Characterization of the gene encoding the glutamic-acid-specific protease of *Streptomyces griseus*

SACHDEV S. SIDHU, GABRIEL B. KALMAR, AND THOR J. BORGFORD[1]

Department of Chemistry and Institute of Molecular Biology and Biochemistry, Simon Fraser University, Burnaby, BC V5A 1S6, Canada

Received September 1, 1993

SIDHU, S.S., KALMAR, G.B., and BORGFORD, T.J. 1993. Characterization of the gene encoding the glutamic-acid-specific protease of *Streptomyces griseus*. Biochem. Cell Biol. **71**: 454–461.

The complete gene sequence (*sprE*) for the glutamic-acid-specific serine protease (SGPE) of the gram-positive bacterium *Streptomyces griseus* is reported. The *sprE* gene encodes a 355 amino acid pre–pro–mature enzyme. The presence of a glutamic acid residue at the junction of the pro and mature segments of the protein suggests that the enzyme is self-processing. SGPE was found to have extensive homology with the *S. griseus* proteases A and B. However, there is an additional segment to the pro region of SGPE, lacking in proteases A and B, which has significant homology to the pro region of the α-lytic protease of the gram-negative bacterium *Lysobacter enzymogenes*. Expression of recombinant SGPE in *Bacillus subtilis* is also reported, and the enzyme is shown to be self-processing.

Key words: serine protease, maturation, expression, propeptide.

SIDHU, S.S., KALMAR, G.B., et BORGFORD, T.J. 1993. Characterization of the gene encoding the glutamic-acid-specific protease of *Streptomyces griseus*. Biochem. Cell Biol. **71** : 454–461.

La séquence complète du gène (*sprE*) de la sérine protéase spécifique de l'acide glutamique (SGPE) de la bactérie à Gram positif *Streptomyces griseus* est présentée. Le gène *sprE* code une enzyme pré–pro–mature de 355 acides aminés. La présence d'un acide glutamique à la jonction du pro-segment et du segment mature de la protéine suggère que l'enzyme s'active elle-même. La SGPE possède une très grande homologie avec les protéases A et B de *S. griseus*. Cependant, il y a un segment supplémentaire après le pro-segment de la SGPE, absent dans les protéases A et B, qui a une homologie significative avec le pro-segment de la protéase α-lytique de la bactérie à Gram négatif *Lysobacter enzymogenes*. Nous décrivons également l'expression d'une SGPE produite par génie génétique dans *Bacillus subtilis* et nous montrons que cette enzyme s'active elle-même.

Mots clés : sérine protéase, maturation, expression, propeptide.

[Traduit par la rédaction]

Introduction

Serine proteases fall into two classes or families: the subtilisin family and the chymotrypsin family. The members of the two families are very different in sequence and tertiary structure, yet they make use of a common catalytic mechanism in the hydrolysis of peptide bonds. Although it was once thought that the subtilisin family of enzymes was exclusively prokaryotic and the chymotrypsin family was exclusively eukaryotic, prokaryotic and eukaryotic members of both families have been identified. The soil microorganism *Streptomyces griseus* is known to secrete a variety of proteases (Jurasek et al. 1971). These include SGT, SGPA, and SGPB (James et al. 1978), and SGPE (Yoshida et al. 1988), which are all members of the chymotrypsin-like family of serine proteases (Fersht 1985) and structurally very similar to the mammalian members of the family (Neurath 1984). The *S. griseus* enzymes are also structurally similar to the α-lytic protease of *Lysobacter enzymogenes* (Fujinaga et al. 1985).

The *S. griseus* enzymes provide a useful system for the study of relationships between protein structure and activity. Despite their structural similarities, they have distinct substrate specificities. SGPA and SGPB have primary specificity for large aliphatic or aromatic amino acid side chains, SGT has primary specificity for positively charged side chains, and SGPE has primary specificity for glutamic acid residues. The closely related α-lytic protease has primary specificity for alanine.

The assortment of bacterial chymotrypsin-like enzymes is also potentially important in understanding protein folding and secretion processes. SGPA, SGPB, and the α-lytic enzyme are all known to be synthesized as pre–pro–mature precursors targeted for secretion. Furthermore, the pro region of the α-lytic protease is now known to be essential to the folding and maturation of that enzyme (Silen and Agard 1989), but its role has not been fully determined.

The amino acid sequence of mature SGPE and a portion of the "mature" gene sequence were reported previously (Svendsen et al. 1991). As a first step to a complete structural and kinetic characterization of SGPE, we report the cloning and sequencing of the complete *sprE* gene, as well as its recombinant expression in *Bacillus subtilis*.

Materials and methods

Enzymes and reagents

T_7 DNA polymerase was purchased from Pharmacia and CIP was from Boehringer Mannheim Biochemicals. Vent DNA polymerase (New England Biolabs) was used for all PCR. All other enzymes for digesting or modifying DNA were purchased from New England Biolabs, Inc., or Bethesda Research Laboratories, Inc. Enzymes were used in accordance with the recommendations of the supplier. [α-^{32}P]dATP (~3000 Ci/mmol; 1 Ci = 37 GBq) was from Amersham. Ampicillin and kanamycin were from Sigma. Zeta-probe was purchased from Bio-Rad and X-ray film was from Kodak. All chemicals and reagents were of the highest grade commercially available.

ABBREVIATIONS: SGPE, *Streptomyces griseus* glutamic-acid-specific serine protease; SGT, *S. griseus* trypsin; SGPA and SGPB, *S. griseus* proteases A and B, respectively; CIP, calf intestinal alkaline phosphatase; PCR, polymerase chain reaction(s); kbp, kilobase pair(s); bp, base pair(s); SFPE, *Streptomyces fradiae* glutamic-acid-specific protease; S1, subsite 1.

