**A secreted aspartic proteinase from *Glomerella cingulata*: purification of the enzyme and molecular cloning of the cDNA**

Sarah J. Clark, Matthew D. Templeton and Patrick A. Sullivan

Author for correspondence: Patrick A. Sullivan. Tel: +64 6 350 6272. Fax: +64 6 350 5688. e-mail: P.A.Sullivan@massey.ac.nz

A secreted aspartic proteinase from *Glomerella cingulata* (GcSAP) was purified to homogeneity by ion exchange chromatography. The enzyme has an $M_r$ of 36 000 as estimated by SDS-PAGE, optimal activity from pH 3.5 to pH 4.0 and is inhibited by pepstatin. The N-terminal sequence, 23 residues long, was used to design a gene-specific primer. This was used in 3' RACE (rapid amplification of cDNA ends) PCR to amplify a 1.2 kb fragment of the gcsap cDNA. A second gene-specific primer was designed and used in 5' RACE PCR to clone the 5' region. This yielded a 600 bp DNA fragment and completed the open reading frame. The gcsap open reading frame encodes a protein with a 78 residue prepro-sequence typical of other fungal secreted aspartic proteinases. Based on the deduced sequence, the mature enzyme contains 329 amino acids and shows approximately 40% identity to other fungal aspartic proteinases. Subsequent cloning and sequencing of gcsap fragments obtained from PCR with genomic DNA revealed a 73 bp intron beginning at nt 728. Southern analyses at medium and high stringency indicated that *G. cingulata* possesses one gene for the secreted aspartic proteinase, and Northern blots indicated that gene expression was induced by exogenous protein and repressed by ammonium salts. GcSAP is a putative pathogenicity factor of *G. cingulata*, and it will now be possible to create SAP$^-$ mutants and assess the role GcSAP plays in pathogenicity.

**Keywords:** *Glomerella cingulata*, fungi, cDNA cloning, 3' and 5' RACE-PCR, secreted aspartic proteinase

**INTRODUCTION**

The plant-pathogenic fungus *Glomerella cingulata* (anamorph *Colletotrichum gloeosporioides*) causes anthracnose diseases of a variety of crops world-wide, but particularly in sub-tropical climates (Mordue, 1971; Irwin & Cameron, 1978). Infection involves the attachment of the spore to the host surface, differentiation to form an appressorium and direct penetration of the cuticle and epidermal cell walls. It is widely accepted that penetration of the cuticle results from both the mechanical force generated by a high internal pressure in the appressorium, and the activity of cutinase secreted by the fungus (Bailey et al., 1993). The subsequent development of the infection depends on the secretion of an array of enzymes, such as endopolygalacturonases, cellulases, pectin lyase, xylanase and β-galactosidase (for reviews see Collmer & Keen, 1986; Bailey et al., 1993; Cooper, 1977) which in concert degrade structural components of plant tissue.

Extracellular proteinases, especially those of animal pathogens, have been widely implicated as pathogenicity factors for several decades (for a review see Ogrydziak, 1993), but remarkably few of these studies have encompassed the proteinases produced by phytopathogenic fungi. Studies of an aspartic proteinase of *Botrytis cinerea* (Movahedi & Heale, 1990) showed that secretion of this enzyme is an early event in infection by *B. cinerea*, occurring before pectic enzyme production. Proteinase activity caused plant cell death and significantly higher levels of activity were secreted by virulent

**Abbreviations:** GcSAP, secreted aspartic proteinase from *Glomerella cingulata*; RACE, rapid amplification of cDNA ends. The GenBank accession number for the nucleotide sequence reported in this paper is U43775.
compared with less virulent strains of the fungus. It is also noteworthy that protease inhibitors are widespread in the plant kingdom (for reviews see Ryan, 1973; Avles, 1993) and it has been shown that inhibitor activity in melon seedlings infected with Colletotrichum lagenarium increased sharply 3 d after spore inoculation (Roby et al., 1987). Similarly, accumulation of a trypsin inhibitor in tomato plants has been correlated with Phytophthora infestans (Peng & Black, 1976). Collectively, these studies suggest, but do not unequivocally establish, that proteases have significant roles in pathogenicity.

