

Identification and characterization of a second polygalacturonase gene of *Aspergillus niger*

H. J. D. Bussink, K. B. Brouwer, L. H. de Graaff, H. C. M. Kester, and J. Visser

Section of Molecular Genetics, Department of Genetics, Agricultural University, NL-6703 HA Wageningen, The Netherlands

Received March 26/May 29, 1991

Summary. The filamentous fungus *Aspergillus niger* produces several endopolygalacturonases that are involved in the degradation of pectin. PGI, the enzyme representing the second most abundant activity in a commercial enzyme preparation, was further characterized and the corresponding gene was isolated. The nucleotide sequence of the *pgaI* gene was determined and the protein coding region was found to be interrupted by two short introns, one of which has a unusual donor splice site. The deduced 368 amino acids long protein with a putative prepropeptide of 31 amino acids shows 60% sequence identity to PGII in the mature protein. PGI overproducing *A. niger* strains were obtained by cotransformation with the cloned gene.

Key words: Pectin degradation – Filamentous fungus – Introns – Overexpression

Introduction

Many fungal species are known to produce multiple molecular forms and isoenzymes of endopolygalacturonase (Rombouts and Pilnik 1980; Cervone et al. 1986). Endopolygalacturonases (endoPG; EC 3.2.1.15) hydrolyse pectate, a de-esterified derivative of pectin, which is primarily found in the middle lamella and primary cell wall of higher plants. Pectin is a heteropolysaccharide consisting of a backbone of D-galacturonic acid residues which are partially esterified and connected by α -1,4 glycosidic bonds and with few L-rhamnose residues in the main chain. Other neutral sugars, such as L-arabinose, D-galactose and D-xylose, compose the side chains often bound to L-rhamnose (de Vries et al. 1986). Some microorganisms, both phytopathogens and saprophytes, produce several distinguishable pectinolytic enzymes involved in the degradation of this complex substrate (Rexová-Benková and Markovic 1976; Rombouts and Pilnik

1980). Recently, genomic or cDNA clones coding for pectinesterase (Khanh et al. 1990) and the pectic depolymerizing enzymes endopectin lyase D and A (Gysler et al. 1990; Harmsen et al. 1990), endopectate lyase (Dean and Timberlake 1989b) and endopolygalacturonase (Bussink et al. 1990; Ruttkowski et al. 1990) have been obtained from *Aspergillus* species.

Five different endoPGs have been purified from a single commercial pectinase preparation derived from *A. niger* (Kester and Visser 1990). The origin of these PG isoenzymes is not known; one possibility considered is that they are encoded by different genes, analogous to the *A. niger* pectin lyases (Harmsen et al. 1990). On the other hand, the occurrence of multiple forms of secreted enzymes is well documented in fungal biology, and for *A. niger* it has been demonstrated that posttranscriptional events such as differential mRNA splicing (Boel et al. 1984), limited proteolysis (Svensson et al. 1986) and glycosylation (Hayashi and Nakamura 1981) can generate enzyme diversity. The gene encoding the most abundant PG, PGII, has been characterized (Bussink et al. 1990, 1991). Therefore, in order to reveal the molecular basis for the existence of PG isoenzymes and to increase our knowledge about the importance of the individual PGs, we started with the characterization of the second most abundant enzyme, PGI. This PG is quite different from PGII in its physicochemical properties, specific activity and mode of action on oligomeric substrates, whereas the physicochemical properties of three minor enzymes purified by Kester and Visser (1990) are quite similar to those of PGI.

Here we report on partial amino acid sequences of PGI and the cloning and characterization of the corresponding gene. The genes encoding the two most abundant PGs found in commercially available pectinolytic preparations have now been isolated and the cloned genes expressed in *A. niger* strains.

Materials and methods

Strains, library and plasmids. *Aspergillus niger* N400 (CBS 120.49), its derivatives N402 (*cspA1*) and N593 (*cspA1*, *pyrA6*) (Goosen

et al. 1987), plasmid pGW1800 which contains the *pgaII* gene and the transformant *A. niger* N593-pGW1800/27 (Bussink et al. 1990) have all been described. The gene library of *A. niger* N400 in the phage lambda replacement vector EMBL4 (Harmsen et al. 1990) was used for gene isolation. pUC9, phage M13 (Yanisch-Perron et al. 1985) and pEMBL (Dente and Cortese 1987) vectors were used for subcloning. The gene library was plated on *Escherichia coli* LE392 and plasmids and recombinant M13 phages were propagated in *E. coli* DH5 α F' or in *E. coli* JM101.

Characterization of PGI. Purified PGI (Kester and Visser 1990) was fragmented using CNBr or trypsin, the resulting fragments were separated by SDS-polyacrylamide gel electrophoresis and electroblotted onto Immobilon-P (Millipore, Bedford, MA) polyvinylidene difluoride membranes (Matsudaira 1987). The appropriate pieces of membrane were recovered and used for sequence analysis, with a gas phase sequencer equipped with a PTH analyzer as described (Amons 1987). Deglycosylation of denatured PGs with *N*-glycanase was carried out according to the recommendations of the supplier (Genzyme, Cambridge, Mass.).

