Curr Genet (1991) 19:467–474

Current Genetics

Expression and sequence comparison of the *Aspergillus niger* and *Aspergillus tubigensis* genes encoding polygalacturonase II *

Hendrik J. D. Bussink [1], Frank P. Buxton [2], and Jaap Visser [1]

[1] Department of Genetics, Agricultural University, NL-6703 HA Wageningen, The Netherlands
[2] Ciba-Geigy AG, Biotechnology Department, CH-4002 Basel, Switzerland

Received December 8, 1990/March 4, 1991

Summary. The structure and expression of the polygalacturonase-encoding *pgaII* genes of two recently recognized species, *Aspergillus niger* and *Aspergillus tubigensis,* was investigated. While the structure of the *pgaII* genes is very similar, showing 83% DNA sequence identity and 94% identity at the amino acid level, they have diverged significantly. The NH_2-terminal sequence suggests that these PGs are made as pre pro-proteins and the secretory propeptide of the PGII precursors shows sequence homology with some other fungal pro-peptides. The expression of the *pgaII* genes is strongly regulated by the carbon source and the *A. tubigensis* gene is expressed and regulated in *A. niger* transformants. The low similarity of the fungal PGs with those of bacterial and plant origin is discussed in relation to the possible functional role of specific amino acids.

Key words: Filamentous fungus – Pectin – Taxonomy – Precursor processing

Introduction

Saprophytic and plant pathogenic fungi and bacteria produce an array of enzymes capable of degrading the complex polysaccharides of the plant cell wall (Cooper 1984; Misaghi 1982); these include the pectic enzymes involved in the degradation of pectin, the main constituent of the middle lamella of higher plant cells. Pectin is a heteropolysaccharide consisting of a linear chain of D-galacturonic acid residues which are partially esterified and connected by α-1-4 glycosidic bonds. L-rhamnose residues are found in the main chain and other neutral sugars like L-arabinose, D-galactose and D-xylose compose the side chains. The microbial degradation of pectin is a complex process; e.g., *Aspergillus niger* synthesizes endopolygalacturonase (endo-PG), exo-PG, endopectin lyase and pectinesterase (Rexová-Benková and Markovic 1976; Rombouts and Pilnik 1980). As opposed to fungal cellulases, which have been extensively studied (Knowles et al. 1987), the other fungal plant cell wall-degrading enzymes are less well characterized at the molecular level. The development of the molecular biology of the pectic enzymes is important in order to increase our knowledge about their specific roles in the degradation of plant cell walls, to develop applications in the processing of agricultural products (Rombouts and Pilnik 1980) and, in the case of phytopathogenic fungi, to understand host-parasite interactions (Collmer and Keen 1986).

The pectate lyase gene of *Aspergillus nidulans* (Dean and Timberlake 1989) and an *A. niger* pectin esterase cDNA (Khanh et al. 1990) have already been cloned but we are particularly interested in pectin lyases (EC 4.2.2.3) and polygalacturonases [poly(1,4-α-D-galacturonide)glycanohydrolase, EC 3.2.1.15] and wish to use the corresponding genes to study fungal gene structure and expression. The molecular cloning and characterization of the *A. niger* pectin lyase gene family (Gysler et al. 1990; Harmsen et al. 1990), the isolation and characterization of five *A. niger* polygalacturonases (Kester and Visser 1990) and the molecular cloning of the polygalacturonaseII (*pgaII*) gene of *A. niger* (Bussink et al. 1990) have recently been described. In the present paper we further characterize the structure and expression of the *pgaII* gene.

A. niger is of considerable economic importance, and is employed for the production of organic acids and extracellular enzymes to be used in the food industry. However, the taxonomy of the black Aspergilli has been difficult and not always reliable. The classification of the black Aspergilli has been recently reinterpreted on the basis of restriction fragment length polymorphisms, and it has been proposed that the *A. niger* aggregate consists of two distinct species, *A. niger* and *A. tubigensis* (Kusters-van Someren et al. 1991). To further demonstrate the divergence between these species, and its consequences for plant cell wall-degrading enzymes, we isolated and characterized the *pgaII* gene of an *A. tubigensis* strain and compared it to that of *A. niger*.