Production of a proteolytic activity by G. cingulata was first detected when sections of apple fruit rotted by G. cingulata were assayed for enzyme activity (Kuc & Williams, 1962). More recently, G. cingulata has become a particularly suitable organism for the assessment of putative pathogenicity factors with the development of protocols for targeted gene disruptions (Rikkerink et al., 1994; Bowen et al., 1995) and the identification of genes selectively expressed during appressorium formation (Hwang & Kolatukudy, 1995). It should now be possible to assess the roles of the secreted protease, and as a first step towards this objective we describe here the purification of a G. cingulata secreted aspartic proteinase (GcASP), some properties of the enzyme, and molecular cloning of the cDNA and the gene encoding the enzyme.

METHODS

Organisms and plasmids. G. cingulata was obtained from the International Collection of Micro-organisms from Plants (ICMP), Maanaki Whenua Landcare NZ. Escherichia coli strain DH5α was the host for recombinant plasmids. The plasmid pBluescript KS II(+) (Stratagene) was used for cloning experiments.

Media and reagents. G. cingulata cultures were maintained on potato dextrose agar. E. coli cultures were grown in Luria broth, supplemented with ampicillin as required (Sambrook et al., 1989). BSA (fraction V) was obtained from Boehringer Mannheim, and bovine haemoglobin (type 11) was supplied by Boehringer Mannheim. Oligonucleotides reverse transcriptase were from Gibco BRL. Oligonucleotides were produced by an Applied Biosystems 380B synthesizer using β-cyanoethyl chemistry.

Enzyme purification. Twenty 200 ml cultures of G. cingulata were grown in a nitrogen-free salts medium (Wickerham, 1946) containing 0·2% (w/v) BSA and 2% (w/v) glucose (static for 3 d in 1 Roux flasks at 26 °C). After harvesting the mycelium by centrifugation, the supernatant was adjusted to pH 7·0 with 6 M NaOH and concentrated to approximately 300 ml by ultrafiltration (molecular mass cut-off 10 000 Da). After dialysis for 17 h against 10 mM sodium citrate buffer, pH 6·8, the concentrate was applied to a DEAE-Sepharose column (2·5 × 17 cm) equilibrated with buffer, and the column washed with similar buffer at a flow rate of 1 ml min⁻¹ until the A₂₈₀ returned to zero. Bound protein was eluted with a linear gradient of 10 to 250 mM sodium citrate, pH 6·3 (400 ml) and was collected in 10 ml fractions. Fractions 15–30 contained protease activity and were pooled and concentrated to approximately 15 ml by ultrafiltration and dialysed against 20 mM Bis-tris buffer, pH 6·0 (fraction 1). Aliquots (0·8 ml) of this concentrate were applied to a Mono Q HR 5/5 column (Pharmacia) equilibrated in the same buffer. The column was eluted with a gradient of 0 to 200 mM NaCl in 20 mM Bis-tris, pH 6·3, over 30 min at a flow rate of 0·75 ml min⁻¹. Fractions containing the enzyme were pooled and stored at −20 °C (fraction 2).

Protease activity was estimated by measuring the hydrolysis of (a) BSA or (b) haemoglobin. Enzyme activity was also detected in the growth medium with Azocoll (Calbiochem).

Haemoglobin assay. Each assay mixture contained 0·2 ml enzyme solution and 1 ml acid-denatured haemoglobin (2% (w/v) in 0·2 M glycine/HCl, pH 3·5). After incubation at 30 °C for 30 min, a 300 μl aliquot was mixed with 200 μl 10% (w/v) trichloroacetic acid and incubated on ice for 10 min and then centrifuged (13 000 r.p.m. for 10 min). The absorbance of the supernatant was read at 280 nm and corrected for background using a zero-time control in which trichloroacetic acid was added prior to the enzyme. One unit of enzyme was defined as that amount which catalysed a change in A₂₈₀ of 1·0 min⁻¹.