Manipulation of DNA. Standard methods were used for plasmid DNA isolation, nick-translation, preparation of plaque lifts, Southern analysis and subcloning (Maniatis et al. 1982). Sequencing of recombinant single-stranded M13 phage DNA and double-stranded plasmid DNA was performed using the ³²P Sequencing Kit (Pharmacia LKB, Uppsala, Sweden) according to the recommendations of the supplier, employing a few additional synthetic oligonucleotide primers. The *pgaI* gene was fully sequenced on both strands of DNA. Nucleotide and amino acid sequences were analyzed using the computer programmes of Devereux et al. (1984) and Higgins and Sharp (1989). *Aspergillus* DNA was isolated according to de Graaff et al. (1988). Cotransformation of *A. niger* N593 was carried out as described before (Bussink et al. 1990).

Screening of the genomic library. The oligonucleotide hybridization probe mixture d(ATGGCIGAYGGIGCIGTIATHGAYGGI-GAYGG) (I = inosine; Y = T,C; H = T,C,A) was synthesized on the basis of amino acid sequence data (see Results and discussion). The hybridization stringency for this probe was calculated using the formula of Davis et al. (1986), disregarding the presence of the deoxyinosines (Ohtsuka et al. 1985). Then 40 pmol of the probe mixture were 5'-end-labelled using T₄ polynucleotide kinase and [γ -³²P]ATP (6000 Ci/mmol) and were added to 150 ml of prewarmed (65°C) hybridization mixture which consisted of 1 × SSC, 10 × Denhardt's solution of 0.1% SDS. Nitrocellulose plaque lifts were prehybridized for 5 h at 65°C in the hybridization mixture of which 0.1 mg/ml sheared and denatured herring sperm DNA had been added. The nitrocellulose filters were transferred to the probe mixture, the thermostat of the waterbath was adjusted to 52°C and the hybridization was allowed to proceed for 14 h. The filters were washed in the hybridization mixture at 52°C for 0.5 h, in two changes of 2 × SSC at room temperature for 45 min each and then in 4 × SSC, 0.05% sodium pyrophosphate at 64°C for 0.5 h, followed by a final wash in 2 × SSC at room temperature for 0.5 h. After exposure of the filters to X-ray film, positive plaques were identified on the original plates and the phages were eluted from agar plugs. These phages were replated and screened with the nick-translated 1.2 kbp *Bam*HI-*Bgl*II fragment of pGW1800 as a heterologous probe, employing the hybridization conditions described by Harmsen et al. (1990), with washes to 2 × SSC, 0.1% SDS and 0.1% sodium pyrophosphate at 60°C.

Analysis of PG production. The medium used to induce PGI synthesis contained per litre: 4.2 g urea, 1.5 g KH₂PO₄, 0.5 g KCl, 0.5 g MgSO₄·7H₂O, trace elements (Vishniac and Santer 1957), 10 g green-band pectin (Obipektin, Bischofszell, Switzerland) and 10 g dried and milled sugar beet pulp. In some cases the urea was replaced by 4.0 g/l NH₄Cl. The liquid cultures were inoculated at 10⁶ spores/ml and incubated at 30°C in a New Brunswick orbital shaker at 250–300 rpm. Medium samples were obtained by centrifugation

or by filtration and stored at –20°C until use. PG activities were measured by a reducing sugar assay (Kester and Visser 1990). Western blots were probed with anti-PGI antiserum (Kester and Visser 1990), or in some cases with a PGI-specific monoclonal antibody (Vissen and coworkers, unpublished) and subsequently with an alkaline phosphatase conjugated secondary antibody (Bio-Rad, Richmond, CA).

Results and discussion

PGI is encoded by a second polygalacturonase gene

Partial NH₂-terminal amino acid sequences of PGI and two CNBr fragments thereof were determined (Table 1). The sequence of the 21-kDa CNBr fragment is identical to the NH₂-terminal amino acid sequence of the intact PGI, and in neither case was any amino acid residue detected at position 4. Since the sequencing programme used only detects cysteine residues if the protein is *S*-pyridylethylated before, which is not the case here, it is likely that a cysteine occurs at this position. Although they have a fair degree of homology, the NH₂-terminal amino acid sequence of PGI is clearly different from the corresponding PGII sequence, indicating that PGI and PGII are encoded by different genes. In the case of the NH₂-terminal amino acid sequence of the 5.5 kDa CNBr fragment of PGI, a region of high sequence identity was not found in the amino acid sequence of PGII. The 5.5 kDa fragment is internally located, and its sequence will thus be preceded by a methionine. This sequence was used to design a specific 32-mer oligonucleotide probe mixture of 24 components comprising all possible codons, reducing the number of oligonucleotides in the mixture by introducing deoxyinosine at the third position of fourfold degenerated codons.

