was digested with *Pst*I and ligated into pUC18 (digested with *Pst*I and *Sma*I and treated with CIP). PCR products (0.2 μg) and vectors (0.2 μg) were ligated in final volumes of 10 μl, and the ligation mixtures were used to transform *E. coli* DH5α/P3. Vectors with inserts of the correct size were identified by restriction analysis and were partially sequenced to verify accurate amplification and ligation. A plasmid containing a correctly amplified fragment of *sprC* was designated pDS-C8, and a plasmid containing a correctly amplified fragment of *sprB* was designated pDS-B8.

pDS-C8 was digested with *Eco*RI and *Sca*I, and treated with T4 DNA polymerase to produce blunt-ended fragments. The fragment containing the pro-mature portion of *sprC* was gel-purified and ligated into pEB11 (4) (digested with *Sma*I and treated with CIP). pDS-B8 was digested with *Fsp*I and *Pst*I, and the fragment containing the pro-mature portion of *sprB* was gel-purified and ligated into pEB11 digested with *Sma*I and *Pst*I and treated with CIP. Each ligation mixture was used to transform *E. coli* DH5α/P3, and vectors containing the correct inserts in the correct orientations were isolated on the basis of restriction enzyme analysis. These vectors were designated pEB-B8 (containing *sprB* pro-mature region) and pEB-C8 (containing *sprC* pro-mature region). pEB-B8 and pEB-C8 were transformed into *B. subtilis* DB104 (7), and transformants containing the correct plasmids were identified using restriction enzyme digests. Expression and secretion of SGPB and SGPC in these transformants was verified using a skim milk clearing assay (8).

Purification of SGPC and SGPB—*B. subtilis* DB104 harboring either pEB-C8 or pEB-B8 were streaked on YT/milk plates containing 50 μg/ml kanamycin and grown overnight at 30 °C. A single colony with a well defined zone of clearing was used to inoculate 2 ml of YTC broth (described above). This 2-ml culture was grown approximately 12 h at 30 °C with vigorous shaking and then used to inoculate 200 ml of YTC broth in a 1-liter Erlenmeyer flask. After an additional 12-h growth period under the same conditions, the culture was used to inoculate 15 liters of YTC broth in a Chemap Fermentor equipped with a model FZ3000 control unit and a 20-liter G-type fermentation vessel.

The fermentor culture was maintained at 30 °C with a stir rate of 200 rpm and an aeration rate of 9 liter/min, and proteolytic activity was monitored. When proteolytic activity had plateaued (approximately 48 h), the culture was cooled to approximately 15 °C.

Bacteria were removed from the culture by ultrafiltration using a Millipore Pellicon apparatus equipped with a HVMP membrane cassette (0.45-μm cutoff). The filtrate (containing active protease) was next concentrated to 1.0 liter with a PTGC membrane cassette (10,000 nominal molecular weight limit). The retentate was centrifuged for 30 min at 10,000 × *g* to remove any additional precipitate. Sodium acetate (3.0 M, pH 4.8) was added to the concentrated retentate to a final concentration of 100 mM.

Acetone was added to the retentate with stirring to a final concentration of 30% (v/v). After stirring for 10 min, the mixture was centrifuged at 4,000 × *g* for 15 min, and the pellet was discarded. A second volume of acetone was added to the supernatant to a final concentration of 70% (v/v), and the mixture was again stirred and centrifuged as above. The pellet from this second fractionation, which contained active protease, was resuspended in 150 ml of 100 mM sodium acetate (pH 4.8). Proteolytic activity was monitored during all fractionations.

In order to remove cationic contaminants, the sample was applied to a Q-Sepharose anion exchange column (60 cm × 3 cm; Pharmacia) equilibrated with 10 mM sodium acetate, pH 4.8 (Buffer A). The column was washed with Buffer A, and the flow-through collected in 25-ml fractions. Fractions with activity toward *N*-succinyl-Ala-Ala-Pro-Phe-*p*-nitroanilide were pooled and dialyzed against 5 mM sodium acetate, 2 mM calcium chloride, pH 4.8 (Buffer B), overnight at 4 °C.