Protein estimations. These were performed using the bichinchoninic acid (BCA) assay (Smith et al., 1985).

DNA and RNA manipulations. Restriction endonucleases and DNA-modifying enzymes were incubated under the conditions recommended by the suppliers. Standard conditions for molecular cloning, hybridization, transformation and electrophoresis were used (Sambrook et al., 1989). Molecular mass markers were HindIII fragments of bacteriophage λ (Gibco...
BRL. Purification of DNA fragments from agarose gels was done as described by Vogelstein & Gillespie (1979). DNA sequencing was performed on an ABI 373A automated sequencer using standard ABI protocols. G. cingulata DNA was isolated as described by Raeder & Broda (1985). G. cingulata RNA for RACE was extracted from either 48 h stationary cultures or 36 h shake cultures by the methods of Carlson & Botstein (1982) and Teeri et al. (1987), respectively. These cultures were grown in the BSA medium described above. The method of Teeri et al. (1987) was also used to extract RNA for Northern analysis. Poly(A+) RNA was purified from 1 mg total RNA in a batchwise manner using 30 mg microcrystalline oligo (dT)-cellulose (New England Biolabs). Binding and elution buffers used were those recommended by the manufacturer. For Southern blotting, the membranes were hybridized at 56 °C and then washed as follows: twice with 2 x SSC (1 x SSC is 0.15 M NaCl, 0.015 M sodium citrate), 0.5% SDS at room temperature for 10 min, and then 2 x SSC, 0.5% SDS at 63 °C for 1 h (medium stringency). This was followed by a higher stringency wash of 0.1 x SSC, 0.5% SDS at 68 °C for 1 h. Northern membranes were hybridized at 65 °C and then washed twice with 2 x SSC, 0.5% SDS at room temperature for 15 min, and then twice at 65 °C for 10 min.

Rapid amplification of cDNA ends (RACE)

3' RACE. The basic protocol for 3' RACE was that described by Ausubel et al. (1992). Three primers were required. A gene-specific oligonucleotide (SAP-primer1) (5'-GGGGTACCGAGATGGACCAAG-G/A/C/T)TA[12] was based on the sequence of amino acid residues QYDVEY in the determined N-terminal sequence of mature GcSAP. SAP-primer1 also contained a KpnI site. A polyT primer (5'-GGGGTACCGTGACATCGA[C/T][A/G]AG-[G/A/C/T]TA) was used as a restriction site for XbaI and was used to make AmpliFINDER anchors were ligated to the 3' ends of anchored cDNA was used in PCR with SAP-primer1. Total RNA was reverse-transcribed under conditions recommended by Gibco BRL. For 20 μl PCR reactions, 2 μl undiluted cDNA was used. Optimized PCR conditions included 5 cycles with an annealing temperature of 56 °C and a further 35 cycles with annealing at 55 °C. The MgCl₂ concentration was 2.5 mM.

5' RACE. This was performed using the 5'-AmpliFINDER RACE kit from Clontech. Two gene-specific primers were synthesized. Oligonucleotide SAP-primer2 (5'-GCTCTAGAAAAATGTTCCTCTGT) contained sequence complementary to the region from nt 652 to 666 as well as a restriction site for XbaI and was used to make first strand cDNA. Oligonucleotide SAP-primer4 (5'-GGGGTACCGTGAAGGATGGACCA) contained sequence complementary to the region from nt 454 to 470 and a restriction site for KpnI. After first strand cDNA had been made, AmpliFINDER anchors were ligated to the 5' ends of the cDNAs under conditions recommended by Clontech. The anchored cDNA was used in PCR with SAP-primer4 and an anchor primer provided in the 5' RACE kit. The anchor primer contained an EcoRI site. Optimized PCR conditions included an annealing temperature of 60 °C and a MgCl₂ concentration of 2.5 mM in a reaction volume of 50 μl.