The PGI-specific oligonucleotide probe was used to screen a genomic DNA library of *A. niger* N400. Nine positive signals were obtained among the about 14000 plaques analyzed, and these phages were purified in a second screening step, using the 1.2 kbp *Bam*HI-*Bgl*II fragment of the *pgaII* gene as a heterologous probe and employing hybridization conditions of moderate stringency. Restriction analysis of the DNA purified from

Table 1. Amino acid sequences for PGI derived peptides and the N-terminus of mature PGI

Sequence ^a	Residue position ^b	Cleavage method ^c
1. ASTXTFTSASEA	32–43	none
2. ASTXTFTSASE	32–42	CNBr
3. ADGAVIDGDGS	107–117	CNBr
4. NQDDXIXNSXESISFT	205–221	trypsin

^a At positions where amino acids could not be unambiguously assigned, this is indicated by an "X"

^b The positions of the amino acid residues are deduced from the DNA sequence

^c Prior to sequencing, fragments were separated by electrophoresis in 15% SDS-polyacrylamide gels, from which apparent molecular masses of 21 kDa (sequence 2), 5.5 kDa (sequence 3) and 3.5 kDa (sequence 4) were estimated. The 21 kDa CNBr fragment is not a final product of the reaction

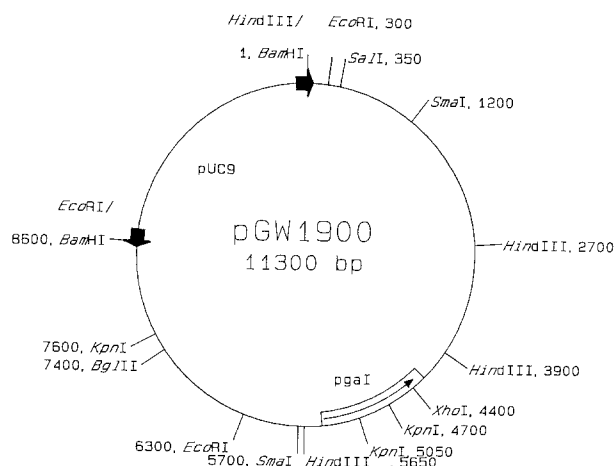


Fig. 1. Restriction map of the *pgaI* gene of *A. niger* N400. Plasmid pGW1900, which is the 8.6 kbp *Bam*HI fragment (thick arrows) containing the *pgaI* gene inserted into pUC9, is shown. Of the restriction sites in the vector, only the *Eco*RI and *Hind*III sites are indicated

four phages indicated that the inserts of these phages are derived from the same region of the *A. niger* genome. Further Southern analyses, using the heterologous *pgaII* probe, suggested that the complete PGI-encoding gene is present on the 8.6-kbp *Bam*HI fragment found in three of the four phages analyzed. This fragment was purified and ligated into the *Bam*HI-digested vector pUC9. The resulting plasmid, pGW1900, was isolated from transformed *E. coli* DH5 α F' cells, and a physical map was constructed (Fig. 1). The nucleotide sequence of the relevant part of this plasmid, from the *Eco*RI site at position 6300 to close to the *Hind*III site at position 3900 of the restriction map of pGW1900, was determined (Fig. 2) and the sequence hybridizing with the oligonucleotide probe was identified. Its complementary sequence, from position 368 to position 399 (Fig. 2), exactly matches the nucleotide sequence of the probe, except for the deoxyinosines present in the latter sequence. The nucleotide sequence also encodes the NH₂-terminal amino acid sequence of the mature PGI, providing additional evidence that the PGI-encoding *pgaI* gene has been cloned. The PGI-encoding nucleotide sequence was designated on the basis of the amino acid sequences determined which identify the open reading frame throughout of gene. The homology of the deduced amino acid sequence with the corresponding PGI sequence provides an additional argument. The nucleotide sequence shown comprises 909 bp of the 5' upstream region, the protein coding region, which is interrupted by two introns, and 368 bp of the 3' nonprotein coding region (see below). The *pgaI* gene encodes a putative protein of 368 amino acids.

Structure of the *pgaI* gene

The *pgaI* promoter sequence contains a putative TATA box (TATAAAA) at position -152 with respect to the translation initiation codon, and a potential CAAT box (CCAAT) is found at position -218 (Fig. 2). A pyrimidine-rich sequence, a frequently observed fungal pro-

moter element that has been shown to be involved in correct initiation of transcription (Punt et al. 1990), is present in the *pgaI* gene from position -98 to -72. The sequence ATCACC immediately preceding the ATG start codon perfectly matches the *Neurospora crassa* consensus sequence ATCAC(C/A) (Legerton and Yanofsky 1985) and closely resembles the consensus sequence for initiation of translation in higher eukaryotes (Kozak 1984).

The protein-coding region is interrupted by two short introns. The distal 62-bp intron B occurs at the same position as the unique 52-bp intron in the *pgaII* gene. The presence and location of this intron is confirmed by NH₂-terminal amino acid sequencing of the 3.5-kDa tryptic peptide (Table 1), which was selected for analysis on the basis of the tryptic fragment expected from the deduced amino acid sequence. The 5' splice site of this intron, GCACGA, does not match the fungal 5' splice consensus GTANGT (Gurr et al. 1987). In particular, the cytosine at the second position is an exception to the GT boundary rule, which has been previously observed in a few *Neurospora* (Gurr et al. 1987) and some non-fungal (Shapiro and Senapathy 1987) introns. On the other hand, the sequences CGCTAAC and CAG are in complete agreement with the fungal intron internal consensus sequence PyGCTAAC and the 3' splice consensus PyAG, respectively (Gurr et al. 1987). The 52-bp intron A changes the reading frame and contains translational stop codons in both reading frames identified by amino acid sequencing. The presence and location of this intron, which is not present in the *pgaII* gene, is supported by a comparison of the deduced *pgaI* and *pgaII* gene products, and in this intron at the appropriate positions sequences are found that are in complete agreement with the fungal intron consensus sequences.