The dialyzed sample was applied to a Waters AP-1 cation exchange column using a Pharmacia FPLC system. The column was washed with Buffer B until the A_{280} base-lined. The enzyme was then eluted in a linear gradient from 0 to 0.25 M NaCl in Buffer B over 60 min. Fractions with activity toward *N*-succinyl-Ala-Ala-Pro-Phe-*p*-nitroanilide were analyzed by SDS-PAGE in 12% gels (9), and those fractions exhibiting a single 26-kDa band (preparation of SGPC) or 19 kDa band (preparation of SGPB) were pooled. Protein concentration was determined using the method of Lowry (10). The amino-terminal sequence of purified SGPC was determined using an Applied Biosystems model 473 protein sequenator at the Microsequencing Center of the University of Victoria, British Columbia, Canada.

Proteolytic Activity Assay—Enzyme activity was determined spectrophotometrically at 412 nm using the following assay mixture: 1.00 ml of 0.1 mg/ml *N*-succinyl-Ala-Ala-Pro-Phe-*p*-nitroanilide (11) in 50 mM Tris buffer, pH 8.0 + 2–20 μl of enzyme solution. Unit activity is defined as the amount of enzyme required to produce 1 mmol of *p*-nitroaniline in 1 h at 20 °C.

Homology Searches—Homologous DNA and protein sequences were searched for and identified using the BLAST network service at the National Center for Biotechnology Information (NCBI).

RESULTS

Southern Blot Analysis—Hybridization of a ^{32}P-labeled DNA fragment encoding amino acids 9–185 of mature SGPB (B-mat) with *Bam*HI-digested genomic DNA of *S. griseus* detected six unique genomic DNA fragments (Fig. 1). The two largest fragments (8.4 and 6.8 kbp) could be attributed to the genes encoding SGPB and SGPA, respectively (2), while the two smallest fragments (1.5 and 0.4 kbp) corresponded in size to fragments containing 5′ and 3′ portions of *sprE* (4). However, two strongly hybridizing fragments (4.8 and 2.3 kbp) could not be correlated with the genes of any known *S. griseus* serine proteases.

Cloning and Sequencing of sprC—The B-mat DNA probe was used to isolate plasmids containing the above mentioned 4.8-kbp DNA fragment from a DNA library prepared from *S. griseus* genomic DNA digested to completion with *Bam*HI. Approximately 20,000 *E. coli* transformants were screened by colony hybridization, and 14 strongly hybridizing clones were selected for further analysis. Southern blot analysis of plasmid DNA isolated from these clones revealed three that contained a 4.8 kbp insert which hybridized strongly with the B-mat probe. One of these plasmids (designated pDS-C) was chosen for sequencing.

Sequence analysis of pDS-C revealed a gene that we designated *sprC* (Fig. 2). This gene contains an open reading frame encoding a polypeptide of 457 amino acids. The putative GTG initiation codon is preceded by a potential ribosome binding site but, since the coding region begins just 34 nucleotides from one end of the insert, the promoter is not contained within the DNA fragment. Comparison of the predicted polypeptide with the gene products of *sprA*, *sprB* (2), *sprE* (4), and the α-lytic protease gene of *Lysobacter enzymogenes* (12) suggested a similar pre-pro-mature organization. The pre region, determined by the method of von Heijne (13) using the computer program PC/Gene (IntelliGenetics, Inc., Mountain View, California), encompasses the first 40 amino acids of the polypeptide and is typical of prokaryotic secretion signals. The pre region is followed by a propeptide region of 162 amino acids, which in turn is followed by a 189-amino acid segment highly homologous (45% identity) to mature SGPB. The open reading frame terminates in a 66-amino acid extension that is not present in

