Characterization and Structure of Genes for Proteases A and B from *Streptomyces griseus*

GRAHAM HENDERSON, PHYLLIS KRYGSMAN, CI JUN LIU, CHERYL C. DAVEY, AND LAWRENCE T. MALEK*

Cangene Corporation, Mississauga, Ontario L4V 1T4, Canada

Received 17 February 1987/Accepted 15 May 1987

Protease A and protease B are extracellular proteins which are secreted by *Streptomyces griseus*. The genes encoding protease A (sprA) and protease B (sprB) were isolated from an *S. griseus* genomic library by using a synthetic oligonucleotide probe. Fragments containing sprA and sprB were characterized by hybridization and demonstration of proteolytic activity in *Streptomyces lividans*. Each DNA sequence contains a large open reading frame with the coding region of the mature protease situated at its carboxy terminus. The amino terminus of each reading frame appears to encode a 38-amino-acid signal peptide followed by a 76- or 78-amino-acid polypeptide, a propeptide, which is joined to the mature protease. Strong homology between the coding regions of the protease genes suggests that sprA and sprB originated by gene duplication.

*Streptomyces griseus*, an organism used for the commercial production of pronase, secretes many extracellular proteins (13). Protease A and protease B, two of the serine proteases secreted by *S. griseus*, have sequences which are 61% homologous on the basis of amino acid identity (8). These proteases also have similar tertiary structure, as determined by X-ray crystallography (6, 8, 11). Although the structures of proteases A and B have been extensively studied, the genes encoding the proteases have not been characterized.

This report describes the structure of two *S. griseus* genes which are responsible for the expression of protease A and protease B. The DNA sequences suggest that each protease is initially secreted as a precursor which is processed to remove an amino-terminal polypeptide (propeptide) from the mature protease. This propeptide may be important for the secretion of proteases in *Streptomyces* spp. We propose the genetic designations sprA and sprB for the unmapped genes for (serine) proteases A and B.

**MATERIALS AND METHODS**

**Strains and plasmids.** *S. griseus* IMRU3499 was obtained from the Waksman Institute of Microbiology, Rutgers University, Piscataway, N.J. *Streptomyces lividans* 66 (4) and plasmid pJ702 (14) were from the John Innes Institute. *Escherichia coli* HB101 (ATCC 33694) was used for all transformations. Plasmids pUC8 (28) and pUC18 and pUC19 (21) were purchased from Bethesda Research Laboratories, Inc.

**Media, growth, and transformation.** Growth of *Streptomyces* mycelium for the isolation of DNA or the preparation of protoplasts was as described previously (10). Protoplasts of *S. lividans* were prepared by lysozyme treatment, transformed with plasmid DNA, and selected for resistance to thiorhopteran, as described previously (10). Transformants were screened for proteolytic activity on LB plates (17) containing 30 μg of thiorhopteran per ml and 1% skim milk. *E. coli* transformants were grown on YT medium (17) containing 50 μg of ampicillin per ml.

**Materials.** Oligonucleotides were synthesized by using an Applied Biosystems 380A DNA synthesizer. Columns, phosphoramidites, and reagents used for oligonucleotide synthesis were obtained from Applied Biosystems, Inc., through Technical Marketing Associates. Oligonucleotides were purified by polyacrylamide gel electrophoresis followed by DEAE-cellulose chromatography. Enzymes for digesting and modifying DNA were purchased from New England Biolabs, Inc., and used in accordance with the recommendations of the supplier. The radioisotopes [α-32P]dATP (–3,000 Ci/mmole) and [γ-32P]ATP (–3,000 Ci/mmole) were from Amersham. Thiorhopteran was donated by E. R. Squibb & Sons.

**Isolation of DNA.** Chromosomal DNA was isolated from *S. griseus* as described previously (5), except that sodium dodecyl sarcosinate (final concentration, 0.5%) was substituted for sodium dodecyl sulfate. Plasmid DNA from transformed *S. lividans* was prepared by an alkaline lysis procedure (10). Plasmid DNA from *E. coli* was purified by a rapid boiling method (9). DNA fragments and vectors used for all constructions were separated by electrophoresis on low-melting-point agarose and purified from the molten agarose by phenol extraction and ethanol precipitation (15).