* The nucleotide sequence reported in this paper will appear in the EMBL, GenBank and DDBJ Nucleotide Sequence Databases under the accession numbers X58893 (*A. niger*) and X58894 (*A. tubigensis*)

Offprint requests to: J. Visser

Materials and methods

Strains, library and plasmids. *Aspergillus niger* N400 (CBS 120.49), and its derivatives N402 (*cspA*) and N593 (*cspA*, *pyrA*), have been described (Goosen et al. 1987). *Aspergillus tubigensis* NW756 is identical to the isolate previously designated *A. niger* NW756 (Kusters-van Someren et al. 1991). *A. niger* N593-pGW1800/27, which is identical to the transformant T27, and plasmid pGW1800 have also been described (Bussink et al. 1990). The gene library of *A. tubigensis* NW756 was constructed by ligating partially *Sau*3AI-digested genomic DNA fragments into the lambda replacement vector EMBL3 cut with *Bam*HI (Flipphi et al., unpublished). Phage M13 (Yanisch-Perron et al. 1985), pEMBL (Dente and Cortese 1987) and pTZ18R (Pharmacia LKB Biotechnology AB, Uppsala, Sweden) vectors were used for subcloning. Recombinant lambda phages were plated on *E. coli* LE392 and plasmids and recombinant M13 phages were propagated in *E. coli* JM109 (Yanisch-Perron et al. 1985) or *E. coli* DH5αF′ (BRL, Life Technologies Inc., Gaithersburg, USA).

Growth conditions and analysis of polygalacturonases. The media used to study PG induction and isoenzyme profiles contained per litre: 6 g $NaNO_3$, 1.5 g KH_2PO_4, 0.5 g KCl, 0.5 g $MgSO_4 \cdot 7H_2O$, plus traces of $MnCl_2$, $ZnSO_4$, and $FeSO_4$, and 10 g of the carbon source indicated. Dried sugar beet pulp was ground in a Waring blender before use. The brown-band pectin was a kind gift of Dr. Rombouts. For the other experiments a slightly different medium was used which contains NH_4Cl as the nitrogen source and both sugar beet pulp (10 g/l) and pectin (10 g/l; d.e. 61.2%, Obipektin, Bischofszell, Switzerland) as the carbon sources (Bussink et al. 1990). The liquid cultures were inoculated at 10^6 spores/ml and incubated at 30 °C in a New Brunswick orbital shaker at 250–300 rpm. Purification of PGII and further protein analyses were as described (Kester and Visser 1990; references therein). PG zymograms were made by isoelectric focusing of culture filtrates (10 µl), followed by overlaying the focusing gel with a polygalacturonic acid-containing agarose gel, incubating it for about 1.5 h and staining it with ruthenium red (Ried and Collmer 1985).

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 ^{T7}Sequencing Kit (Pharmacia LKB) or the Sequenase DNA Sequencing Kit (United States Biochemical Corporation Cleveland, USA), employing additional synthetic oligonucleotide primers. Nucleotide sequences were confirmed from 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). Co-transformation of *A. niger* N593 was carried out as previously described (Bussink et al. 1990).

Northern blot analysis. Total RNA was isolated essentially as described (Cathala et al. 1983). Mycelium was ground in a mortar to a fine powder under liquid nitrogen. The powder was added to preheated (50 °C) extraction buffer to which β-mercaptoethanol had been freshly added. The mixture was immediately vigorously agitated and further treated as described. Equal amounts of total RNA, as judged from the intensity of ribosomal RNA bands in ethidium bromide-stained agarose gels, were glyoxylated and electrophoresed in a 1.3% agarose gel (Thomas 1983). The RNA was blotted onto GeneBind 45 membrane (Pharmacia LKB) using 10 × SSC as the transfer buffer. After boiling, to remove the glyoxal (Thomas 1983), the membrane was hybridized at 65 °C using the SDS buffer described by Church and Gilbert (1984). Unbound probe was removed by washing at 65 °C in a series of SSC buffers with 0.1% sodium pyrophosphate and 0.1% SDS, with a final wash in 0.2 × SSC.

Results and discussion

Polygalacturonase isoenzyme patterns

Five different endo-polygalacturonases have been isolated from a single commercial enzyme preparation derived from *A. niger*, PGI and PGII being the major activities (Kester and Visser 1990). The PG activities in the culture filtrates of *A. niger* N400 and *A. tubigensis* NW756 were, therefore, also analyzed, using isoelectric focusing in combination with activity staining (Ried and Collmer 1985). When *A. niger* N400 and *A. tubigensis* NW756 were cultivated in a minimal medium with simple sugars, such as sucrose (Fig. 1), glucose, xylose or galacturonic acid, as the carbon source, they produced only very little, or no, PG activity. However, when pectin or pectin-containing substances, such as sugar beet pulp, were used as the carbon source, PG production was induced (Fig. 1). In both strains the PG activity can be attributed to at least two distinct endo-PGs, with isoelectric points of between 5 and 6, and about 3.5, respectively. The activity with the lower isoelectric point probably corresponds to PGI, since this was also the most abundant low-isoelectric point PG in the commercial enzyme preparation. The PGs with the higher isoelectric points were partially purified from the culture filtrates of *A. tubigensis* NW756 and *A. niger* N400 and analyzed by Western blotting using an antibody raised against PGII (Kester and Visser 1990). Both enzymes react with this antibody and have an apparent molecular mass identical to that of PGII, i.e., 38 kDa (data not shown), which is much lower than the apparent molecular mass observed for other PGs. However, these two enzymes are not identical, since the isoelectric point of the *A. tubigensis* NW756 PGII, about 6.0, is slightly higher than that of the *A. niger* N400 PGII, 5.2.