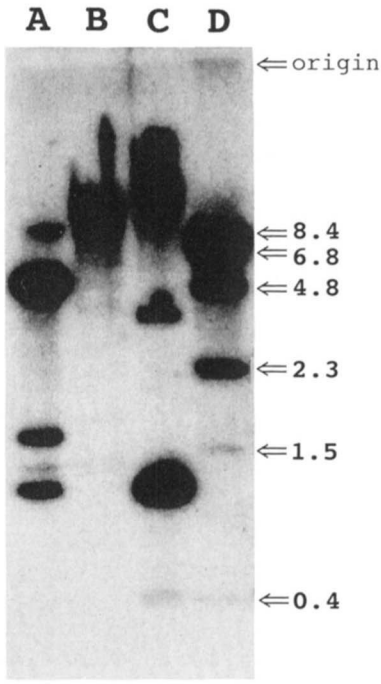

FIG. 1. **Southern blot analysis of *S. griseus* genomic DNA digested with the restriction enzymes *Pvu*II (*A*), *Pvu*I (*B*), *Eco*RI (*C*), and *Bam*HI (*D*).** The blot was hybridized with radiolabeled B-mat DNA, a probe prepared from the mature region of *sprB* as described under "Materials and Methods." Significant hybridization to six different genomic fragments was observed in *Bam*HI digests (*lane D*). The fragments at 8.4 and 6.8 kbp correspond to *sprB* and *sprA*, respectively (1), and fragments at 1.5 and 0.4 kbp correspond to 5′ and 3′ portions of *sprE* (4). The two strongly hybridizing fragments at 4.8 and 2.3 did not correlate with any previously known protease genes of *S. griseus*. The larger of the two is designated *sprC*, the smaller fragment is as yet uncharacterized.

SGPB nor in any other known protease.

A computer search of the complete non-redundant DNA/protein data base revealed that the sequence of the final 47 amino acids of the carboxyl-terminal extension is highly homologous to the carboxyl-terminal domain of chitinase A1 (57% identity) and the amino-terminal domain of chitinase D (49% identity) of *Bacillus circulans* (Fig. 3). The intervening sequence, which connects the small domain to the protease domain, is very rich in threonines and prolines; the homologous domains of chitinases A1 and D are connected to larger chitinolytic domains by similar linker regions (14).

Expression of SGPC and SGPB in B. subtilis—The polymerase chain reaction (PCR) was used to amplify a fragment of the *sprC* gene extending from codon 35 to the termination codon of pre-pro-mature SGPC. This fragment was then cloned into pEB11 (4) to produce the expression vector pEB-C8. pEB-C8 contains an open reading frame encoding a fusion protein composed of the pre region of subtilisin connected in-frame to residues 35–457 of SGPC by a tripeptide sequence (Pro-Ser-Thr). It has been reported that the deletion of as few as 6 amino acids from the amino terminus of the pro region of α-lytic protease completely abolishes the secretion of active protease in *E. coli* (15). Thus, pEB-C8 was designed to express a fragment of SGPC initiating from within the putative pre region to ensure inclusion of the complete pro-mature enzyme. Transcription and translation are initiated from the subtilisin BPN promoter and ribosome binding site, respectively, and the subtilisin pre region serves to direct the translated polypeptide for extracellular secretion (4). Similarly, the enzyme SGPB was expressed from pEB-B8, a vector designed with the subtilisin pre region fused to pro-mature SGPB. In each case, secretion of active protease was evidenced on YT/milk plates by the appearance of zones of clearing (caused by the degradation of milk proteins) around *B. subtilis* transformants harboring pEB-C8 or pEB-B8. No zones of clearing were observed around pEB11 transformants (data not shown).

Purification of SGPC and SGPB—Optimal protease expression in *B. subtilis* required a medium containing tryptone, yeast extract, and $CaCl_2$. Because yeast extract contains a pigment that interferes with subsequent purification steps, the medium was treated with activated charcoal prior to autoclaving. Pretreatment resulted in the removal of the majority of the pigment without diminishing protease expression. Any remaining traces of the pigment were effectively removed during ion exchange chromatography on Q-Sepharose, since the pigment bound strongly to the matrix while the proteases were not adsorbed.

The final purified yields of both SGPB and SGPC were approximately 5 mg/liter with specific acitivities toward *N*-succinyl-Ala-Ala-Pro-Phe-*p*-nitroanilide of 4.3 and 11.7 units/mg, respectively. Each purified enzyme showed a single band by SDS-PAGE with a mobility corresponding to a molecular mass of 26 kDa for SGPC or 19 kDa for SGPB. Amino-terminal sequence analysis of purified SGPC confirmed the location of the pro-mature junction shown in Fig. 2.

DISCUSSION

There are two branches to the family of serine proteases, the subtilisin branch and the chymotrypsin branch. Although all serine proteases employ a charge relay system involving three catalytic residues (aspartic acid, histidine, and serine), they fall into one or the other of the two branches according to their three-dimensional structures (16). The dependence of many important biological processes (including digestion, hormone activation, and blood clotting) on proteases of the chymotrypsin branch (17) underscores the need to better understand how these enzymes have become adapted to diverse activities.