**Construction of genomic library.** Chromosomal DNA of *S. griseus* 3499 was digested to completion with BamHI and fractionated by electrophoresis on a 0.8% low-melting-point agarose gel. DNA fragments ranging in size from 4 to 12 kilobase pairs (kbp) were isolated from the agarose gel. The plasmid vectors pUC8 and pUC19 were digested with BamHI and treated with calf intestinal alkaline phosphatase (Boehringer Mannheim Biochemicals). The *S. griseus* BamHI fragments (0.3 μg) and vectors (0.8 μg) were ligated in a final volume of 20 μl as described previously (15). Approximately 8,000 transformants of *E. coli* HB101 were obtained from each ligation reaction.

**Subcloning of protease gene fragments.** A hybrid *Streptomyces-E. coli* vector was constructed by ligating pIJ702, which had been linearized with BamHI, into the *BamHI* site of pUC8. The unique *BglII* site of this vector was used for subcloning BamHI and BglII fragments of the protease genes. Other fragments were adapted with BamHI linkers to facilitate ligation into the *BglII* site. The hybrid vector, with pUC8 inserted at the BamHI site of pIJ702, was incapable of...
replicating in *S. lividans*. However, the *E. coli* plasmid could be readily removed before transformation of *S. lividans* by digestion with BamHI followed by recirculation with T4 DNA ligase.

**Hybridization.** A 20-mer oligonucleotide (5'-TTCCCGGACAACGAGACTACGG-3') was designed from an amino acid sequence (FPNNDYG) which is common to both proteases. For use as a hybridization probe, the oligonucleotide was end labeled by using T4 polynucleotide kinase (New England Biolabs) and [γ-32P]ATP. Digested genomic or plasmid DNA was transferred to a Hybond-N nylon membrane (Amersham) by electroblotting and hybridized in the presence of formamide (50%) as described previously (10). The filters were hybridized with the labeled oligonucleotide probe at 37°C for 18 h and washed at 47°C. The *S. griseus* genomic library was screened by colony hybridization as described previously (29).

**DNA sequencing.** The sequences of *sprA* and *sprB* were determined by using a combination of the chemical cleavage (16) and dideoxy (22) sequencing methods. Restriction fragments were end labeled by using either polynucleotide kinase or the large fragment of DNA polymerase I (Amersham), with the appropriate radiolabeled nucleoside triphosphate. Labeled fragments were either digested with a second restriction endonuclease or strand separated, followed by electrophoresis from a polyacrylamide gel. Subclones were prepared in the M13 bacteriophages mp18 and mp19 (21), and the dideoxy sequencing reactions were run by using the -20 universal primer (New England Biolabs). In some areas of strong secondary structure, compressions and polymerase failure necessitated the use of either inosine (18) or 7-deazaguanosine (19) analogs in the dideoxy reactions to clarify the sequence. The sequences were compiled with the software of DNASTAR (7).

**RESULTS**

**Screening for protease genes by hybridization.** Proteases A and B are homologous proteins containing several segments of identical sequence (8). An oligonucleotide probe was designed from one of these consensus amino acid sequences by using the known codon bias for *Streptomyces* spp. (2, 3, 24). The utility of the oligonucleotide probe was demonstrated by hybridization to genomic DNA of *S. griseus*. Since the oligonucleotide was designed from an identical amino acid sequence of both proteases, two DNA fragments should have been detected by hybridization analysis. As anticipated, the probe hybridized equally to two fragments generated by either BamHI (8.4 and 6.8 kbp) or BglII (11 and 2.8 kbp). The same fragments were detected in the genomic DNA of other *S. griseus* strains, but no such hybridization was observed with DNA from *S. lividans*.

By using the oligonucleotide probe, plasmids containing *sprA* and *sprB* were isolated from a genomic DNA library prepared from *S. griseus* DNA, which had been digested to completion with BamHI. Of 15,000 *E. coli* transformants that were screened by colony blot hybridization, 12 were detected by the oligonucleotide probe and isolated for further characterization. These colonies contained two distinct classes of plasmids, based on restriction analysis. As expected based on the hybridization of genomic DNA, the plasmids contained either a 6.8- or a 8.4-kbp BamHI fragment.

**Characterization of protease genes.** The DNA fragments isolated by hybridization screening were tested for expression of proteolytic activity. The 6.8- and 8.4-kbp BamHI fragments were ligated into the BglII site of the vector pIJ702, to allow transformation of *S. lividans*, with selection for thiostrepton resistance (14). Transformants containing these constructions were then tested on a milk plate for secretion of proteases. A clear zone, representing the degradation of milk proteins, surrounded each transformant that contained either BamHI fragment. The clear zones were not found around *S. lividans* colonies which contained either pIJ702 only or no plasmid. Thus, the two BamHI fragments each appear to contain a protease gene which is capable of effecting secretion in a different *Streptomyces* species.