The PG isoenzyme pattern depends both on the strain analyzed and the inducing carbon source used. For instance *A. tubigensis* NW756 always produced more PGII than *A. niger* N400 (Fig. 1) and PGII is the major activity when sugar beet pulp is used as the carbon source, where-

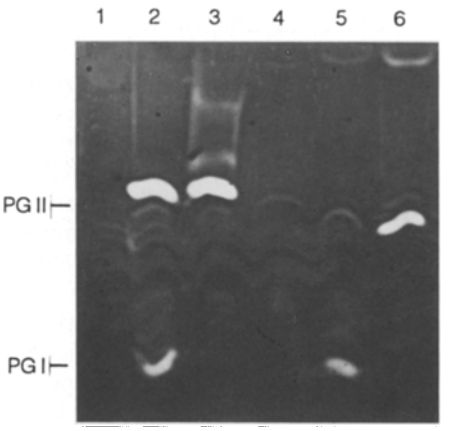

Fig. 1. PG isoenzyme profiles produced by *A. tubigensis* NW756 (*lanes 1–3*) and *A. niger* N400 (*lanes 4–6*) on sucrose (*lanes 1 and 4*), a brown-band pectin (*lanes 2 and 5*) and sugar beet pulp (*lanes 3 and 6*). The isoenzymes were separated by isolectric focusing (pH 4–9) and visualized by activity staining

as brown-band pectin favours PGI production. In other experiments we have used a commercially available pectin (Obipektin) as the sole carbon source; on this *A. niger* N400 secreted similar amounts of PGI and PGII activity, whereas *A. tubigensis* NW756 produced significantly more PGII than PGI activity.

Phytopathogenic fungi are also known to synthesize multiple forms of PG (Cervone et al. 1986), whose expression may be differentially regulated (Durrands and Cooper 1988). In order to reveal the molecular basis for the existence of the PG isoenzymes we are continuing with the characterization of PGI (to be published elsewhere).

Molecular cloning of the A. tubigensis NW756 PGII gene

The genomic DNA of both *A. niger* N400 and *A. tubigensis* NW756 was analyzed by Southern blotting using the 1.2 kbp *Bam*HI-*Bgl*II fragment of plasmid pGW1800 as a probe (Bussink 1990). This fragment contains most of the structural *pgaII* gene of *A. niger* N400 and about 200 bp of the 5′ upstream region. In the restricted genomic DNA of both strains a single hybridizing sequence is detected (Fig. 2), but the size of the hybridizing fragments is not identical. The *Hinc*II restriction site which is present in the structural *pgaII* gene of *A. niger* N400 is absent in the *pgaII* gene of *A. tubigensis* NW756, whereas the *Eco*RV restriction site which occurs in *A. tubigensis* NW756 is absent in *A. niger* N400, indicating divergence of these genes. The *pgaII* gene of *A. tubigensis* NW756 was isolated by screening a genomic library of *A. tubigensis* NW756 DNA in the phage lambda replacement vector EMBL3. After a second screening step, DNA was isolated from four positive recombinant phages. These phages all contained the hybridizing 5.5 kbp *Xho*I-*Bgl*II fragment which was also observed on Southern blots of restricted *A. tubigensis* NW756 DNA (data not shown). This fragment was isolated and ligated into the *Bam*HI and *Sal*I-digested vector pEMBL18. The ligation mixture was used to transform *E. coli* JM109, positive clones were identified by colony hybridization and plasmid DNA was then isolated. The resulting plasmid which carries the *pgaII* gene of *A. tubigensis* NW756, designated as pGW1756, was further characterized with restriction enzymes (Fig. 3). The restriction map of pGW1756 is in agreement with the hybridizing fragments detected in digested chromosomal *A. tubigensis* DNA (Fig. 2), but it is very different from the map of the *A. niger* N400 *pgaII* gene (Bussink et al. 1990).

Plasmid pGW1756 was used to co-transform *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 previously optimized growth medium with both 1% sugar beet pulp and 1% pectin as PG-inducing carbon sources, and the amount of PG in the culture filtrates was subsequently assayed by activity measurements. PG activities over ten-fold higher than that of *A. niger* N402 were observed in *A. niger* transformants, e.g., N593-pGW1756/7, from which it is concluded that the cloned *A. tubigensis* NW756 *pgaII* gene is functional.