The *pgaI* and *pgaII* genes have significantly diverged, showing 66% nucleotide sequence identity in the regions coding for the mature proteins and 60% at the amino acid level and only a low sequence identity in the non-protein coding sequences. The codon usage in the two genes is different, with a higher bias observed in the *pgaI* gene (Table 2). In this gene only 38 sense codons are used, as against 51 in the *pgaII* gene, and there is a stronger tendency to avoid A and a higher preference for C at the third nucleotide position. Fungal genes, which are expressed at low level, show much less codon bias than highly expressed genes (Bennetzen and Hall 1982; May et al. 1987; Gurr et al. 1987), but in contrast to this observation PGI is considered to be the most abundant PG secreted by *A. niger* (Kester and Visser 1990; Keon and Waksman 1990). It is, however, conceivable that PGI is the more abundant enzyme under natural growth conditions, as opposed to those applied in the laboratory or in industry (see below).

The deduced primary structure of PGI

The deduced amino acid sequence of PGI has an NH₂-terminal extension of 31 amino acids before the start of the mature protein. It is very unlikely that processing by a signal peptidase occurs after Lys 31 and a probable

GAATTCGGAAGAATGCTCGCTTCGTCGCGCACACACGCTTCCTAAAATTTATCCGATGATCAACCGAG -842

GAACACGGGTCGAACCCCTGAAAGGAATCTCGCCGAAAGGATCAATCATATTGATTCGCGGACCCCTGTAAGAGTCCCTAAAGGGTCTTTGGCAGAGTCTCGATAGTGGGATCGGCACCTGG -721

CAGTTTCTAGTCTCGTCAGTGCAGCCATTCCGACCGATCCGAGTGTGCTGGTGGGTCAGAGCAGTCTATATCGATATCATCACCAAAATCCCAAGGACCGTGGGATGACGAGTATGCAT -601

TGGCACTGAGAATTGCCGAGATCATGCCCCCTCTGGAGTGCATCCATCAAGCATGTTGGCAGAGTGAACACGCGTGGATAAGCCCGATGGGGCAGATGGGGGAAGAGACCGGAGAC -481

ACGGCATCAGCAGTGGAGATGCCGAGAAGTGCAGAGCCGCGGATGCTTCGACCTTCAGCCCAAGCGTCAATCCATGCCTTCTTGGGTACAGGGTGGCCGAGCTGAAGCCCGTTAGCG -361

TCTGAGATACCCGGATGACACGAACCTTGGTGTCTACCAATGCTGTGATGCATGCAGTGCAGGCGGATCCCTGCCTGAAGCGCCCAICGCGCTTAGAGATTGATACCCGGTGGATCG -241

GAGGGTCCCCCTTGGTCTTTCCACCAATAATGGAGGTATCTCACTGCTTTGTTTCAGGAACAGAAGCTTAGCATGACCAGTCTTTTCAGCTATAAACTCTCCAGGACTTCCGCTGTTGAGA -121

TGCTCTATCCAACAACACCTCCTCTTCCACTCTTGTCTTTCTTTCTGTTGACCAGACCCCATCCATTTCTTTGGCCGAAACCTGTTCTGTTGACCAGAACCTATACCAACAATCACC -1

ATG CAC TCT TAC CAG CTT CTT GGC CTG GCC GCT GTC GGC TCC CTC GTC TCT GCC GCC CCC GCT CCT TCT CGC GTC TCC GAG TTC GCT AAG 90
M H S Y Q L L G L A A V G S L V S A A P A P S R V S E F A K 30

AAG GCC TCT ACC TGC ACC TTC ACC TCT GCC TCT GAG GCC AGC GAG AGC ATC TCC AGC TGC TCC GAT GTT GTC CTG AGC AGC ATC GAG GTC 180
K * A S T C T F T S A S E A S E S I S S C S D V V L S S I E V 60

CCC GCT GGC GAG ACC CTC GAC CTG TCC GAT GCT GCT GAT GGC TCC ACC GTATGTGCTCTCAGCCGCTTCCATCCTGGCTATCAGCTAACACCATCTAG 280
P A G E T L D L S D A A D G S T intron A 76

ATC ACC TTC GAG GGC ACC ACT TCC TTC GGA TAC AAG GAA TGG AAG GGC CCC CTG ATC CGC TTC GGT GGT AAG GAT CTG ACT GTC ACC ATG 370
I T F E G T T S F G Y K E W K G P L I R F G G K D L T V T M 106

GCC GAC GGC GCT GTC ATC GAC GGT GAC GGT TCC CGC TGG TGG GAC AGC AAG GGT ACC AAC GGT GGC AAG ACC AAG CCC AAG TTC ATG TAC 460
A D G A V I D G D G S R W W D S K G T N G G K T K P K F M Y 136

ATC CAC GAT GTT GAG GAC TCG ACC TTC AAG GGC ATC AAC ATC AAG AAC ACT CCC GTC CAG GCC ATC AGT GTC CAG GCT ACC AAC GTC CAC 550
I H D V E D S T F K G I N I K N T P V Q A I S V Q A T N V H 166

CTG AAC GAC TTC ACC ATC GAC AAC TCC GAC GGT GAT GAC AAC GGT GGC CAC AAC ACC GAC GGT TTC GAC ATC AGC GAG TCT ACC GGT GTC 640
L N D F T I D N S D G D D N G G H N T D G F D I S E S T G V 196