[1]Author to whom all correspondence should be addressed. E-mail: thor_borgford@sfu.ca.

Printed in Canada / Imprimé au Canada

FIG. 1. Oligonucleotide PCR primers. EF1 and EF2 hybridize to the noncoding strand, while ER1 and ER2 hybridize to the coding strand. The 5′ terminus of EF2 contains a *Sma*I site (CCCGGG) and the 5′ terminus of ER2 contains an *Xba*I site (AGATCT) to facilitate cloning.

Strains and plasmids

Streptomyces griseus IMRU3499 was obtained from the Waksman Institute of Microbiology, Rutgers University, Piscataway, N.J. *Escherichia coli* DH5α/P3 was a gift from Dr. R. Kay, Terry Fox Laboratories, Vancouver, and the plasmid pPT30 (Thomas et al. 1985) was a gift from Professor A.R. Fersht, Cambridge University, U.K.

Media and growth conditions

Growth of *S. griseus* mycelium for the isolation of DNA was as described previously (Hopwood et al. 1985). *Escherichia coli* was transformed as described previously (Sambrook et al. 1989) and grown on YT medium (1.0% tryptone, 0.5% yeast extract, 0.5% NaCl) (Miller 1972) containing 200 μg/mL of ampicillin. Protoplasts of *B. subtilis* were prepared by lysozyme treatment, transformed with plasmid DNA, and selected for resistance to kanamycin as described previously (Chang and Cohen 1979). *Bacillus subtilis* transformants were grown on YT medium containing 50 μg/mL of kanamycin and 1.5% w/v skim milk powder.

Isolation of DNA

Chromosomal DNA was isolated from *S. griseus* as described previously (Hopwood et al. 1985). *Escherichia coli* plasmid DNA was purified by an alkaline lysis procedure (Sambrook et al. 1989) and *B. subtilis* plasmid DNA was purified by a modified alkaline lysis procedure (Lovett and Amulos 1989). DNA fragments and vectors were gel purified for each cloning step by established methods (Sambrook et al. 1989).

Construction of genomic library

Streptomyces griseus IMRU3499 chromosomal DNA was digested to completion with *Bgl*II and size fractionated on a 1.0% agarose gel. DNA fragments ranging in size from 0.5 to 12 kbp were isolated and cloned into pUC18 (Norrander et al. 1983), which had been linearized with *Bam*HI and treated with CIP. The *S. griseus Bgl*II 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.

Hybridization

In all colony blot and Southern blot procedures, DNA was covalently cross-linked to Zeta-probe membranes using a Stratagene UV Stratalinker 2400 in accordance with the manufacturer's instructions. Filters were hybridized with labeled probe for 18 h at 42°C in the presence of 50% formamide as described previously (Vogelli and Kaytes 1987). Filters were subsequently washed at 55°C and hybridization was detected by autoradiography.

Cloning of the gene encoding SGPE

Two oligonucleotides (EF1 and ER1, Fig. 1) having sequences complementary to the previously reported partial DNA sequence of SGPE (Svendsen et al. 1991) were synthesized on an Applied Biosystems 392 DNA/RNA Synthesizer and used as PCR primers. A 234-bp DNA fragment, encoding amino acids 44–121 of mature SGPE, was amplified from *Bgl*II-digested *S. griseus* genomic DNA (Fig. 2). The PCR-amplified fragment was then used in the preparation of an ^{32}P-labeled hybridization probe (E-mat) according to protocols provided in New England Biolabs' Random Priming System I. A previously described colony hybridization method (Vogelli and Kaytes 1987) was used to select clones from the *S. griseus* genomic library for further analysis.

Plasmid DNA was purified from individual clones and digested with appropriate restriction endonucleases. The digested plasmid DNA was electrophoresed on a 1.0% agarose gel and centrifugally transferred (Wilkins and Snell 1987) to Zeta-probe membranes at 1000 rpm in an International Equipment Co. HN-S centrifuge equipped with a microtiter plate rotor. Filters were hybridized with the E-mat probe as described above. Four plasmids, each containing a 4.7-kbp insert, were found to hybridize strongly with the probe. One of these plasmids, designated pDS-E, was chosen for further characterization.

DNA sequencing

pDS-E restriction fragments were selected for sequencing by Southern blot hybridization with the E-mat probe. Various sequencing subclones were prepared by cloning the appropriate restriction fragments into pUC18. Subclones were sequenced by a combination of manual and automated methods. Manual sequencing involved the use of [α-^{32}P]dATP in accordance with protocols provided in Pharmacia's ^{T7}Sequencing Kit. Severe compressions in the DNA sequence caused by regions of high secondary structure were alleviated by the addition of dimethyl sulfoxide to the sequencing reaction and by the substitution of 7-deaza-dGTP and 7-deaza-dATP for dGTP and dATP. Automated sequencing involved the use of an Applied Biosystems 370A DNA Sequencer with reagents and protocols supplied by the manufacturer.

Construction of pEB11

A 3.4-kbp fragment of pUB110 (digested with *Eco*RI and *Pvu*II and blunt ended with T_4 DNA polymerase) and a 2.3-kbp fragment of pUC18 (digested with *Pvu*II and treated with CIP) were gel purified and ligated (0.2 μg of each fragment) in a final volume of 10 μL. The ligation mixture was used to transform *E. coli* DH5α/P3. A recombinant plasmid (pEB1) was selected by restriction enzyme analysis such that both the ampicillin gene and the kanamycin gene had the same polarity.