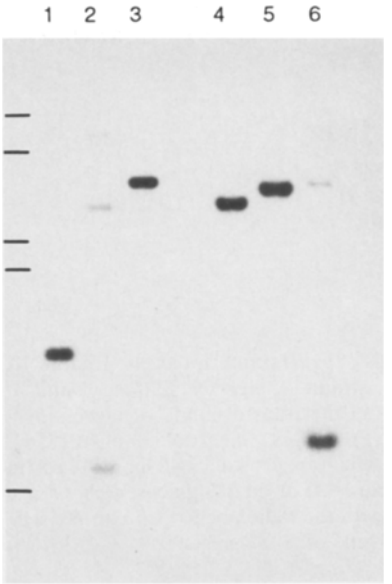

Fig. 2. Southern blot analysis of *A. niger* N400 (*lanes 1–3*) and *A. tubigensis* NW756 (*lanes 4–6*) DNA. Genomic DNA was digested with *Bam*HI and *Bgl*II (*lanes 1 and 4*), *Hinc*II (*lanes 2 and 5*) and *Eco*RV lanes (*lanes 3 and 6*) and electrophoresed in an 0.85% agarose gel. The blot was hybridized with the radiolabelled 1.2 kbp *Bam*HI-*Bgl*II fragment of the *pgaII* gene of *A. niger* N400. Size markers (6.6, 4.4, 2.3, 2.0 and 0.56 kbp) are phage lambda DNA restricted with *Hin*dIII

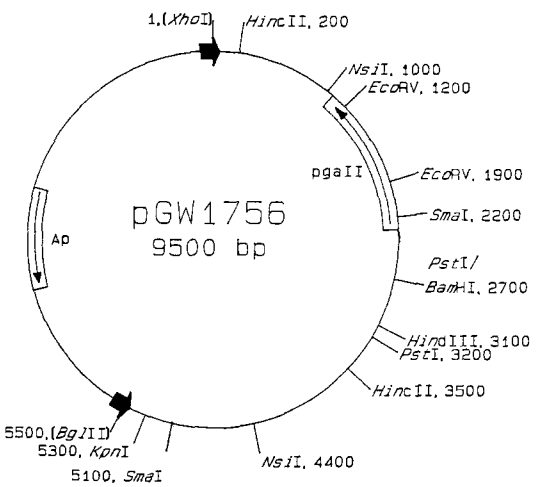

Fig. 3. Restriction map of the *pgaII* gene of *A. tubigensis* NW756. Plasmid pGW1756, which is the 5.5 kbp *Xho*I-*Bgl*II fragment (*thick arrows*) containing the *pgaII* gene inserted into pEMBL18, is shown. The ampicillin resistance gene (*Ap*) is indicated, but restriction sites in the vector are not shown

Regulated pgaII gene expression in PGII over-producing A. niger transformants

To further investigate the regulation of *pgaII* gene expression, transcript levels were analyzed by Northern blotting (Fig. 4). As described in Materials and methods equal amounts of total RNA were taken. Both *A. niger* and *A. tubigensis* synthesized a single PGII mRNA when cells were grown with pectin and sugar beet pulp as the carbon sources (the hybridizing smear in lane 2 is likely to be the

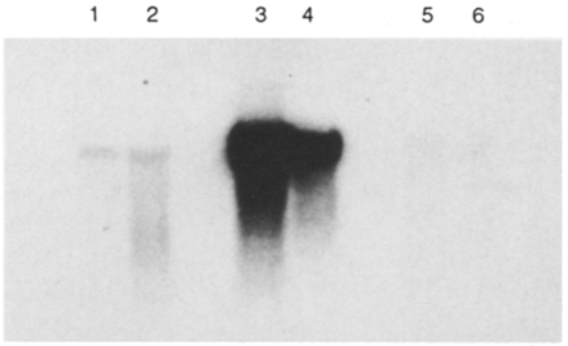

Fig. 4. Northern blot analysis of *pgaII* transcript levels. Total RNA extracted from the wild-type strains *A. niger* N402 (*lane 1*) and *A. tubigensis* NW756 (*lane 2*) and the transformants *A. niger* N593-pGW1800/27 (*lanes 3 and 5*) and *A. niger* N593-pGW1756/7 (*lanes 4 and 6*) was used. Mycelia were grown for 42 h on 1% pectin and 1% sugar beet pulp (*lanes 1–4*) or on 2% glucose (*lanes 5 and 6*). The blot was hybridized with the radiolabelled 1.2 kbp *Bam*HI-*Bgl*II fragment of the *pgaII* gene of *A. niger* N400

result of RNA degradation during the isolation of the RNA). Transcript levels were increased in *A. niger* strains transformed with either the cloned homologous gene (*A. niger* N593-pGW1800/27) or the cloned *A. tubigensis* gene (*A. niger* N593-pGW1756/7). When glucose (2%) was used as the sole carbon source, very low transcript levels were observed in the multicopy transformants. These levels were even lower than those obtained for the wild-type strains grown on pectin and sugar beet pulp medium.