PCR of a genomic sap fragment. G. cingulata genomic DNA was used as the template in a PCR reaction with primers SAP-primer1 and SAP-primer3 (5'-GGGGTACCGATGACGAAGCAACGAA). SAP-primer3 contained sequence complementary to the region from nt 1238 to 1255. Annealing was done at 57 °C for 5 cycles, then at 56 °C for 30 cycles. The MgCl₂ concentration was 2.5 mM.

Computer programs. Editing, analysis and alignment of sequences was carried out using the ABI SeqEd Macintosh program, and PILEUP, LINEUP and SEQUED programs from the Genetics Computer Group (Wisconsin Sequence Analysis Package, version 8). Autoradiograms were scanned with a Bio-Rad GS-670 Imaging densitometer and the digital images were processed with Adobe Photoshop 3.0 and Canvas 3.5.1 from Denaba systems.

RESULTS AND DISCUSSION

Induction and purification of GcSAP

Providing protein as the sole nitrogen source in fungal cultures has been widely documented as a system for induction of proteinase production (Ross et al., 1990, and references cited therein) and this strategy was used for G. cingulata proteinase induction. G. cingulata grew well when cultured in salts medium containing either (NH₄)₂SO₄ or BSA as the sole nitrogen source. Proteinase activity in the culture medium was monitored for 5 d and detected in the BSA cultures in assays with either BSA or Azocoll as the substrate. Maximum proteinase activity was observed on day 3 in the BSA medium and there was no proteinase activity in the medium in which (NH₄)₂SO₄ was the nitrogen source.

For proteinase enzyme purification (see Methods for details), G. cingulata was cultured in BSA medium for 3 d. Following concentration by ultrafiltration, the growth medium was dialysed against 10 mM sodium citrate buffer, pH 6.8, and the crude enzyme was applied to a column of G-25 desalting resin and the enzyme was reapplied to a column of DEAE-Trisacryl. Elution of the enzyme from the DEAE-Trisacryl column was performed by a linear NaCl gradient (0-0.1 M). Fractions containing the proteinase were pooled and dialysed against 10 mM sodium citrate buffer, pH 6.8, and applied to a Tris-HCl-Tricine gel filtration column (Pharmacia). The enzyme was eluted from this column at a molecular mass of 29 kDa.

Fig. 1. SDS-PAGE of GcSAP. Lane 1, purified GcSAP (10 μg of fraction 2, as described in Table 1); lane 2, molecular mass standards as indicated on the right.
to a DEAE-Sepharose column. Most of the protein, as judged by the $A_{280}$ profile, was eluted during loading, washing with buffer, and in the 10–100 mM sodium citrate gradient. The proteinase was recovered in fractions 15–30 from 100–200 mM sodium citrate, pH 6.3. These fractions were pooled, concentrated and dialysed against 20 mM Bis-tris buffer, pH 6–7.0) and at pH 7.0 was 40% of the maximum value ever, in contrast with other secreted aspartic proteinases, which has an activity of 6–8 U mg$^{-1}$ (Wright et al., 1992). However, the specific activity of GcSAP protein, the lengths of typical 3’ untranslated sequences and poly(A$^+$) tails associated with eukaryotic mRNA transcripts, it was anticipated that the PCR product would be approximately 1.2 kb. PCR with the adaptor primer and gene-specific primer SAP-primer1 yielded two products close to the expected size (1.2 kb and 990 bp). Control PCR reactions were performed with only one primer (the adaptor primer) and cdna was derived from either (NH$_2$)$_2$SO$_4$ or BSA cultures. These eliminated the 990 bp PCR product as a cdna fragment, and the 1.2 kb product was reamplified from a band stab. The amplified fragment was digested with KpnI, gel-purified and ligated into the KpnI site of pBluescript KS II(+) yielding pSJ1. Sequencing of pSJ1 confirmed that the PCR product coded for the gcsap cdna because the deduced amino acid sequence matched exactly with the N-terminus of GcSAP determined by the Edman degradation. The gcsap cdna fragment consisted of 960 bp of coding sequence, which represented 97% of the mature protein, and contained 200 bp of downstream sequence (Fig. 2).