TAC ATC AGC GGT GCT ACC GTC AAG AAC CAG GAC GAC TGC ATT GCC ATC AAC TCT GGC GAG GCACGATATCCCTATCCACTATCATCTCCATTCAT 739
Y I S G A T V K N Q D D C I A I N S G E intron B 216

ATCGCTAACAAATCAAAACCCACAG AGC ATC TCT TTC ACC GGC GGT ACC TGC TCC GGT GGC CAC GGT CTC TCC ATC GGC TCT GTC GGT GGC GGT GAT 834
S I S F T G G T C S G G H G L S I G S V G G R D 240

GAC AAC ACC GTC AAG AAC GTG ACC ATC TCC GAC TCC ACT GTC AGC AAC TCC GCC AAC GGT GTC CGC ATC AAG ACC ATC TAC AAG GAG ACC 924
D N T V K N V T I S D S T V S N S A N G V R I K T I Y K E T 270

GGT GAT GTC AGC GAG ATC ACC TAC TCT AAC ATC CAG CTC TCC GGA ATC ACC GAC TAC GGT ATC GTC ATC GAG CAG GAC TAC GAG AAC GGC 1014
G D V S E I T Y S N I Q L S G I T D Y G I V I E Q D Y E N G 300

TCT CCC ACC GGC ACC CCC TCC ACC GGT ATC CCC ATC ACT GAT GTC ACC GTT GAC GGT GTC ACC GGT ACT CTC GAG GAT GAC GCC ACC CAG 1104
S P T G T P S T G I P I T D V T V D G V T G T L E D D A T Q 330

GTC TAC ATT CTC TGC GGT GAC GGC TCT TGC TCT GAC TGG ACC TGG TCC GGT GTT GAC CTC TCT GGT GGC AAG ACC AGC GAT AAA TGC GAG 1194
V Y I L C G D G S C S D W T W S G V D L S G G K T S D K C E 360

AAC GTT CCT TCC GGT GCT TCT TGC TAA ATCGTTCTCCGGATGCGAGGCAACGCTCTGTAGGACGTCGGTGTATATATCGATCCACACTTGACATGTACTAGTTAGGTG 1304
N V P S G A S C --- 368

GTTTACTGCCATAGTGTAGTGAATAATAGGAGCTTTGCTCAATATGTGAATTTGTAGCAGAAGTAAATGAGTAGACTGATAGAGGTAATGATGAAAGATAGAATTTAATACCAGGTG 1423

ATGAAGTTAATAACGGGTGAATTAGGGCGCTAGGCTTAGGTATATAAGTTGTCATCCCTCACACTCGAAATCTCTTCTCTTTCTATATCTTCAACCTCTGAACCACCGTGGC 1542

TCCACAGACTAACAAGATTCTCTATATCGATGCTTGATCTCAA 1586

Fig. 2. Nucleotide and deduced amino acid sequences of the *A. niger* N400 *pgaI* gene. The start of the mature PGI (*) and potential *N*-glycosylation sites (#) are indicated

cleavage site is after Ala-18 (von Heijne 1986). We therefore hypothesize that the NH₂-terminal extension of the PGI precursor comprises a signal peptide as well as a pro-peptide with a dibasic cleavage site as is also found, for example, in the precursors of the *A. awamori* glucoamylase (Innes et al. 1985) and the *Phanerochaete chrysosporium* ligninases (Tien and Tu 1987; de Boer et al. 1987). The prepro sequence of the PGI precursor does not have significant sequence homology with the corresponding PGII sequence, and this result is unexpected, since the putative PGII pro-peptide shows homology to some other fungal pro-peptides with a monobasic cleavage site found in precursors of otherwise unrelated proteins (Bussink et al. 1991).

The molecular mass of mature PGI calculated from the deduced amino acid sequence is 35 kDa, which is significantly lower than the previously reported values (Kester and Visser 1990), namely 55 kDa and 47 kDa as determined by polyacrylamide gel electrophoresis in the presence of SDS and by gel permeation chromatography, respectively. The deduced amino acid sequence of PGI has two potential *N*-glycosylation sites (Asn-X-Thr/Ser), one of which (Asn246-Val-Thr248) corresponds to the unique *N*-glycosylation site in PGII. Upon deglycosylation with *N*-glycanase the apparent molecular mass of PGI decreased by about 2-3 kDa and that of PGII by 2 kDa, from 38 kDa to 36 kDa, as determined by electrophoresis in 10% polyacrylamide gels in the presence of

```

PGI 32 ASTCTFTSASEASESISSSCDVVLSSIEVPAGETLDLSDAADGSTITFEGTTSFGYKEWKGPLIRFGGKDLVTMADGAVIIGDGSRWWSKGTNGGKTKPKFMYIHVDVST 144
PGII 28 DS.CFTTAAAAKAGKAKCSTITLNVIEVPAGTTLDTGLTSGTKVIFEGTTTFQYEWAGPLISMSGEHITVTGASGHLINCDGARWWDGKGT.SGKKPKFFYAHGLDSSS 138

PGI 145 FKGINIKNTPVQAISSVQATNVHLNDFITDNSDGDNGGHNTDGFDISETSGVYISGATVKNQDDCIAINSGESISFTGGTCSGGHLSIGSVGGDDNTVKNVTISDSTVSNS 257
PGII 139 ITGLNLIKNTPLMAFSVQANDITFTDVTINNADGDTQGGHNTDAFDVGNVSGVNIKIPVWHNQDDCLAVNSGENIWF TGGTICGGHLSIGSVGDRSNNVKNVTIEHSTVSNS 251
<----->

PGI 258 ANGVRIKTIYKETGDVSEIITYSNLQSGITDYGVIEQDYENGSTGTPSTGIPITDVTVDGVTGLLEDDATQVYILCGDGSQSDWTWSGVOLSGGKTSKDCENVPASGASC 368
PGII 252 ENAVRIKTIISGATGVSVEIITYSNIVMSGISDYGVVIQQDYEDGKPTGKPTNGVTIQDVKLESVTSVDSGATEIYLLGSGSQCSDWTWDDVKVTGGKSTACKNFPASVASC 362
<----->