These results show, firstly, that *pgaII* gene expression is strongly regulated, most likely at the level of transcription, and secondly, that the regulatory sequences in the *A. tubigensis pgaII* gene are recognized by the components of the *A. niger* regulatory apparatus. They also indicate that sufficient 5′ upstream sequences are present in both pGW1800 and pGW1756 to obtain regulated gene expression. In case of pGW1800, the 5′ upstream region preceding the translation initiation codon is 1356 bp.

Polygalacturonase gene structure and sequence divergence

The nucleotide sequence of the complete *pgaII* gene of *A. tubigensis* NW756 was determined and the sequence of the *A. niger* N400 gene was completed by sequencing of the 5′ and 3′ flanking sequences of the structural gene. The nucleotide sequences, as well as the derived amino acid sequences, were aligned (Fig. 5). The aligned sequences comprise about 900 bp of the upstream region, the structural gene, one intron, and 534 bp and 386 bp of the 3′ noncoding regions of the *pgaII* genes of *A. niger* N400 and *A. tubigensis* NW756, respectively.

The structure of the two different *pgaII* genes is very similar and the nucleotide sequences shown are 83% identical. The nucleotide sequence identity in the protein coding regions, viz. 90%, is higher than in the flanking sequences. Like the *A. niger* N400 *pgaII* (Bussink et al. 1990), the coding sequence of the NW756 *pgaII* gene is interrupted by a 52 bp intron.

Upstream of the ATG there is a possible TATA box at position-116 (numbering refers to the N400 sequence), conserved in both sequences. The possible candidates for a CAAT box (to about − 200 before the TATA box) are not conserved between the *A. niger* and *A. tubigensis* sequences, suggesting the absence of a functional CAAT box, which is often observed for fungal promoters (Gurr et al. 1987). On the other hand pyrimidine-rich sequences, which are frequently observed fungal promoter elements, are found in the *pgaII* genes from position −38 to −18 as well as at several other locations further upstream. The nucleotide sequence identity in the 5′ upstream non-protein coding region is 81% with deletions or insertions of at most four nucleotides at 12 locations. The degree of homology does not appear to decrease as one moves upstream of the ATG.

The intron of both genes shows 79% sequence identity. Its position is confirmed by the cDNA sequence of Ruttkowski et al. (1990). The 5′ splice site of the intron, GTAAGC, is identical to the splice site found for an intron of the *A. awamori pepA* gene (Berka et al. 1990) and it resembles the filamentous fungal 5′ splice consensus GTANGT (Gurr et al. 1987). The 3′ splice sitze, TAG, is in complete agreement with the fungal consensus 3′ splice site PyAG. The fungal intron internal consensus sequence, NPuCTPuAC (Rambosek and Leach 1987) or PyGCTAAC (Gurr et al. 1987), is not found; but a perfect match to a more general form of these consensus sequences, i.e., the consensus sequence PyNPyPyPuAPy of higher eukaryotes (Lewin 1987), is found in the introns of both genes with the conserved A at position −11 with respect to the 3′ splice site. Thus, the putative lariat sequences TATTGAT (N400) and TGTTGAT (NW756) differ by one nucleotide at a position which is not usually highly conserved. It can be seen that the mutations affect the central part of the intron, whereas the borders remain unchanged.

The predicted PGII amino acid sequences are 94% identical, differing at 23 positions. Most of these amino acid substitutions can be explained by either a single point mutation (14 times) or by the combination of a point mutation that changes the amino acid and an additional silent point mutation (six times). However, at positions 44, 307 and 321 the amino acid substitutions cannot be the result of single, or multiple, point mutations which directly change the encoded amino acid. Thus, it is likely that *Aspergillus* strains exist, or have existed, which produce a PGII with different amino acids at these positions.