3’ RACE. Total RNA was extracted from G. cingulata grown in BSA medium and used to make cdna templates for PCR. Based on the size of the GcSAP protein, the lengths of typical 3’ untranslated sequences and poly(A$^+$) tails associated with eukaryotic mRNA transcripts, it was anticipated that the PCR product would be approximately 1.2 kb. PCR with the adaptor primer and gene-specific primer SAP-primer1 yielded two products close to the expected size (1.2 kb and 990 bp). Control PCR reactions were performed with only one primer (the adaptor primer) and cdna was derived from either (NH$_2$)$_2$SO$_4$ or BSA cultures. These eliminated the 990 bp PCR product as a cdna fragment, and the 1.2 kb product was reamplified from a band stab. The amplified fragment was digested with KpnI, gel-purified and ligated into the KpnI site of pBluescript KS II(+) yielding pSJ1. Sequencing of pSJ1 confirmed that the PCR product coded for the gcsap cdna because the deduced amino acid sequence matched exactly with the N-terminus of GcSAP determined by the Edman degradation. The gcsap cdna fragment consisted of 960 bp of coding sequence, which represented 97% of the mature protein, and contained 200 bp of downstream sequence (Fig. 2).

5’ RACE. Poly(A$^+$) RNA was purified from RNA enriched with the transcript of GcSAP and used to make the cdna template for PCR. Following the ligation of an anchor oligonucleotide to the 3’ ends of the cdna preparation, PCR was performed using the anchor primer and the gene-specific primer SAP-primer4. One product of 600 bp was observed on agarose gels and this was subsequently cloned between the KpnI and EcoRI sites of pBluescript KS II(+) (+). Two clones were sequenced and this confirmed that the PCR product was the 5’ region of the gcsap cdna. The 5’ cdna fragment contains 104 bp of upstream noncoding sequence and 234 bp of sequence downstream of the initiation ATG that encodes a typical prepro-sequence of 78 amino acids, and this is followed by sequence coding for the mature protein. This sequence matched exactly that previously determined by 3’ RACE.

### Table 1. Purification of GcSAP

<table>
<thead>
<tr>
<th>Fraction*</th>
<th>Volume (ml)</th>
<th>Activity (U)†</th>
<th>Protein (mg)</th>
<th>Specific activity (U mg$^{-1}$)</th>
<th>Yield (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$F_{o-c}$</td>
<td>544</td>
<td>10.5</td>
<td>6800</td>
<td>0.0015</td>
<td>100</td>
</tr>
<tr>
<td>$F_1$</td>
<td>14.5</td>
<td>6.3</td>
<td>8.7</td>
<td>0.7</td>
<td>60</td>
</tr>
<tr>
<td>$F_2$</td>
<td>10.5</td>
<td>1.5</td>
<td>2.0</td>
<td>0.8</td>
<td>14</td>
</tr>
</tbody>
</table>

* $F_{o-c}$; concentrated growth medium; $F_1$, concentrated dialysed fraction from DEAE-Sepharose chromatography; $F_2$, pooled active fractions (peak 3) from Mono Q chromatography.
† Activity was measured in the BSA assay (see Methods).

### Properties of GcSAP

GcSAP exhibited properties typical of a fungal secreted aspartic proteinase. In assays with haemoglobin as substrate maximum activity was at pH 3.5–4.0. However, in contrast with other secreted aspartic proteinases, significant activity was observed over a wide range (pH 2.0–7.0) and at pH 7.0 was 40% of the maximum value (data not shown). The activity was completely inhibited by pepstatin, a specific aspartic proteinase inhibitor. The purified GcSAP also degraded BSA, keratin and casein, but haemoglobin was the best substrate. The specific activity of purified GcSAP in the BSA assay was 0.8 U mg$^{-1}$ which is low compared to SAP2 from *C. albicans*, aspartic proteinase 1, as shown in Fig. 3.