Chymotrypsin-like enzymes are commonly found in higher eukaryotes, but there are few simple prokaryotic systems in which to study the evolution and divergence of this branch of the serine protease family. The soil bacterium *S. griseus* may be the most primitive organism to produce proteases of the chymotrypsin branch, and hence it is ideally suited to the study of protease evolution. Many distinct proteolytic activities have been detected in Pronase (a commercial preparation of *S. griseus* culture supernatants). Four chymotrypsin-like serine proteases have been identified, namely *S. griseus* trypsin (SGT), *S. griseus* protease A (SGPA), *S. griseus* protease B (SGPB) (1), and *S. griseus* protease E (SGPE) (3). The gene sequences of the enzymes have been published (2, 4, 5), and high resolution crystal structures have been determined (18–21).

The previous studies of *S. griseus* proteases proceeded along similar lines. Hydrolytic activities toward specific synthetic substrates were first detected in Pronase. The enzymes were then isolated on the basis of their activity toward that substrate (1, 3). Primary structure determinations provided the information necessary to design oligonuceotides that could be used as either radiolabeled probes or PCR primers in the isolation of the corresponding genes (2, 4, 5).

Hybridization Studies of S. griseus—The degree of homology between the known *S. griseus* protease genes suggested that members of the family should cross-hybridize with one another. Therefore, we employed a genetic approach designed to detect novel genes homologous to *S. griseus* protease B. A DNA fragment encoding the mature SGPB was used to generate radiolabeled probes using a random-priming method. Southern blot analysis of genomic *S. griseus* DNA detected the genes *sprA*, *sprB* (2), and *sprE* (4), and the band intensities were propor-