The particular protease gene contained within each cloned
The BamHI fragment could be readily determined by dideoxy sequencing of the plasmids by using the oligonucleotide probe as a primer. The 8.4-kbp BamHI fragment was found to contain sprB, since a polypeptide deduced from the DNA sequence matched a unique segment of protease B (8). The 6.8-kbp BamHI fragment, which could not be sequenced by this method, was assumed to contain sprA.

The protease genes were localized within each BamHI fragment by determining which restriction fragments were capable of hybridizing to the oligonucleotide probe. Detailed restriction maps of the 6.8- and 8.4-kbp BamHI fragments are shown in Fig. 1. Hybridization to the oligonucleotide probe was confined to a 0.9-kbp PvuII-Stul fragment of sprA and a 0.6-kbp PvuII-PvuI fragment of sprB (Fig. 1). Hybridization to the cloned BamHI fragments and the 2.8-kbp BglII fragment of sprB was in essential agreement with the hybridization to BamHI and BglII fragments of genomic DNA. Thus, rearrangement of the BamHI fragments containing the protease genes is unlikely.

The functional limits of the genes were determined by subcloning restriction fragments into pIJ702, transforming S. lividans, and testing for proteolytic activity. The 3.2-kbp BamHI-BglII fragment of sprA and the 2.8-kbp BglII fragment of sprB, when subcloned into pIJ702 in either orientation, resulted in the secretion of a protease from S. lividans.

The intact protease genes could be further delimited to a 1.9-kbp Stul fragment for sprA and a 1.4-kbp BssHII fragment for sprB. Each of these functionally active subclones (Fig. 1) contained the region which hybridized to the oligonucleotide probe.

DNA sequence of protease genes. The 3.2-kbp BamHI-BglII fragment of sprA and the 2.8-kbp BglII fragment of sprB were subcloned into pUC18 to facilitate further structural characterization. The restriction maps of these subclones and the strategies used to sequence the 1.4-kbp SalI fragment containing sprA and the 1.4-kbp BssHII fragment containing sprB are shown in Fig. 2. The DNA sequences of these fragments are shown in Fig. 3. The predicted sequence of protease A differed from the published amino acid sequence (8) by the amidation of amino acid 133, whereas that of protease B was identical to the published sequence (8).

Each sequence contains a large open reading frame with the coding region for the mature protease situated at its 3' end. For each gene, the sequence encoding the carboxy terminus of the protease is followed immediately by a translation stop codon. However, at the other end, the predicted amino acid sequences appear to extend beyond the amino termini of the mature proteases by an additional 116 amino acids for sprA and 114 amino acids for sprB. The putative GTG initiation codons at these positions are each
preceded by a potential ribosome-binding site and followed by a sequence which could encode a signal peptide. A processing site for a signal peptidase can be predicted at 38 amino acids from the amino terminus of the putative precursor of each protease. The remaining sequence between the processing site and the mature amino terminus of each protease appears to represent the mature peptide. The genomic organization of sprA and sprB, based on the interpretation of the DNA sequence data, is shown schematically in Fig. 2.

The translation stop codons at the end of each coding region are followed by inverted repeated sequences which could form stable hairpin loops. Two of these structures were found after the coding region of sprA. The positions of the inverted repeats and their free energies are indicated in Fig. 3. These structures, which are similar to those characteristic in other Streptomyces genes (3, 24), are believed to be involved with termination of transcription. Interestingly, an inverted repeat was found 178 base pairs before the initiation codon of sprB. Preceding this structure was a potential coding region, with the characteristic Streptomyces codon usage, which was followed in frame by a TGA stop codon. If the inverted repeated sequence preceding the sprB coding region represents a transcription terminator, then the entire transcription unit of sprB would be defined. This corroborates the functional activity of the BssHII subclone.

**DISCUSSION**

**Codon usage and base content.** The codon selection for sprA and sprB (Table 1), which is similar to that of other

---

**FIG. 3.** DNA sequences of sprA (A) and sprB (B). The deduced amino acid sequences, which are numbered from the amino termini of the mature proteases, are shown in the DNA sequences. Stop codons are indicated by asterisks. The processing sites for the signal peptide (thin arrow) and the mature protease (thick arrow) are shown. Bases involved with the putative ribosome-binding sites are indicated by dots. Inverted repeat sequences are overlined. Indicated above these structures are the stabilities of the hypothetical hairpin loops (free-energy change in kilocalories [1 kcal = 4,184 J] per mole), calculated as described previously (25). DNA sequences which hybridized to the oligonucleotide probe are underlined.
sequenced *Streptomyces* genes (2, 3, 24), shows a clear bias for the use of C or G in position 3. The G+C content for each position with a codon is normal (54 to 60%) for positions 1 and 2 but anomalously high (92 to 96%) for position 3 (Table 2). In comparison, the base composition outside of the proposed open reading frames is uniformly distributed within each nucleotide triplet. The maintenance of a biased codon usage throughout each coding region and the disappearance of the triplet periodicity outside of the reading frames strengthens the assignment of initiation and termination codons for the proposed protease precursors.