### Cloning and sequencing of the gcsap cdna

**3’ RACE.** Total RNA was extracted from *G. cingulata* grown in BSA medium and used to make cdna templates for PCR. Based on the size of the GcSAP protein, the lengths of typical 3’ untranslated sequences and poly(A$^+$) tails associated with eukaryotic mRNA transcripts, it was anticipated that the PCR product would be approximately 1.2 kb. PCR with the adaptor primer and gene-specific primer SAP-primer1 yielded two products close to the expected size (1.2 kb and 990 bp). Control PCR reactions were performed with only one primer (the adaptor primer) and cdna was derived from either (NH$_2$)$_2$SO$_4$ or BSA cultures. These eliminated the 990 bp PCR product as a cdna fragment, and the 1.2 kb product was reamplified from a band stab. The amplified fragment was digested with KpnI, gel-purified and ligated into the KpnI site of pBluescript KS II(+) (+), yielding pSJ1. Sequencing of pSJ1 confirmed that the PCR product coded for the gcsap cdna because the deduced amino acid sequence matched exactly with the N-terminus of GcSAP determined by the Edman degradation. The gcsap cdna fragment consisted of 960 bp of coding sequence, which represented 97% of the mature protein, and contained 200 bp of downstream sequence (Fig. 2).

**5’ RACE.** Poly(A$^+$) RNA was purified from RNA enriched with the transcript of GcSAP and used to make the cdna template for PCR. Following the ligation of an anchor oligonucleotide to the 3’ ends of the cdna preparation, PCR was performed using the anchor primer and the gene-specific primer SAP-primer4. One product of 600 bp was observed on agarose gels and this was subsequently cloned between the KpnI and EcoRI sites of pBluescript KS II(+) (+). Two clones were sequenced and this confirmed that the PCR product was the 5’ region of the gcsap cdna. The 5’ cdna fragment contains 104 bp of upstream noncoding sequence and 234 bp of sequence downstream of the initiation ATG that encodes a typical prepro-sequence of 78 amino acids, and this is followed by sequence coding for the mature protein. This sequence matched exactly that previously determined by 3’ RACE.

*Rhizopus chinensis* rhizopinepsin II and *Rhizopus niveus* aspartic proteinase 1, as shown in Fig. 3.
Secreted aspartic proteinase from G. cingulata

Fig. 2. Nucleotide sequence of gcsap and derived amino acid sequence. These data were obtained from cDNA and genomic clones. The predicted signal peptidase and KEX2 processing sites in the prepro-sequence are indicated by (V) and underlined text (KR) respectively. The N-terminal sequence data determined by Edman degradation are double-underlined. The conserved and repeated nucleotide sequences in untranslated regions, as discussed in the text, are underlined. The intron (nt 728-800) is in lower case letters.

Intervening sequence

Genomic sequence of the gcsap gene was obtained by performing PCR with SAP-primer1 and SAP-primer3 using genomic DNA as the template. A single product of 950 bp was produced and cloned into the KpnI site of pBluescript KS II (+). Sequencing of the resulting clone revealed a 73 bp intron beginning at nt 728 (Fig. 2). The intron/exon splice junctions (GTAAG and CAG) and putative lariat structure (TAGTGACA) fit the consensus sequences found in other filamentous fungi (Gurr et al., 1987).