Based on the sequence of the *Bacillus amyloliquefaciens* subtilisin gene (subtilisin BPN) (Wells et al. 1983), two oligonucleotides (SF1 and SR1) were synthesized and used as PCR primers with *Eco*RI-digested pPT30 (Thomas et al. 1985) (a plasmid carrying the subtilisin BPN gene) as template. A 330-bp PCR-amplified

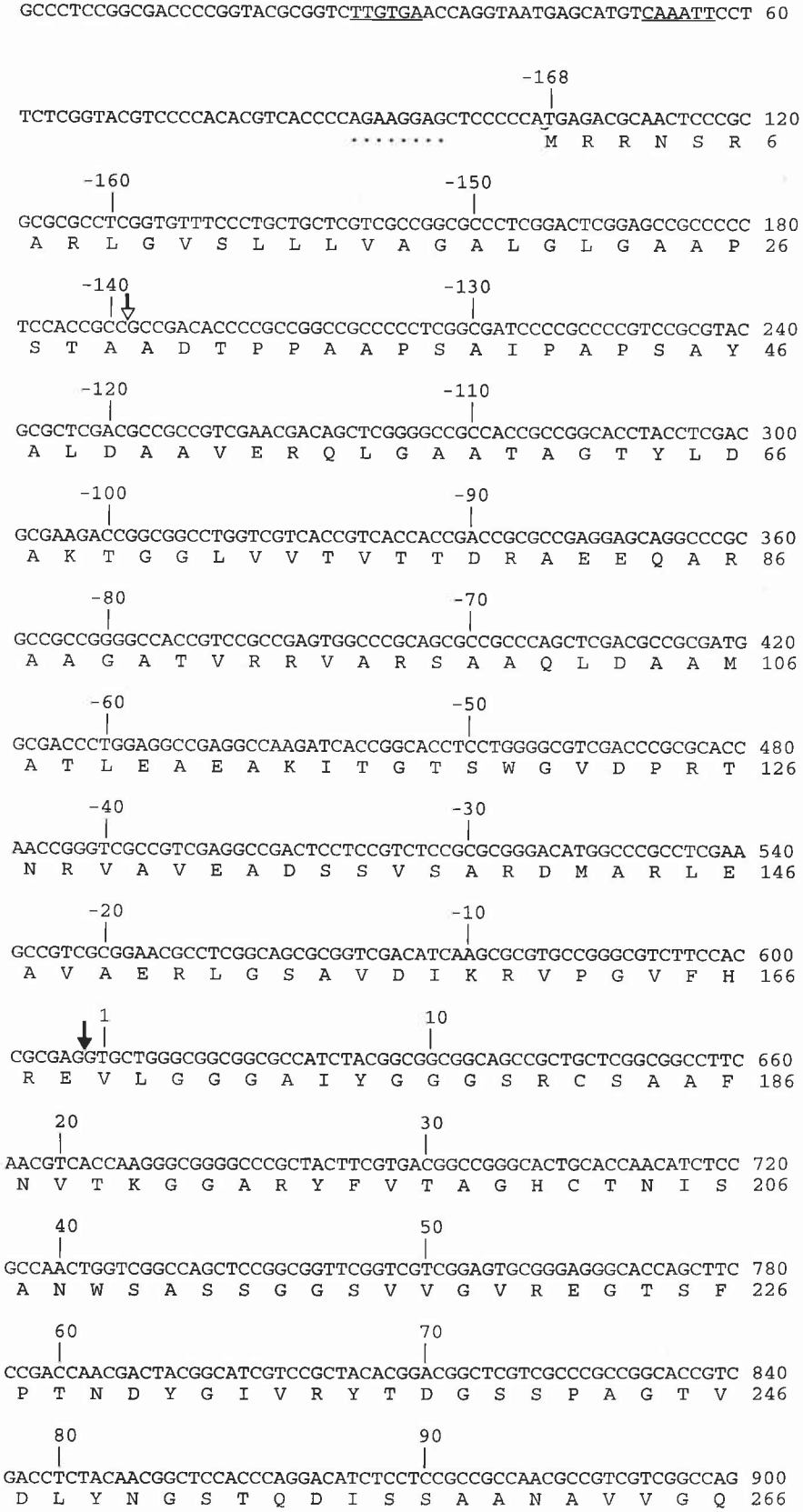

FIG. 2. DNA sequence of *sprE*. The deduced amino acid sequence is shown below the DNA sequence, with the stop codon indicated by an asterisk. The pre–pro junction (hollow arrow) and pro–mature junction (solid arrow) are shown. The putative ribosome-binding site is indicated by dots, putative promoter sites are underlined, and a 3′ inverted repeat sequence is double underlined. Numbering to the right of the figure is relative to the first nucleotide and 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.