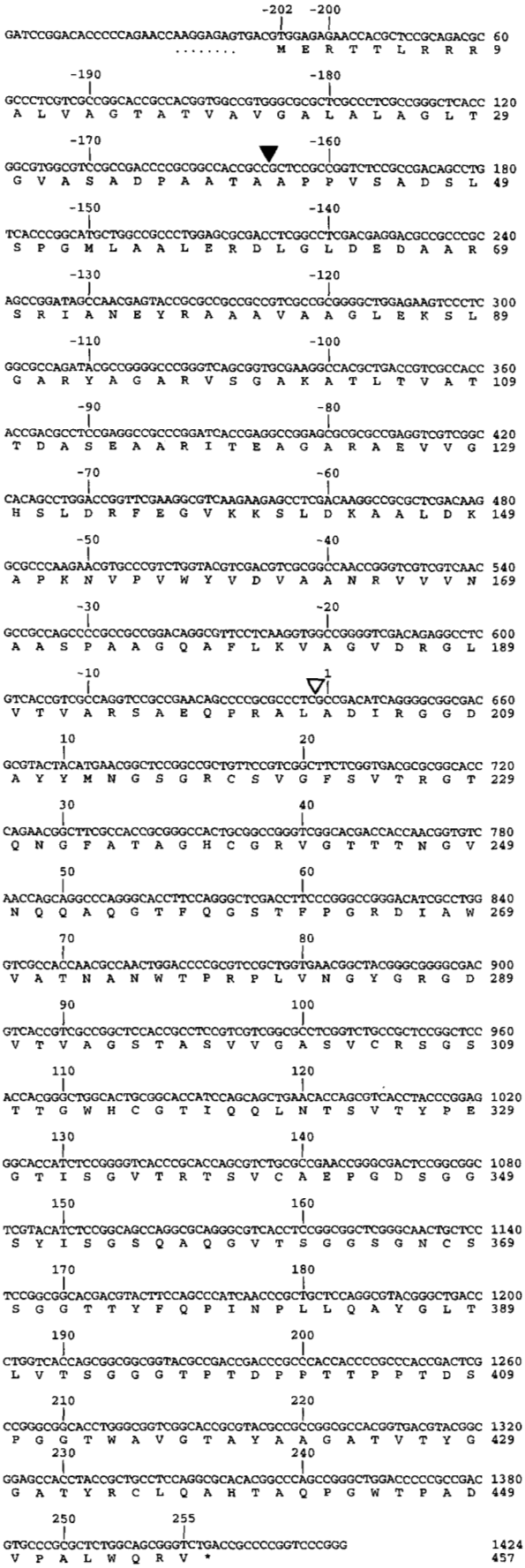

FIG. 2. **DNA sequence of *SprC*, the gene encoding the protease SGPC.** *SprC* encodes a 457-amino acid pre-pro-mature protein, of which 255 residues comprise the mature protein. The amino acid sequence deduced from the gene is shown beneath the nucleotide sequences in *one-letter code*. A termination codon is indicated by *asterisks*. Pre-pro (*solid triangles*) and pro-mature (*open triangles*) junctions are indicated. A putative ribosome binding site is indicated by *dots* below the nucleotide sequence. *Numbering* to the *right* of the figure is relative to the first nucleotide and first amino acid of each sequence. In addition, amino acid numbering with respect to the amino terminus of the mature protease is shown above the sequence. The pre region of SGPC is typical of bacterial secretion signals, suggesting that SGPC is an extracellular enzyme (13). Mature SGPC consists of an amino-terminal protease domain (amino acids 1–189) connected to a smaller carboxyl-terminal domain (amino acids 209–255) by a proline/threonine-rich linker (amino acids 190–208).

tional to homology (Fig. 1). The *S. griseus* trypsin gene was not detected, but this is not surprising since SGT's homology to SGPB is significantly less than that of the other proteases (19). Two putative genes, which could not be accounted for on the basis of extant literature, were also detected; an analysis of one of these genes (*sprC*) and its protein product (SGPC) is the subject of this report.

Organization of sprC Gene Product—The *sprC* gene contains an open reading frame encoding a polypeptide of 457 amino acids (Fig. 2). Sequence comparisons suggest this polypeptide has a pre-pro-mature organization similar to that of the α-lytic protease of *L. enzymogenes* and *S. griseus* proteases A, B, and E (2, 4, 12). The pre region extends from residue 1 to residue 40 and is typical of bacterial secretion signals. It consists of a positively charged amino terminus followed by a hydrophobic region capable of forming a membrane spanning helix, and it terminates in a signal peptidase recognition site (13). The pre region is followed by a propeptide of 162 amino acids; propeptides are found in most bacterial proteases and have been shown to be essential for correct folding in subtilisin (22) and the α-lytic protease (23). Mature SGPC is 255 amino acids in length and consists of two domains connected by a 19-amino acid linker: an amino-terminal chymotrypsin-like protease domain of 189 amino acids and a 47-amino acid carboxyl-terminal domain with significant homology to putative chitin-binding domains of chitinases A1 and D of *B. circulans* (Fig. 3).

The maturation of bacterial chymotrypsin-like proteases is believed to involve two proteolytic cleavages. The first cleavage, between the pre and pro regions, requires the action of a leader peptidase (13), while the second cleavage, at the pro-mature junction, is believed to be autocatalytic for *S. griseus* proteases A, B (2), and E (4) and the α-lytic protease (12). SGPC being homologous to these enzymes, its maturation would be expected to follow a similar mechanism. The boundary between the pro and mature regions of SGPC was originally predicted on the basis of comparisons with the pro-mature junction of the α-lytic protease and was confirmed by amino-terminal analysis of the recombinantly expressed enzyme.

Primary Specificity of SGPC—The presence of a leucine residue on the amino-terminal side of the SGPC pro-mature junction (Fig. 2) suggested that the primary specificity of the enzyme should be similar to that of SGPA and SGPB, as these enzymes contain leucine residues in the equivalent position (2). Kinetic studies of SGPC and SGPB with the chromogenic substrate *N*-succinyl-Ala-Ala-Pro-Phe-*p*-nitroanilide (11) confirmed the prediction. SGPC exhibited a specific acitivity of 11.7 units/mg, while the specific activity of SGPB was 4.3 units/mg.