### Homology of sprA and sprB

The alignment of the amino acid sequences translated from the coding regions of the *sprA* and *sprB* genes (Fig. 4) indicates an overall homology of 54% based on amino acid identity. However, the sequence homology is not uniformly distributed throughout the coding regions (Table 3). The carboxy-terminal domains of the proteases (6) are 75% homologous, whereas the average

![Table 1. Codon usage in sprA and sprB genes](data:image/png;base64,iVBORw0KGgoAAAANSUhEUgAHgAAAJwCAIAAABCFz6AAAgAElEQVR42u2wE+8B7QnCBkA90F...)

<table>
<thead>
<tr>
<th>Amino acid</th>
<th>Codon</th>
<th>No. of codons in:</th>
<th>Amino acid</th>
<th>Codon</th>
<th>No. of codons in:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>sprA</td>
<td></td>
<td>sprB</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Amino acid</th>
<th>Codon</th>
<th>No. of codons in:</th>
<th>Amino acid</th>
<th>Codon</th>
<th>No. of codons in:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*^a^ Stop codon.

### Table 2. G+C content of sprA and sprB

<table>
<thead>
<tr>
<th>Position in codon</th>
<th>Before coding*</th>
<th>G+C content (%) in region:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>sprA</td>
<td>sprB</td>
</tr>
<tr>
<td></td>
<td>sprA</td>
<td>sprB</td>
</tr>
<tr>
<td>1</td>
<td>75</td>
<td>71</td>
</tr>
<tr>
<td>2</td>
<td>61</td>
<td>69</td>
</tr>
<tr>
<td>3</td>
<td>66</td>
<td>66</td>
</tr>
<tr>
<td>All</td>
<td>67</td>
<td>69</td>
</tr>
</tbody>
</table>

*^a^ 177 base pairs preceding the initiation codon.

*^b^ 132 base pairs following the termination codon.
signal peptide in directing the secretion of correctly processed extracellular amylase (P. Krygsman, G. Henderson, L. Escote-Carlson, and L. Malek, manuscript in preparation).

The propeptide of the proposed protease precursor is probably not required for the initial membrane translocation stage of secretion, since this function was adequately performed by the signal peptide. However, the propeptide could be involved with facilitating secretion at a later stage, such as crossing the cell wall. The overall length and hydrophilic character of the propeptide could allow it to assume an extended conformation through a peptidoglycan matrix. However, amino acid homology suggests that the propeptide is not simply an inert spacer spanning the cell wall but probably has a definite, conserved structure. The propeptide may interact with the mature protease portion of the precursor, possibly inhibiting its activity until secretion is complete. Consistent with this hypothesis is the occurrence of the same consensus amino terminus, IXGG (Fig. 4C), in many eucaryotic serine proteases which have amino-terminal activation peptides (23, 27, 31).

The processing of each precursor protease is probably autocatalytic, since the amino acid sequence near the scissile bond should provide a good substrate (12, 20). Less likely would be the involvement of host proteases, which would have to be present in both *Streptomyces* species. The precursor proteases predicted from the structures of *sprA* and *sprB* are similar in general organization to those of the alkaline and neutral proteases of *Bacillus* spp., which have amino-terminal propeptides of 77 and 194 amino acids, respectively (26, 30). Also common is the proposed autocatalytic processing of the propeptides from the mature proteases. These similarities suggest that a precursor with an amino-terminal propeptide may be a general feature of proteases which are secreted by gram-positive bacteria.

**ACKNOWLEDGMENTS**

We are grateful to Robert T. Garvin and Eric James for their support of this work, helpful discussions, and critical review of the manuscript. We thank Gisela Soostmeyer for screening *Streptomyces* strains and preparing and testing transformants of *S. lividans*. We also acknowledge Donald Stewart and Burton Pogell for both helpful discussions and assays of proteolytic activity.

**LITERATURE CITED**