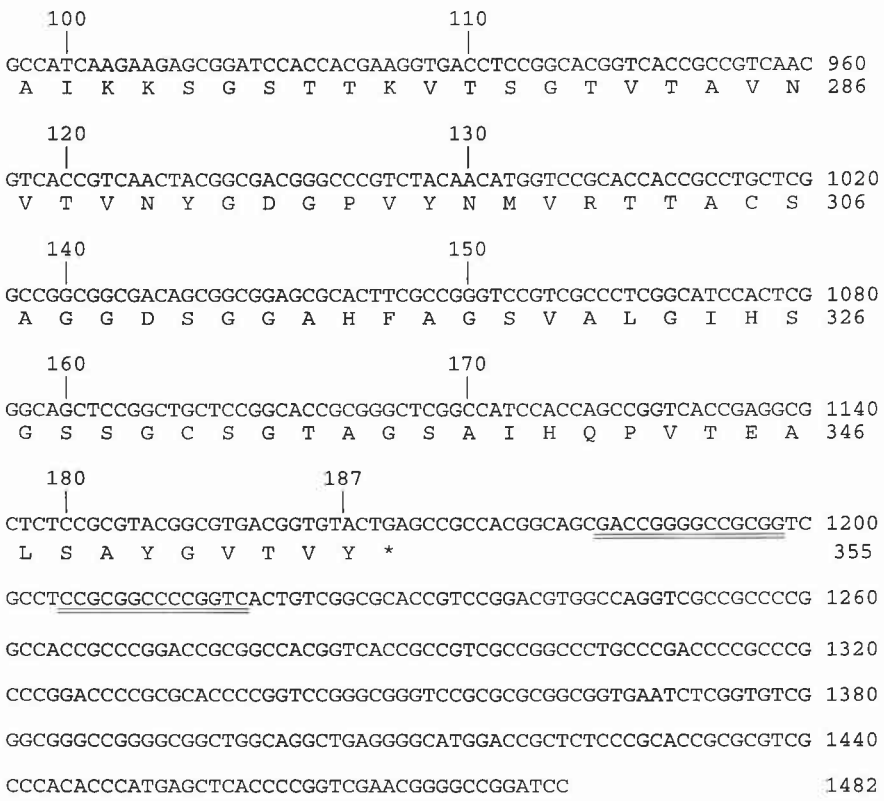

FIG. 2 (*concluded*)

fragment was isolated encompassing a region of the subtilisin gene extending from 210 bp upstream of the initiation codon to the last codon of the pre region. After an initial cloning step involving pUC18, the PCR-generated fragment was blunt-end ligated into pEB1, which was *Pvu*II digested and treated with CIP. The ligation mixture was used to transform DH5α/P3. A clone (pEB11) in which the subtilisin pre region has the same polarity as antibiotic resistance genes was selected by restriction analysis (Fig. 3).

Expression of SGPE in B. subtilis

Two oligonucleotides were synthesized (EF2 and ER2, Fig. 1) based on the sequence of *sprE* and used as PCR primers with *Eco*RI-digested pDS-E as template. The amplified product was digested with *Xba*II and ligated into pUC18, which had been digested with *Xba*I and *Sma*I and treated with CIP. The PCR product (0.2 μg) and vector (0.2 μg) were ligated in a final volume of 10 μL, and the ligation mixture was used to transform *E. coli* DH5α/P3. Vectors with inserts of the correct size were identified by restriction analysis. One of these (pDS-E2) was partially sequenced to verify accurate amplification and ligation. pDS-E2 was then digested with *Pst*I and *Sma*I, and the fragment containing the pro-mature portion of *sprE* was gel purified and ligated into pEB11 (digested with *Pst*I and *Sma*I and treated with CIP). *Escherichia coli* DH5α/P3 was transformed with the ligation mixture, and a vector containing the correct insert in the correct orientation was isolated on the basis of restriction enzyme analysis and designated pEB-E (Fig. 3). *Bacillus subtilis* DB104 (Kawamura and Doi 1984) was transformed with pEB-E and transformants containing the correct plasmid were identified using restriction enzyme digests. Expression of SGPE in these transformants was verified using a skim milk clearing assay (Wells et al. 1983).

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

Cloning and sequencing of sprE

Taking advantage of the previously published partial gene sequence of *sprE* (Svendsen et al. 1991), we used PCR to amplify a fragment of the gene encoding amino acids 44–121 of the mature protease. The amplified fragment was then used as a probe to isolate plasmids containing *sprE* from a genomic DNA library prepared from *S. griseus* DNA digested to completion with *Bgl*II. From approximately 20 000 *E. coli* transformants screened by colony hybridization, 48 potential positive clones were selected for further analysis. Southern blot analysis of plasmid DNA isolated from the 48 clones revealed four that contained inserts which hybridized strongly with the probe. Restriction analysis showed that each of the four plasmids contained an identical 4.7-kbp insert, and one of these plasmids (designated pDS-E) was chosen for sequencing.

Sequence analysis confirmed that *sprE* was contained within the 4.7-kbp insert of pDS-E. The complete sequence of *sprE* revealed an enzyme organized into a pre-pro-mature precursor (Fig. 2) consisting of 187 amino acid mature SGPE preceded by a 168 amino acid leader sequence. The pre region, determined by the method of von Heijne (1986) using the computer program PC/Gene (IntelliGenetics, Inc., Mountain View, Calif.), encompasses the first 29 amino acids of the extension and is typical of prokaryotic secretion signals. The remaining 139 residues constitute a propeptide.

The 5′ untranslated region of *sprE* contains putative promoter and ribosome-binding sites which were identified by comparison with other *Streptomyces* gene sequences (Henderson et al. 1987; Hutter and Eckhardt 1988). The translation stop codon is followed by a pair of inverted