In addition to the mutations that lead to amino acid substitutions, there are 81 silent mutations. We have,

Fig. 5. The nucleotide sequences of the *pgaII* genes of *A. niger* N400 (N400) and *A. tubigensis* NW756 (NW756). The deduced amino acid sequence of the *A. niger* PGII is shown, whereas for the *A. tubigensis* PGII only those amino acids are shown which are different in the *A. niger* PGII. The start of the mature proteins is indicated by "⋙". Conserved nucleotides are indicated by *asterisks*. Alignment improvements by the introduction of insertions are marked by *points*. The "+" and "=" characters indicate the region of repeats in the *A. niger* gene and the *open spaces* in this string indicate the nucleotides which deviate from the consensus TCGTCTGCT

Table 1. Codon utilization in the *pgaII* genes of *A. tubigensis* NW756 and *A. niger* N400

	NW756	N400		NW756	N400		NW756	N400		NW756	N400
UUU Phe	4	4	UCU Ser	11	11	UAU Tyr	2	1	UGU Cys	1	2
UUC Phe	5	8	UCC Ser	13	12	UAC Tyr	6	6	UGC Cys	8	7
UUA Leu	0	1	UCA Ser	1	0	UAA –	0	0	UGA –	0	0
UUG Leu	2	1	UCG Ser	3	5	UAG –	1	1	UGG Trp	7	7
CUU Leu	6	7	CCU Pro	2	4	CAU His	3	3	CGU Arg	0	0
CUC Leu	6	5	CCC Pro	6	3	CAC His	4	5	CGC Arg	1	2
CUA Leu	1	0	CCA Pro	0	1	CAA Gln	0	0	CGA Arg	0	2
CUG Leu	5	3	CCG Pro	1	1	GAG Gln	4	7	CGG Arg	3	0
AUU Ile	6	8	ACU Thr	15	8	AAU Asn	3	6	AGU Ser	3	3
AUG Ile	20	17	ACC Thr	18	25	AAC Asn	17	16	AGC Ser	8	7
AUA Ile	0	0	ACA Thr	1	0	AAA Lys	5	4	AGA Arg	0	0
AUG Met	4	4	ACG Thr	6	6	AAG Lys	17	16	AGG Arg	0	0
GUU Val	4	6	GCU Ala	10	9	GAU Asp	14	12	GGU Gly	16	10
GUC Val	15	14	GCC Ala	10	10	GAC Asp	11	10	GGC Gly	17	23
GUA Val	0	0	GCA Ala	2	1	GAA Glu	5	8	GGA Gly	8	5
GUG Val	7	8	GCG Ala	7	8	GAG Glu	6	5	GGG Gly	2	5

therefore, looked for differences in codon usage. Codon selection has been correlated with the level of gene expression (Bennetzen and Hall 1982; May et al. 1987) and the codon usage in different species such as *N. crassa* and *A. nidulans* is not identical (Gurr et al. 1987). The silent mutations do not significantly change the over-all frequency of the four individual bases at the third nucleotide position, but the frequency of occurrence of specific threonine, glycine and arginine codons is markedly changed (Table 1). For glycine (GGN) and threonine (ACN) the codons with C and U at the wobble position are preferred and they are used with roughly equal probability in the NW756 *pgaII* gene, whereas in the N400 gene the codons with a C are used more frequently. These changes are largely due to directional U-C transitions at the third position in the glycine and threonine codons, and are thus independent of selective pressure on the protein sequence. For example, there are six U(NW756) to C(N400) transitions in the threonine codons, whereas there is only one transition in the opposite direction. The CGG arginine codon is used preferentially in the *A. tubigensis pgaII* gene, but it is not present in the *A. niger* gene which contains two CGA codons not present in the *A. tubigensis* gene. In other fungal genes both codons are often under-represented or not used at all (Gurr et al. 1987).

The highest sequence divergence between the *A. niger* and *A. tubigensis pgaII* genes is observed in the 3′ non-protein coding sequences. In the last 100 nucleotides of the aligned sequences the sequence identity is relatively high (82%), which may indicate its importance for the termination of transcription. However, the sequence in between this region and the structural gene has substantially diverged, and includes two deletions or insertions of 76 and 77 base pairs, respectively. Of particular interest is a region of repeats in the *A. niger* sequence, which is absent in the *A. tubigensis* gene. The symmetric element TCGTCTGCT is repeated six times and six additional mutant forms of this element with up to two substitutions are found, whereas two of the four mutant forms are also duplicated. Since the *A. tubigensis pgaII* gene is also expressed and regulated in transformed *A. niger* strains, this region of repeats is clearly not essential for *pgaII* gene expression.

The following results support the recognition of two different closely related species in the *Aspergillus niger* aggregate, viz., *A. niger* and *A. tubigensis* (Kusters-van Someren et al. 1991). The expression of the *pgaII* genes is regulated in essentially the same way, although there are quantitative differences which may reflect differences in the *pgaII* genes, but more probably are due to the different genetic backgrounds (unpublished result). The nucleotide sequences, and notably also the amino acid sequences, have diverged substantially and this is in striking contrast to the glucoamylase genes of *A. niger* (Boel et al. 1984) and *A. awamori* (Nunberg et al. 1984). These strains could, however, both be *A. niger* species or *A. tubigensis* species. It is also possible that the extreme conservation of the glucoamylase sequences reflects different selective pressures on glucoamylases and PGs. Korman et al. (1990) have reported a high degree of identity (greater than 98%) of the α-amylase genes of *A. niger* var. *awamori* and *A. oryzae,* which are not considered to be closely related species, whereas the respective acid protease genes are only approximately 70% homologous.