Relationship to the α-Lytic Protease—Sequence comparisons suggest that, of all the *S. griseus* proteases, SGPC is the one most closely related to the α-lytic protease of *L. enzymogenes*. The pro region of SGPC (162 amino acids) is almost identical in length to the pro region of the α-lytic protease (166 amino acids), while in contrast, the pro regions of SGPA and SGPB contain only 78 and 76 residues, respectively. The pro region of SGPE (139 amino acids) lies between these two extremes. In

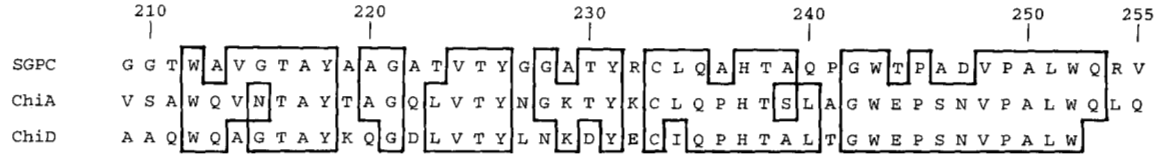

FIG. 3. **Alignment of SGPC's carboxyl-terminal domain and the homologous domains of *Bacillus circulans* chitinases (14).** *Numbering above* the sequences is relative to the amino-terminal end of mature SGPC. The sequence of the carboxyl-terminal SGPC domain appears at the *top* of the figure with sequences of the chitinase A1 (*ChiA*) and D (*ChiD*) domains aligned *below*. The chitinase A1 sequence is the putative chitin-binding domain at the carboxyl-terminal end of chitinase A1 (amino acids 653–699), while the chitinase D sequence is the putative chitin-binding domain at the amino terminal end of chitinase D (amino acids 31–77). Sequence identities are *boxed* for clarity. Notably, the three sequences are colinear; gaps were not introduced in the process of alignment.

addition, the protease domain of SGPC and the mature α-lytic protease each contain 6 cysteine residues, which occur in homologous positions. In the crystal structure of the α-lytic protease, they form three disulfide bridges. SGPA, SGPB, and SGPE each contain only four of the homologous cysteine residues. The crystal structures of SGPA (20) and SGPB (18) reveal that these residues are involved in disulfide bridges that are topologically equivalent to two of the three bridges in the α-lytic protease. Thus it seems likely that SGPC, like the α-lytic protease, contains a third disulfide bridge that is absent in *S. griseus proteases* A, B, and E. Finally, SGPC and the α-lytic protease possess an additional 2 amino acids at their mature amino-terminal ends as compared to SGPA, SGPB, and SGPE. The sequence of SGPC begins Ala-Asp and the α-lytic protease begins with the homologous sequence Ala-Asn.

Relationship to Chitinases—The carboxyl-terminal domain of SGPC is homologous to putative chitin-binding domains of chitinases A1 and D of *B. circulans* (Fig. 3) (14). Recombinant SGPC purified from *B. subtilis* has a molecular mass of 26 kDa as determined by SDS-PAGE. This is in good agreement with the mass predicted from the deduced amino acid sequence, demonstrating clearly that the carboxyl-terminal domain remains stably attached to the active protease. The small linker connecting the domain to the proteolytic domain is rich in threonines and prolines, and the presence of prolines is noteworthy as this structure may serve to resist proteolysis.

What is the relationship between SGPC and chitin? Chitin, a eukaryotic structural polysaccharide composed of repeating β(1–4)-linked *N*-acetyl-D-glucosamine units, is found complexed with proteins in insects, fungi, and nematodes. For example, chitin and protein are present in insect cuticles in equal proportions by mass (24). Hence, the efficient degradation of chitin in the environment of *S. griseus* is expected to require the participation of proteases. Notably, *S. griseus* secretes a combination of chitinolytic and proteolytic enzymes. SGPC seems to be an enzyme that possesses features of both enzymatic classes, but it is not immediately clear why a protease requires a carbohydrate-binding domain in order to degrade chitin-linked proteins. An answer to this question may be found in the observation that the chitin-binding domains of *B. circulans* chitinases are essential in the hydrolysis of insoluble crystalline chitin but are not required for the hydrolysis of soluble chitin (25). SGPA, SGPB, and SGPC have very similar substrate specificities, and we speculate that chitin-linked polypeptides may represent poor substrates for these proteases. Hence, the acquisition of a chitin-binding domain by SGPC would be an expedient method of improving specificity and activity. Perhaps, as has been observed with the cellulose-binding domains of certain cellulases (26, 27), the chitin-binding domain also participates in unraveling the polymers, thereby giving the degradative enzymes access to their substrates. We are currently examining the carbohydrate binding capabilities of SGPC and its activity in the degradation of chitin-linked proteins.