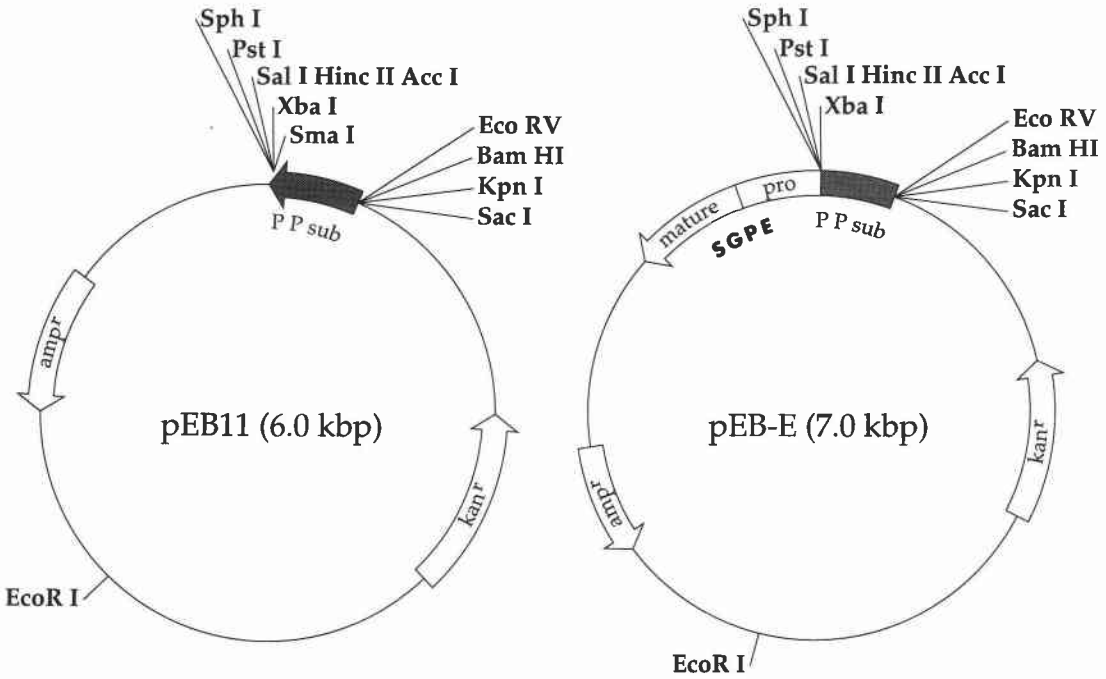

FIG. 3. Plasmid maps of pEB11 and pEB-E. Genes are represented by either open or solid boxes with arrowheads indicating the direction of transcription. The ampicillin resistance gene of pUC18 (*amp*^{r}), the kanamycin resistance gene of pUB110 (*kan*^{r}), and a 330-bp fragment from the subtilisin *BPN* gene (PP sub) are present in both vectors. pEB-E also contains the pro–mature region of *sprE* (SGPE) fused in frame with the subtilisin pre region, thus allowing recombinant expression of SGPE in *B. subtilis*.

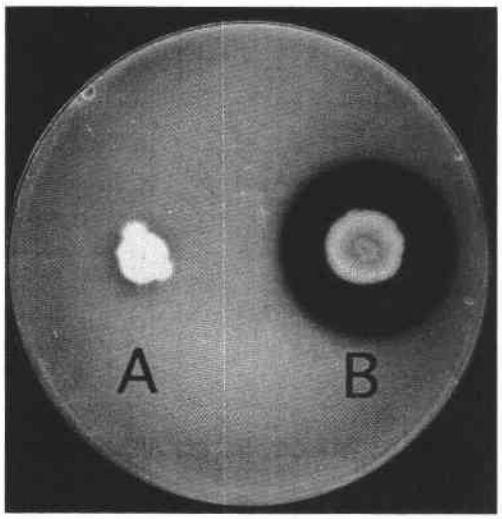

FIG. 4. Recombinant expression of SGPE. *Bacillus subtilis* DB104 transformed with either (A) pEB11 or (B) pEB-E was plated on YT plates containing 50 μg/mL of kanamycin and 1.5% skim milk. Plates were photographed after a 72-h incubation at 25°C. A zone of clearing indicating the secretion of active protease is visible only in the case of pEB-E transformant.

repeated sequences capable of forming a stable hairpin loop. Such structures, believed to be involved in transcription termination, have been identified in other *Streptomyces* genes (Henderson et al. 1987; Hutter and Eckhardt 1988).

Construction of an E. coli – B. subtilis *shuttle vector*

Bacillus subtilis is a gram-positive organism adapted for the secretion of hydrolytic enzymes (Harwood 1989); the protease-deficient DB104 strain of *B. subtilis* (Kawamura and Doi 1984) is an ideal host for recombinant expression of SGPE. Low-efficiency labour-intensive transformations make the genetic manipulation of DNA in *B. subtilis* generally more difficult than in *E. coli*. Therefore, we assembled the *E. coli* – *B. subtilis* shuttle vector pEB11 (Fig. 3) from the *E. coli* plasmid pUC18 and the *B. subtilis* plasmid pUB110. pEB11 contains origins of replication recognized by both organisms, an ampicillin resistance gene for selection in *E. coli* and a kanamycin resistance gene for selection in *B. subtilis*. The transformation efficiencies of pEB11 in *E. coli* and *B. subtilis* were found to be comparable to those of pUC18 and pUB110 in the respective hosts. pEB11 also contains a 330-bp PCR-amplified fragment derived from the *B. amyloliquefaciens* subtilisin gene. This fragment contains the region encoding the subtilisin secretion signal (pre region) preceded by 210 bp of 5′ untranslated region containing promoter and ribosome-binding sites recognized by *B. subtilis* (Wells et al. 1983). pEB11 was designed in such a way that insertion into the unique *Sma*I site of the vector produces an open reading frame consisting of the subtilisin pre region joined in-frame to the inserted DNA by a proline residue.

Expression of SGPE in B. subtilis

The pro–mature segment of *sprE* was cloned into pEB11 to produce the expression vector pEB-E (Fig. 3). The vector contains an open reading frame encoding a fusion protein composed of the pre region of subtilisin followed in-frame by the pro–mature segment of SGPE. Our cloning strategy added three additional codons to the *sprE* fragment and therefore three additional amino acids (Pro-Gly-Ala) to the amino terminus of the SGPE pro region. *Bacillus subtilis* transformants harbouring either pEB-E or pEB11 were tested on skim milk plates for secretion of proteases (Wells et al. 1983). Secretion of active SGPE was evidenced by the appearance of a zone of clearing (representing the degrada-

tion of milk proteins) around transformants containing pEB-E (Fig. 4).