Apart from the absence of the intron, the sequence of a PG cDNA clone of *A. niger* RH5344, which has been recently published (Ruttkowski et al. 1990), differs by only two bases out of the 1320 from the *A. tubigensis* NW756 *pgaII* sequence. Both differences are in the non-protein coding region. This indicates that *A. niger* RH5344 belongs to the *A. tubigensis* group and not to the *A. niger* group. The predicted amino acid sequence from this cDNA clone is identical to the amino acid sequence (Ruttkowski et al. 1990) of the PG isolated from the commercial pectinase ROHAPECT D5L (Röhm, Darmstadt, FRG). The PGII isolated from the commercial Pectinase K2B 078 (Rapidase, Seclin, France) is identical to the PGII produced by *A. niger* N400 (Bussink et al. 1990). Consequently, both *A. niger* and *A. tubigensis* have been

```
PGII-N400   176  GHNTDAFDVGNSVGVNIIKPWVHNQDDCLAVNS------GENIWFTGGTCIGGHGLSIGSVG-DR  233
PGII-NW756  176  GHNTDAFDVGNSVGVNIIKPWVHNQDDCLAINS------GENIWFTSGTCIGGHGLSIGSVG-GR  233
PEH         224  ARNTDGIDPMSSKNITIAHSNISTGDDNVAIKAYKGRSETRNISILHNEFGTGHGMSIGS----E  284
PEHA        224  ARNTDGIDPMSSKNITIAYSNIATGDDNVAIKAYKGRAETRNISILHNDFGTGHGMSIGS----E  284
PGA         280  VKNTDGFDPGQSTNVVLAYSYINTGDDHVAVKASSGP--TRNLLFAHNHFYYGHGLSIGS----E  338
PG-2A       245  SPNTDGVHVSNTQYIQISDTIIGTGDDCISIVS-----GSQNVQATNITCGPGHGISIGSLGSGN  304
                   ***. . . . . . . . ** ... .    . *.       ***.****

PGII-N400   234  SNNVVKNVTIEHSTVSNSE----NAVRIKTISGATGSVSEITYSNIVMSGISDYGVVIQQDYEDG  294
PGII-NW756  234  SNNVVKNVTIEHSTVSNSE----NAVRIKTVSGATGSVSEITYSNIVMSGISDYGVVIQQDYEDG  294
PEH         285  T-MGVYNVTVDDLIMTGTT----NGLRIKSDKSAAGVVNGVRYSNVVMKNVAK-PIVIDTVYEKK  343
PEHA        285  T-MGVYNVTVDDLKMNGTT----NGLRIKSDKSAAGVVNGVRYSNVVMKNVAK-PIVIDTVYEKK  343
PGA         339  TNTGVSNMLVTDLTMDGNDSSAGNGLRIKSDASRGGKVTNIVYDGICMRNVKE-PLVFDPFYSSV  402
PG-2A       305  SEAYVSNVTVNEAKIIGAE----NGVRIKTWQGGSGQASNIKFLNVEMQDVKY-PIIIDQNYCDR  364
                   . * *. ... . .     *..***.  . .* .... . .. * ..    .... * .
```

Fig. 6. The homologous domain in the PGs of *A. niger* N400 (PGII-N400), *A. tubigensis* NW756 (PGII-NW756), *E. carotovora* subsp. *carotovora* strain SCRI193 (PEH), *E. carotovora* subsp. *carotovora* strain SCC3193 (PEHA), *P. solanacearum* (PGA) and tomato (PG-2A). Conserved amino acids are indicated by *asterisks*, conservative substitutions by *points* and where alignment has been improved by the introduction of insertions, this is marked by *dashes*

used for the production of pectinases which are employed commercially in the food and beverage industry. Thus, both appear to have obtained GRAS status before they were recognized as two distinct species.