Acknowledgment—We thank J. Baardsnes for assistance in protein purification.

REFERENCES

1. Awad, W. M., Jr., Soto, A. R., Siegel, S., Skiba, W. E., Bernstrom, G. G., and Ochoa, M. S. (1972) *J. Biol. Chem.* **247,** 4144–4154
2. Henderson, G., Krygsman, P., Liu, C. J., Davey, C. C., and Malek, L. T. (1987) *J. Bacteriol.* **169,** 3778–3784
3. Yoshida, N., Tsuruyama, S., Nagata, K., Hirayama, K., Noda, K., and Makisumi, S. (1988) *J. Biochem.* (*Tokyo*) **104,** 451–456
4. Sidhu, S., Kalmar, G., and Borgford, T. (1993) *Biochem. Cell Biol.* **71,** 454–461
5. Kim, J. C., Cha, S. H., Jeong, S. T., Oh, S. K., and Byun, S. M. (1991) *Biochem. Biophys. Res. Commun.* **181,** 707–713
6. Norrander, J., Kempe, T., and Messing, J. (1983) *Gene* (*Amst.*) **26,** 101–106
7. Kawamura, F., and Doi, R. H. (1984) *J. Bacteriol.* **160,** 442–444
8. Wells, J. A., Ferrari, E., Henner, D. J., Estell, D. A., and Chen, E. Y. (1983) *Nucleic Acids Res.* **11,** 7911–7925
9. Laemmli, U. K. (1970) *Nature* **227,** 680–685
10. Lowry, O. H., Rosebrough, N. J., Farr, A. L., and Randall, R. J. (1951) *J. Biol. Chem.* **193,** 265–275
11. Nakajima, K., Powers, J. C., Ashe, B. M., and Zimmerman, M. (1979) *J. Biol. Chem.* **254,** 4027–4032
12. Silen, J. L., McGrath, C. N., Smith, K. R., and Agard, D. A. (1988) *Gene* (*Amst.*) **69,** 237–244
13. von Heijne, G. (1986) *Nucleic Acids Res.* **14,** 4683–4690
14. Watanabe, T., Oyanagi, W., Suzuki, K., Ohnishi, K., and Tanaka, H. (1992) *J. Bacteriol.* **174,** 408–414
15. Fujishige, A., Smith, K. R., Silen, J. L., and Agard, D. A. (1992) *J. Cell Biol.* **118,** 33–42
16. Fersht, A. R. (1985) *Enzyme Structure and Mechanism*, 2nd Ed., pp. 22, W. H. Freeman and Co., New York
17. Neurath, H. (1984) *Science* **224,** 350–357
18. Read, R. J., Fujinaga, M., Sielecki, A. R., and James, M. N. G. (1983) *Biochemistry* **22,** 4420–4433
19. Read, R. J., and James, M. N. G. (1988) *J. Mol. Biol.* **200,** 523–551
20. James, M. N. G., Sielecki, A. R., Brayer, G. D., Delbaere, L. T. J., and Bauer, C. A. (1980) *J. Mol. Biol.* **144,** 43–88
21. Nienaber, V. L., Breddam, K., and Birktoft, J. J. (1993) *Biochemistry* **32,** 11469–11475
22. Ohta, Y., Hojo, H., Aimoto, S., Kobayashi, T., Zhu, X., Jordan, F., and Inouye, M. (1991) *Mol. Microbiol.* **5,** 1507–1510
23. Silen, J. L., Frank, D., Fujishige, A., Bone, R., and Agard, D. A. (1989) *J. Bacteriol.* **171,** 1320–1325
24. Ruiz-Herrera, J. (1991) *Fungal Cell Wall: Structure, Synthesis and Assembly*, CRC Press, London
25. Watanabe, T., Oyanagi, W., Suzuki, K., and Tanaka, H. (1990) *J. Bacteriol.* **172,** 4017–4022
26. Tapani, R., Ruohonen, L., Nevanen, T., Laaksonen, L., Kraulis, P., Jones, T. A., Knowles, J. K. C., and Teeri, T. T. (1992) *Proteins* **14,** 475–482
27. Din, N., Gilkes, N. R., Tekant, B., Miller, R. C. J., Warren, A. J., and Kilburn, D. (1991) *Bio/Technology* **9,** 1096–1099