Discussion

The *sprE* gene contains an open reading frame that encodes a 355 amino acid protein. The deduced amino acid sequence shows that the enzyme SGPE is organized as a pre-pro-mature enzyme. This form of organization is found in other bacterial members of the chymotrypsin family of serine proteases including SGPA and SGPB (Henderson et al. 1987), and the α-lytic protease of *L. enzymogenes* (Silen et al. 1988). Members of the subtilisin family of serine proteases are also organized into pre-pro-mature proteins (Wells et al. 1983). The mature enzyme is 187 amino acids in length, and notably, differs from the previously published amino acid sequence (Svendsen et al. 1991) at three positions. Residues Val132 and Glu177 were previously reported to be Gly and Lys, respectively, and the published amino acid sequence contained an additional carboxy-terminal Leu residue not present in our sequence. We have repeatedly sequenced the regions in which the discrepancies occur with sequencing templates derived from both strands of the gene. We, therefore, have a high degree of confidence in the derived amino acid sequence presented here.

The maturation of SGPA, SGPB, and the α-lytic protease is believed to involve two proteolytic cleavages. The first cleavage takes place at the junction of the pre and pro regions of the immature enzyme and requires the action of a leader peptidase (von Heijne 1986). A second cleavage occurs at the pro-mature junction and is thought to take place autocatalytically in all three enzymes. The boundary between the pro and mature regions of SGPE could be defined because the amino acid sequence (and therefore the amino terminus) of the mature protein has been determined (Svendsen et al. 1991). Notably, the *sprE* gene encodes a glutamic acid residue at the amino terminal side of the pro-mature junction, strongly suggesting that SGPE is a self-processing enzyme. In contrast to the bacterial members of the chymotrypsin family of serine proteases, the mammalian members of this family require the catalytic assistance of other enzymes for their maturation.

The complete amino acid sequence of SGPE is presented in Fig. 5 in an alignment with the sequences of enzymes SGPA, SGPB, the glutamic-acid-specific protease of *Streptomyces fradiae*, and the α-lytic protease. The gene sequence of SFPE was recently deposited in the GenBank (acquisition No. D12470), although to date a published characterization has not appeared. Mature SGPE is 83% identical with the *S. fradiae* enzyme. However, given that SGPE and SFPE are both of *Streptomyces* origin, this degree of homology is not surprising. Mature SGPE also shares considerable homology with SGPA (58% identity), SGPB (54% identity), and the α-lytic protease (36% identity). Thus, SGPA and SGPB (with a mutual identity of 61%) are as closely related to SGPE as to one another. The enzymes SGPA and SGPB have primary substrate specificity for large hydrophobic amino acids, whereas the primary specificity of the α-lytic protease is for alanine. Crystal diffraction studies suggest that primary specificity is a function of the amino acids that line the S1 binding pocket in the active sites of the enzymes. In SGPA and SGPB the S1 pocket is partially defined by the amino acid residues Ala190 and Thr213 (chymotrypsin numbering), whereas in α-lytic protease methionine residues occupy the homologous positions (Fig. 5). According to the sequence alignment presented in Fig. 3, both SGPE and SFPE contain Ser and His residues at positions equivalent to 190 and 213, respectively. Consequently, these residues are likely to play a role in determining the specificity of the glutamic-acid-specific proteases. This supposition is reasonable given the polar nature of Ser and His and the ability of these residues to form favourable interactions with a glutamic acid side chain.

The junction of the pre and pro regions of SGPE was determined by searching for potential prokaryotic secretion signals (von Heijne 1986). The most probable cleavage site was predicted to lie between residues 29 and 30. In general there is a poor consensus among the pre regions aligned in Fig. 5, although they have several common characteristics. Each pre region begins with a sequence rich in positively charged residues. This sequence is followed by a span of hydrophobic residues and the pre region culminates in a leader peptidase cleavage site. When the pre sequences were compared pairwise, they fell into three groups wherein the enzymes SGPA and SGPB formed one group, SGPE and SFPE formed a second, and the pre sequence of the α-lytic protease occupied a group of its own. Clearly, there is much less selective evolutionary pressure on pre region sequences than on the mature enzyme. Therefore, the extent to which pre region sequences diverge may correlate well with the evolutionary divergence of the various enzymes.

Comparison of the pro regions was complicated by large variations in the lengths of these sequences. Whereas the pro regions of SGPE, SFPE, and the α-lytic protease are 139, 141, and 166 amino acids in length, respectively, those of SGPA and SGPB contain only 78 and 76 residues, respectively. Consequently, the pro regions of SGPE, SFPE, and the α-lytic enzyme may be divided into two segments or domains of roughly equal length: a carboxy terminal domain (pro-C) homologous with the pro regions of SGPA and SGPB proteases and an amino terminal domain (pro-N) completely absent from SGPA and SGPB (Fig. 5). The deletion of only six amino acids from the amino terminus of the pro region of the α-lytic protease has been shown to completely abolish the secretion of active protease in *E. coli* (Fujishige et al. 1992). Interestingly, SGPA and SGPB, when expressed in a gram-positive host (Henderson et al. 1987), have no requirement for the pro-N domain whatsoever. It is clear that the pro region is involved in both the folding and secretion of active enzymes; hence, the presence of a pro-N domain may reflect differing requirements in either of these processes. The α-lytic protease is divergent from SGPA and SGPB to such an extent that it may require a larger and more divergent pro region for proper folding. Alternatively, the presence of the pro-N domain of the α-lytic protease may have been necessitated by the need to transverse two cellular membranes rather than a single membrane as in the case of the *S. griseus* proteases. The presence of a pro-N domain in the *S. griseus* protease E is at odds with both these explanations. Given that SGPE is secreted by a gram-positive organism and is closely related to SGPA and SGPB, what is the role of the pro-N domain? We are currently examining the involvement of the pro regions of SGPA, SGPB, and SGPE in folding and secretion in both gram-positive and gram-negative hosts.