```

Fig. 3. Comparison of the deduced amino acid sequences of PGI and PGII. Identical amino acids (:) and the region which is relatively well conserved amongst PGs of plant, fungal and bacterial origin (<----->) are indicated

Table 2. Codon utilization in the *pgaI* and *pgaII* genes of *A. niger* N400

		<i>pgaI</i>		<i>pgaII</i>		<i>pgaI</i>		<i>pgaII</i>		<i>pgaI</i>		<i>pgaII</i>			
UUU	Phe	0	4	UCU	Ser	16	11	UAU	Tyr	0	1	UGU	Cys	0	2
UUC	Phe	10	8	UCC	Ser	18	12	UAC	Tyr	9	6	UGC	Cys	8	7
UUA	Leu	0	1	UCA	Ser	0	0	UAA	----	1	0	UGA	----	0	0
UUG	Leu	0	1	UCG	Ser	1	5	UAG	----	0	1	UGG	Trp	5	7
CUU	Leu	2	7	CCU	Pro	2	4	CAU	His	0	3	CGU	Arg	1	0
CUC	Leu	7	5	CCC	Pro	8	3	CAC	His	5	5	CGC	Arg	4	2
CUA	Leu	0	0	CCA	Pro	0	1	CAA	Gln	0	0	CGA	Arg	0	2
CUG	Leu	6	3	CCG	Pro	0	1	CAG	Gln	7	7	CGG	Arg	0	0
AUU	Ile	2	8	ACU	Thr	6	8	AAU	Asn	8	6	AGU	Ser	1	3
AUC	Ile	25	17	ACC	Thr	32	25	AAC	Asn	17	16	AGC	Ser	12	7
AUA	Ile	0	0	ACA	Thr	0	0	AAA	Lys	1	4	AGA	Arg	0	0
AUG	Met	3	4	ACG	Thr	0	6	AAG	Lys	16	16	AGG	Arg	0	0
GUU	Val	5	6	GCU	Ala	10	9	GAU	Asp	11	12	GGU	Gly	25	10
CUC	Val	21	14	GCC	Ala	11	10	GAC	Asp	23	10	GGC	Gly	19	23
GUA	Val	0	0	GCA	Ala	0	1	GAA	Glu	1	8	GGA	Gly	2	5
GUG	Val	1	8	GCG	Ala	0	8	GAG	Glu	15	5	GGG	Gly	0	5

SDS (not shown). This indicates that PGI and PGII are *N*-glycosylated and that the cleaved *N*-linked glycans are small, probably having a structure similar to those found in e.g. Taka-amylase A of *A. oryzae* (Yamaguchi et al. 1971) and cellobiohydrolase I of *Trichoderma reesei* (Salovuori et al. 1987). Therefore, it is unlikely that the discrepancy between the calculated molecular mass of PGI and the values obtained by experiment is due to *N*-glycosylation.

The highest sequence identity in PGI and PGII (Fig. 3) is found in the region that is relatively well conserved amongst PGII, a tomato PG and bacterial PGs and contains possible candidates for the essential amino acids involved in the hydrolysis reaction that these enzymes catalyze (Bussink et al. 1991). These candidates are also found in PGI, namely His(229) and Asp(186), Asp(207) and Asp(208). The deduced mature PGI and PGII amino acid sequences differ by two amino acids in length. PGI contains one additional amino acid in the NH₂-terminal sequence and a second single amino acid insertion with respect to the PGII sequence is at position 126, or possibly at position 127, in a region where there is otherwise a high degree of sequence identity. Keon and Waksman (1990) have identified a highly homologous amino acid domain in the NH₂-terminal amino acid sequences of the major PG produced by *Colletotrichum lindemuthianum*, *Sclerotinia sclerotiorum* and *A. niger*, the last of which probably corresponds to PGII. This

domain stands out in the alignment of the PGI and PGII sequences, from Ile(58) to Leu(68) in PGI. However, only 10 out of the 17 amino acids conserved amongst the NH₂-terminal sequences of the three fungal major PGs are found in PGI, which further demonstrates the divergence between the *pgaI* and *pgaII* genes.