The untranscribed sequences

The 5' untranscribed region contains a triple repeat of the motif CTCATC between nt -80 and -61 (Fig. 2). Similar repeats have been detected in the same region of many Aspergillus nidulans genes (Ward & Turner, 1986)
Fig. 3. Alignment of fungal aspartic proteinases. GcSAP, secreted aspartic proteinase of G. cingulata (GenBank accession number U43775); RacSAP, R. chinensis rhizopuspepin II (L33857); Rnasap, R. niveus aspartic proteinase 1 (Horiuchi et al., 1988, M19100); Casap2, C. albicans SAP2 (Wright et al., 1992, M83663); Mmasp, M. miehei aspartic proteinase (Gray et al., 1986, M15267). Paired basic amino acids in the prepro-sequence are underlined and the N-terminal residues of the mature enzymes are in bold type and underlined. The catalytic aspartic residues are in bold type, and conserved cysteine residues are indicated by shading.

and also 5' of the initiation ATG of the gpdA gene of G. cingulata (Templeton et al., 1992). The sequence directly upstream of the start codon (CAAGATGAC) closely resembles the consensus sequence (CAC/AA/cCATGGC) for filamentous fungi, although a G in the -1 position is rare (Gurr et al., 1987).

The 3' region of the gcsap cDNA was cloned using a poly(T) primer which presumably annealed to the poly(A) region of the mRNA. However, there is no consensus eukaryotic polyadenylation signal (AAUAAA) in the 3' untranslated region. But this region (nt 1309-1504) contains a number of intriguing features: the sequence from nt 1309 to 1328 contains eight GA repeats; there are two copies of a GAGGAC motif between nt 1380 and 1394; and a palindrome of the motif TATAC is located 5-20 nt before the end of the sequence. The significance of these features is not known.

Analysis of genomic DNA

Southern analysis was used to determine the copy number and to detect any closely homologous genes in the genome. Genomic DNA from G. cingulata was digested with EcoRI, BamHI and KpnI, which do not cut within the cloned sequence, and with SalI, PstI, SacI and BglII, which each cut once within the sequence. Hybridization with a 950 bp probe and washing at both medium and high stringency revealed one band in each digest, except for the BglII digest which had a major
Secreted aspartic proteinase from *G. cingulata*

**Fig. 4.** Southern analysis of gcsap. A Southern blot of restriction-enzyme-digested genomic DNA was probed with a 950 bp gene fragment. The region of the gene covered by the probe is shown in Fig. 5. Lanes: 1, EcoRI; 2, BamHI; 3, PstI; 4, SalI; 5, KpnI; 6, SalI; 7, BglII. (a) Membrane washed at medium stringency; (b) membrane washed at high stringency (see Methods).

**Fig. 5.** Restriction map of the gcsap gene sequence. The PCR product amplified from genomic DNA using SAP-primer1 and SAP-primer3 (see Methods) was cloned and used as a probe, and this is represented by the solid line below the map. The coding regions are represented by the black boxes. The intron and other untranslated regions are shown as a narrow line.

band at 10.6 kb and a faint band at 1.7 kb (Fig. 4a, b). These results are consistent with the restriction map of the gcsap gene and the region covered by the probe (Fig. 5). The SalI site is upstream from the region which hybridized to the probe and the second PstI and SalI fragments only overlap the probe by 70 bp and were not detected. These results indicate that the genome does not contain other genes that are closely related to gcsap.

**Expression of gcsap**

The secretion of GcSAP is induced by exogenous protein and repressed by ammonium salts. A Northern analysis indicated that this regulation is at the level of transcription. As shown in Fig. 6, a transcript of 1.7 kb was detected when BSA was the sole nitrogen source but was absent when NH₄Cl was the sole nitrogen source. These results are consistent with the pattern of proteinase secretion under these growth conditions. There was also no gcsap transcript detected when the growth media contained both BSA and NH₄Cl, indicating that ammonium salts repressed expression of the gene.

**The amino acid sequence**

The deduced amino acid sequence of GcSAP is 38 % and 51 % identical with rhizopuspepsin (Delany *et al.*, 1987) and penicillopepsin (James & Sielecki, 1983) respectively. The alignment in Fig. 3 indicates in particular that the regions surrounding the two catalytic aspartic acid residues are absolutely conserved.