As a first step in the study of protease folding and secretion, a *B. subtilis* expression system was developed for the

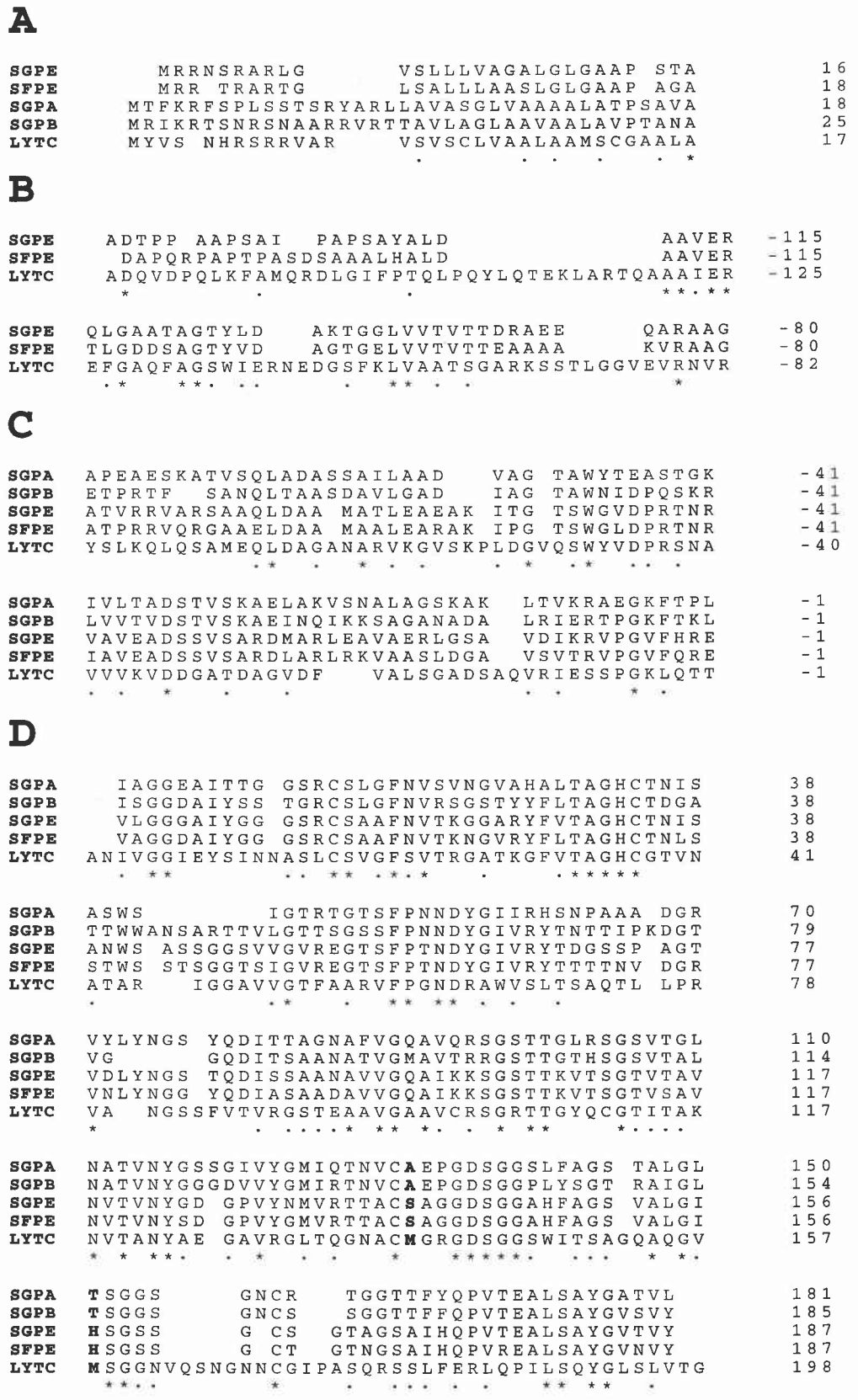

FIG. 5. Alignment of the amino acid sequence of promature SGPE with SGPA, SGPB (Henderson et al. 1987), SFPE (GenBank accession No. D12470), and α-lytic protease (Silen et al. 1988). Aligned separately are (A) the pre regions, (B) the amino terminal pro-N domains, (C) the carboxy terminal pro-C domains, and (D) the mature proteases. Numbering is from the amino terminus of the mature proteases. Residues 190 and 213 (chymotrypsin numbering) (Fujinaga et al. 1985) are shown in boldface. Positions with identical residues in all aligned sequences are indicated by asterisks, whereas conservatively substituted residues (Dayhoff 1972) are indicated with dots.

purpose of examining SGPE's ability to self-process. The shuttle vector pEB11 was created by recombining, within one plasmid, *E. coli* and *Bacillus* origins of replication as well as genes for resistance to the antibiotics ampicillin and kanamycin (Fig. 3). Also present within the pEB11 vector is a portion of the 5′ end of the subtilisin gene that includes subtilisin promoter and ribosome-binding sites, sequences encoding the subtilisin pre region, and a polylinker site. The pro–mature segment of *sprE* was amplified by PCR methods (simultaneously introducing restriction enzyme sites) and ligated into the polylinker site of pEB11, thereby creating the expression vector pEB-E. SGPE expressed from pEB-E is a fusion in which the subtilisin pre region replaces the SGPE pre region. Expression of SGPE was tested in the protease-deficient *B. subtilis* DB104. Clones harbouring the recombinant *sprE* produced a zone of clearing (Fig. 4) on skim milk plates and we took this as evidence of proper folding and autocatalytic maturation of SGPE.