Expression of the *pgaI* gene

Plasmid pGW1900 was used to cotransform *A. niger* N593, a uridine auxotrophic mutant of *A. niger* N402, employing a plasmid containing the functional *pyrA* gene as the selective marker. The transformants were grown in a medium with both 1% sugar beet pulp and 1% pectin as PG-inducing carbon sources and with urea as the nitrogen source. The amount of PGI in the culture filtrates was assayed by Western blotting using a polyclonal antibody raised against PGI (Fig. 4). Amongst the 23 transformants tested, 65% produced significantly more PGI than *A. niger* N402 and a *pyrA* transformant of *A. niger* N593. The high producing transformants, e.g. *A. niger* N593-pGW1900/3, produced about tenfold higher PG activity than the control strains. The cloned *pgaI* gene is therefore a functional gene. Southern blot analysis of genomic *A. niger* N593-pGW1900/3 DNA restricted with *Bg/III*, which cuts pGW1900 once, showed a strongly hybridizing band of about 11 kbp, indicating tandem integration

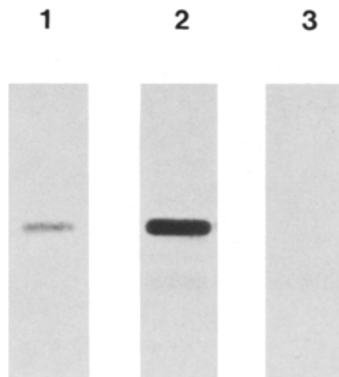


Fig. 4. Western blot analysis of PGI levels. Culture media were obtained from *A. niger* N402 (lane 1) and the transformant *A. niger* N593-pGW1900/3 (lanes 2 and 3) grown for 48 h on 1% pectin and 1% sugar beet pulp as carbon source and with urea (lanes 1 and 2) or ammonium chloride (lane 3) as nitrogen source

of multiple copies of pGW1900 in the *A. niger* genome (not shown).

A. niger N402 produces little PGI activity on a medium optimized for the production of PGII, viz. with 1% pectin and 1% sugar beet pulp as carbon source and NH_4Cl as nitrogen source. This is indicated by Western blot analysis using an antibody raised against PGI and by isoelectric focusing in combination with activity staining (not shown). However, when NH_4Cl was replaced by urea a PG was synthesized which comigrated with purified PGI upon SDS-polyacrylamide gel electrophoresis, whereas the PGII production became reduced with urea as nitrogen source. In order to further demonstrate that the PGI production is enhanced by using urea, the transformant *A. niger* N593-pGW1900/3 was grown both on NH_4Cl and on urea. The transformant was indeed found to synthesize much less PGI on NH_4Cl as compared to the medium with urea (Fig. 4). On the other hand, the PGII overproducing transformant *A. niger* N593-pGW1800/27 synthesizes more PGII with NH_4Cl as the nitrogen source, which is in part due to proteolytic degradation in the medium containing urea (data not shown).

Conclusions

The *A. niger* endoPGs are encoded by at least two diverged genes, in contrast to the fungal maize pathogen *Cochliobolus carbonum* in which case there is evidence for a single endoPG gene encoding a single PG (Scott-Craig et al. 1990; Walton and Cervone 1990) and this may reflect differences between saprophytic fungi and specialized phytopathogenic fungi. However, other phytopathogenic fungi produce sometimes complex PG isoenzyme patterns which are influenced by the substrate. For example, differences have been found between the isoenzymes produced by *Sclerotinia sclerotiorum* in culture and in infected tissues (Marciano et al. 1982) and when grown in vitro *Verticillium albo-atrum* produces a single PG isoenzyme on glucose but numerous PGs on tomato plant cell walls (Durrands and Cooper 1988). *A. niger* synthesizes PGI and PGII in substantial amounts only when it is

grown on pectin or pectin containing substances and the PG isoenzyme pattern depends on the inducing carbon source used (Bussink et al. 1991), whereas the effect of urea observed in the present study shows that medium constituents which are not derived from pectin can strongly influence the levels of individual PGs. The regulation of PG synthesis in *Aspergilli* has been previously studied (e.g. Shinmyo et al. 1978; Dean and Timberlake 1989a), and there are numerous reports on the optimization of PG production (e.g. Hermersdörfer et al. 1987; Friedrich et al. 1990). However, with respect to PG gene expression little attention has been paid to the occurrence of isoenzymes. Since two *A. niger* PG genes have now been isolated, it is possible to address specific questions regarding their individual expression and regarding the molecular basis for the differences in enzymic properties. Therefore, we are currently characterizing PGI and PGII in detail.

Acknowledgements. We are grateful for the contribution made by Mr. B. Fraaye. This work was part of a joint programme between the Biotechnology Department of Ciba-Geigy, Basle (Switzerland) and the Agricultural University of Wageningen. The nucleotide sequence reported in this paper will appear in the EMBL, GenBank and DDBJ Nucleotide Sequence Databases under the accession number X58892.