The N-terminus of the mature enzyme starts at position 79 (Fig. 2), as determined by Edman sequencing of pure GcSAP. Most aspartic proteinases possess a pro-sequence of 50 residues (Davies, 1990) and a prepro-sequence of 70–80 residues and the predicted prepro-sequence in GcSAP (78 residues) is typical. It contains a hydrophobic core signal sequence from position 3 to 12.
and there is a putative signal peptidase cleavage site between residues 16 and 17 (Fig. 2). Cleavage here would yield a protein with a pro-sequence of 62 amino acids. The pro-sequence contains multiple basic residues including three pairs of basic residues; basic residues are common in pro-sequences of aspartic proteinases (Davies, 1990). The last of the paired basic residues (Lys<sub>16</sub>-Arg<sub>17</sub>) is adjacent to the mature N-terminus and is probably the cleavage site of a KEX2-type processing enzyme (Bussey, 1988). The other basic amino acid residues in the pro-sequence may bind to the active site aspartic acid residues, as proposed for pepsinogen (James & Sielecki, 1986), thus giving the pro-sequence a possible role in self-inhibition of pro-GcSAP. Based on comparisons with pepsinogen, probable candidates are Arg<sub>53</sub>, Arg<sub>53</sub> and Lys<sub>54</sub> as they are each adjacent to a tyrosine residue, which in pepsinogen stabilizes the interaction by hydrogen-bonding to the active site aspartic acids.

There are no Asn-Xxx-Ser/Thr motifs in the sequence and hence no potential sites for N-glycosylation. The SDS-PAGE-estimated $M_r$ of 36000 is consistent with a deduced $M_r$ of 34200.

GcSAP contains only two Cys residues, at positions 332 and 365 (Fig. 2). These residues align with conserved Cys residues in other aspartic proteinases (Fig. 3) which are known to form a disulphide bond (Cutfield et al., 1995). Interestingly, GcSAP does not possess the first pair of conserved Cys residues (positions 133 and 145, Fig. 3) beyond the first active site aspartic residue. This disulphide bond loop is common in many aspartic proteinases, but it is clearly not essential for activity because it is not present in some other aspartic proteinases, for example endothiopepsin and penicillopepsin (Cutfield et al., 1995).

Many filamentous fungi and some yeasts secrete aspartic proteinases; for example, C. albicans, the common human fungal pathogen, possesses at least eight genes encoding enzymes of this type (Hube et al., 1994) and Rhizopus species such as R. niveus also have a similar family (Delaney et al., 1987). Other examples include Aspergillus awamori, Mucor miehei and Endothia parasitica (for a review see Ward & Kodama, 1991). The properties of the GcSAP as determined in this study, for example $M_r$ conserved regions of amino acid sequence, inhibition by pepstatin and optimum activity at acidic pH values, are consistent with those for other fungal aspartic proteinases.

Although the sequence identity of aspartic proteinases from diverse sources is often no greater then 20–27%, the crystal structures of numerous members of the family have revealed common features such as a predominance of $\beta$ structure and organization of the polypeptide into two domains with the two DTG motifs in the extended active site cleft (Cutfield et al., 1995). GcSAP will undoubtedly have these general features, together with unique detailed features. With the molecular cloning of the cDNA for GcSAP it will now be possible to assess the role of the enzyme in pathogenicity by a targeted gene disruption.

ACKNOWLEDGEMENTS

This research was supported by contract UOO 401 provided by the New Zealand Foundation for Research Science and Technology. We are grateful for the assistance of the Protein Micro-Chemistry Facility, Biochemistry Department, University of Otago, for the protein sequencing work. Nucleotide sequencing was performed by the Centre for Gene Research, Biochemistry Department, University of Otago. We thank Mrs Lana Ellis for technical assistance with the Northern analysis.

REFERENCES


Secreted aspartic proteinase from G. cingulata


Received 22 October 1996; accepted 4 December 1996.