Acknowledgments

This work was supported by a grant from the Natural Sciences and Engineering Research Council of Canada. S.S. is the recipient of a Medical Research Council of Canada Studentship and T.B. is the recipient of a British Columbia Health Research Foundation Scholarship. We thank also Mr. L. Willis for his assistance in DNA sequencing.

Chang, S., and Cohen, S.N. 1979. High frequency transformation of *Bacillus subtilis* protoplasts by plasmid DNA. Mol. Gen. Genet. **168**: 111–115.

Dayhoff, M.D. 1972. Atlas of protein sequence and structure 1972. National Biomedical Research Foundation, Washington, D.C.

Fersht, A.R. 1985. Enzyme structure and mechanism. W.H. Freeman and Company, New York.

Fujinaga, M., Delbaere, L.T.J., Brayer, G.D., and James, M.N.G. 1985. Refined structure of α-lytic protease at 1.7 Å resolution: analysis of hydrogen bonding and solvent structure. J. Mol. Biol. **183**: 479–502.

Fujishige, A., Smith, K.R., Silen, J.L., and Agard, D.A. 1992. Correct folding of α-lytic protease is required for its extracellular secretion from *Escherichia coli*. J. Cell Biol. **118**: 33–42.

Harwood, C.R. 1989. Introduction to the biotechnology of *Bacillus*. *In Bacillus. Edited by* C.R. Harwood. Plenum Press, New York. pp. 1–4.

Henderson, G., Krygsman, P., Liu, C.J., Davey, C.C., and Malek, L.T. 1987. Characterization and structure of genes for proteases A and B from *Streptomyces griseus*. J. Bacteriol. **169**: 3778–3784.

Hopwood, D.A., Bibb, M.J., Chater, K.F., Kieser, T., Bruton, C.J., Kieser, H.M., Lydiate, D.J., Smith, C.P., Ward, J.M., and Schrempf, H. 1985. Genetic manipulation of *Streptomyces*: a laboratory manual. The John Innes Foundation, Norwich.

Hutter, R., and Eckhardt, T. 1988. Genetic manipulation. *In* Actinomycetes in biotechnology. *Edited by* M. Goodfellow, S.T. Williams, and M. Mordarski. Academic Press, London. pp. 89–184.

James, M.N.G., Delbaere, L.T.J., and Brayer, G.D. 1978. Amino acid sequence alignment of bacterial and mammalian pancreatic serine proteases based on topological equivalences. Can. J. Biochem. **56**: 396–402.

Jurasek, L., Johnson, P., Olafson, R.W., and Smillie, L.B. 1971. An improved fractionation system for pronase on CM-sephadex. Can. J. Biochem. **49**: 1195–1201.

Kawamura, F., and Doi, R.H. 1984. Construction of a *Bacillus subtilis* double mutant deficient in extracellular alkaline and neutral proteases. J. Bacteriol. **160**: 442–444.

Lovett, P., and Amulos, N.P.J. 1989. Plasmid DNA isolation. *In Bacillus. Edited by* C.R. Hardwood. Plenum Press, New York. pp. 355.

Miller, J.H. 1972. Experiments in molecular genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor. pp. 433.

Neurath, H. 1984. Evolution of proteolytic enzymes. Science (Washington, D.C.), **224**: 350–357.

Norrander, J., Kempe, T., and Messing, J. 1983. Construction of improved M13 vectors using oligonucleotide-directed mutagenesis. Gene, **26**: 101–106.

Sambrook, J., Fritsch, E.F., and Maniatis, T. 1989. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor.

Silen, J.L., and Agard, D.A. 1989. The α-lytic protease pro-region does not require a physical linkage to activate the protease domain *in vivo*. Nature (London), **341**: 462–464.

Silen, J.L., McGrath, C.N., Smith, K.R., and Agard, D.A. 1988. Molecular analysis of the gene encoding α-lytic protease: evidence for a preproenzyme. Gene, **69**: 237–244.

Svendsen, I., Jensen, M.R., and Breddam, K. 1991. The primary structure of the glutamic acid-specific protease of *Streptomyces griseus*. FEBS Lett. **292**: 165–167.

Thomas, P.G., Russell, A.J., and Fersht, A.R. 1985. Tailoring the pH dependence of enzyme catalysis using protein engineering. Nature (London), **318**: 375–376.

Vogelli, G., and Kaytes, P.S. 1987. Amplification, storage and replication of libraries. Methods Enzymol. **152**: 407–415.

von Heijne, G. 1986. A new method for predicting signal sequence cleavage sites. Nucleic Acids Res. **14**: 4683–4690.

Wells, J.A., Ferrari, E., Henner, D.J., Estell, D.A., and Chen, E.Y. 1983. Cloning, sequencing, and secretion of *Bacillus amyloliquefaciens* subtilisin in *Bacillus subtilis*. Nucleic Acids Res. **11**: 7911–7925.

Wilkins, R.J., and Snell, R.G. 1987. Centrifugal transfer and sandwich hybridisation permit 12-hour Southern blot analyses. Nucleic Acids Res. **15**: 7200.

Yoshida, N., Tsuruyama, S., Nagata, K., Hirayama, K., Noda, K., and Makisumi, S. 1988. Purification and characterization of an acidic amino acid specific endopeptidase of *Streptomyces griseus* obtained from a commercial preparation (pronase). J. Biochem. (Tokyo), **104**: 451–456.