References

- Amons R (1987) FEBS Lett 212: 68–72
- Bennetzen JL, Hall BD (1982) J Biol Chem 257: 3026–3031
- Boel E, Hjort I, Svensson B, Norris F, Norris KE, Fiil NP (1984) EMBO J 3: 1097–1102
- Boer HA de, Zhang YZ, Adinarayana Reddy C (1987) Gene 60: 93–102
- Bussink HJD, Kester HCM, Visser J (1990) FEBS Lett 273: 127–130
- Bussink HJD, Buxton FP, Visser J (1991) Curr Genet 19: 467–474
- Cervone F, De Lorenzo G, Salvi G, Camardella L (1986) In: Baily J (ed) Molecular biology of plant-pathogen interactions (NATO ASI series, vol H1). Springer, Berlin Heidelberg New York, pp 385–392
- Davis GD, Dibner MD, Battey JF (1986) Basic methods in molecular biology. Elsevier, Amsterdam
- Dean RA, Timberlake WE (1989a) The Plant Cell 1: 265–273
- Dean RA, Timberlake WE (1989b) The Plant Cell 1: 275–284
- Dente L, Cortese R (1987) Methods Enzymol 155: 111–119
- Devereux J, Haeberli P, Smithies O (1984) Nucleic Acids Res 12: 387–395
- Durrands PK, Cooper RM (1988) Physiol Mol Plant Pathol 32: 343–362
- Friedrich J, Cimerman A, Steiner W (1990) Appl Microbiol Biotechnol 33: 377–381
- Goosen T, Bloemheuvel G, Gysler C, de Bie DA, van den Broek HWJ, Swart K (1987) Curr Genet 11: 499–503
- Graaff L de, Broek H van den, Visser J (1988) Curr Genet 13: 315–321
- Gurr SJ, Unkles SE, Kinghorn JR (1987) The structure and organization of nuclear genes of filamentous fungi. In: Kinghorn JR (ed) Gene structure in eukaryotic microbes. IRL Press, Oxford, pp 93–139
- Gysler C, Harmsen JAM, Kester HCM, Visser J, Heim J (1990) Gene 89: 101–108
- Harmsen JAM, Kusters-van Someren MA, Visser J (1990) Curr Genet 18: 161–166

- Hayashi S, Nakamura S (1981) *Biochim Biophys Acta* 657:40–51
- Heijne G von (1986) *Nucleic Acids Res* 14:4683–4690
- Hermersdörfer H, Leuchtenberger A, Wardsack C, Ruttloff H (1987) *J Basic Microbiol* 27:309–315
- Higgins DG, Sharp PM (1989) *CABIOS* 5:151–153
- Innes MA, Holland MJ, McGabe PC, Cole GE, Wittman VP, Tal R, Watt KWK, Gelfand DH, Holland JP, Meade JH (1985) *Science* 228:21–26
- Keon JPR, Waksman G (1990) *Appl Environ Microbiol* 56:2522–2528
- Kester HCM, Visser J (1990) *Biotechnol Appl Biochem* 12:150–160
- Khanh NQ, Albrecht H, Ruttkowski E, Löffler F, Gottschalk M, Jany KD (1990) *Nucleic Acids Res* 18:4262
- Kozak M (1984) *Nucleic Acids Res* 12:857–872
- Legerton TL, Yanofsky C (1985) *Gene* 39:129–140
- Maniatis T, Fritsch EF, Sambrook J (1982) *Molecular cloning, a laboratory manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
- Marciano P, Di Lenna P, Magro P (1982) *Physiol Plant Pathol* 20:201–212
- Matsudaira P (1987) *J Biol Chem* 262:10035–10038
- May GS, Tsang MLS, Smith H, Fidel S, Morris NR (1987) *Gene* 55:231–243
- Ohtsuka E, Matsuki S, Ikehara M, Takahashi Y, Matsubara K (1985) *J Biol Chem* 260:2605–2608
- Punt PJ, Dingemans MA, Kuyvenhoven A, Soede RDM, Pouwels PH, van den Hondel CAMJJ (1990) *Gene* 93:101–109
- Rexová-Benková L, Markovic O (1976) *Adv Carbohydr Chem Biochem* 33:323–385
- Rombouts FM, Pilnik W (1980) In: Rose AH (ed) *Microbial enzymes and bioconversions*. *Econ Microbiol* 5:227–282
- Ruttkowski E, Labitzke R, Khanh NQ, Löffler F, Gottschalk M, Jany K-D (1990) *Biochim Biophys Acta* 1087:104–106
- Salovuori I, Makarow M, Rauvala H, Knowles J, Kääriäinen L (1987) *BIO/TECHNOLOGY* 5:152–156
- Scott-Craig JS, Panaccione DG, Cervone F, Walton JD (1990) *The Plant Cell* 2:1191–1200
- Shapiro MB, Senapathy P (1987) *Nucleic Acids Res* 15:7155–7174
- Shinmyo A, Davis IK, Nomoto F, Tahara T, Enatsu T (1978) *Eur J Appl Microbiol Biotechnol* 5:59–68
- Svensson B, Larsen K, Gunnarsson A (1986) *Eur J Biochem* 154:497–502
- Tien M, Tu CPD (1987) *Nature* 326:520–523
- Vries JA de, Voragen AGJ, Rombouts FM, Pilnik W (1986) In: Fishman ML, Jen JJ (eds) *Chemistry and function of pectins* (ACS Symposium series, vol 310). American Chemical Society, Washington DC, pp 38–48
- Vishniac S, Santer M (1957) *Bacteriol Rev* 21:195–213
- Walton JD, Cervone F (1990) *Physiol Mol Plant Pathol* 36:351–359
- Yamaguchi H, Ikenanka T, Matsushima Y (1971) *J Biochem* 70:587–594
- Yanisch-Perron C, Vieira J, Messing J (1985) *Gene* 33:103–119

Communicated by A. Hinnen