A homologous domain in polygalacturonases of plant, bacterial and fungal origin

We have previously reported a low, but significant, homology between the mature tomato polygalacturonase PG-2A (Grierson et al. 1986) and the *A. niger* PGII (Bussink et al. 1990). Recently, the amino acid sequences of the polygalacturonases produced by *Erwinia carotovora* subsp. *carotovora* strain SCRI193 (PehA; Hinton et al. 1990), by *Erwinia carotovora* subsp. *carotovora* strain SCC3193 (PehA; Saarilahti et al. 1990) and by *Pseudomonas solanacearum* (PGA; Huang and Schell 1990), were derived from the corresponding DNA sequences and compared individually to the tomato polygalacturonase. The homology between the plant and the bacterial PGs is low, e.g., 26% and 28% with the *Erwinia* PGs, as is the homology between the bacterial and the fungal PGs. However, a multiple sequence alignment according to Higgins and Sharp (1989) reveals a domain of about 125 amino acids which is conserved among all the PGs sequenced (Fig. 6). The sequence identity in this region is 18% and the degree of homology rises to 50% if conservative substitutions are allowed. The localized homology suggests a function for this domain in the catalytic reaction, or in the binding of the substrate, or in both. Chemical modification studies, and an analysis of the kinetic parameters as a function of the pH, have indicated the presence of an essential histidine in the *A. niger* PG (Rexová-Benková and Slezárik 1970; Cooke et al. 1976; Rexová-Benková and Mracková 1978). As there is only one histidine conserved amongst all these PGs, and this histidine is present in the most conserved part of the conserved domain, namely at position 223 in PGII, it is likely that this histidine is involved in the catalytic reaction. Rexová-Benková and Mracková (1978) have suggested a carboxylate group may be a component of the catalytic site. Possible candidates are the carboxyl functions of the aspartic acid residues at positions 180, 201 and 202, respectively, in PGII. The relatively well conserved and positively charged Arg(256)-Ile-Lys(258) sequence could be involved in ionic interactions with the carboxylate groups present in the substrate.

Homology in fungal secretory pro-peptides

The start of the mature *A. niger* N400 PGII, namely Asp28, is not immediately preceded by a signal peptidase cleavage site (von Heijne 1986) and it has been hypothesized, therefore, that the N-terminal extension of the PGII precursor may be removed in two processing steps (Bussink et al. 1990). The mature *A. tubigensis* PGII starts with Gly28 (Ruttkowski et al. 1990) and it is likely that the 27 amino acid-long N-terminal extension of the precursor comprises a signal peptide as well as a pro-peptide. The optimum signal peptidase cleavage site calculated according to von Heijne (1986) is at Ala21. Thus, a possible sequence of the hypothetical pro-peptide is Ser(22)-Pro-Ile-Glu-Ala-Arg(27). This sequence is conserved in the *A. niger* and *A. tubigensis* PGII precursors, whereas there are three amino acid substitutions in the signal peptide and there are also two substitutions within the first six amino acids of the mature proteins. The cDNA sequence (Oka et al. 1990) of α-sarcin, a cytotoxin with ribonuclease activity produced by *Aspergillus giganteus*, indicates that it may also be preceded by a pro-sequence hexapeptide Ser(22)-Pro-Leu-Glu-Ala-Arg(27), which is almost identical to the putative PGII pro-peptide. In the case of the cellobiohydrolase II of *Trichoderma reesei* it has also been hypothesized that the precursor is proteolytically processed in two steps (Teeri et al. 1987). Depending on the location of the signal peptidase cleavage site, a possible pro-peptide could be Val-Pro-Leu-Glu-Glu-Arg. Thus, the precursors of the otherwise unrelated fungal proteins PGII, α-sarcin and cellobiohydrolase II have a sequence motif in common before the start of the mature protein, namely ser-PRO-leu-GLU-ala-ARG (amino acids in uppercase letters designate complete conservation and lowercase letters partial conservation). This similarity in these putative pro-peptides may indicate a specific biological function. A characteristic feature of these pro-peptides is the presence of only one basic amino acid, always an arginine, preceding the cleavage site. This distinguishes them from the pro-peptides with a dibasic cleavage site found, for example, in the precursors of the *A. awamori* glucoamylase (Innis et al. 1985) and the *Phanerochaete chrysosporium* ligninases (Tien and Tu 1987; de Boer et al. 1987) which have no other sequence homology. Recently, it has been shown by heterologous expression in *S. cerevisiae* that the glucose oxidase precursor of *A. niger* possesses both a cleavable signal peptide and a pro-peptide with a monobasic cleavage site (Frederick et al. 1990). Its pro-peptide, Leu-Pro-His-Tyr-Ile-Arg, does not have extensive homology with the other pro-peptides. This may suggest that the sequence requirement of the monobasic processing enzyme is not very strict with respect to the primary amino acid sequence cleaved and, thus, that the sequence similarity between the pro-peptides of, for exam-

ple, PGII and α-sarcin is not (only) the result of selective pressure on their proper removal. However, it is also conceivable that the PGII and glucose oxidase precursors are processed by different enzymes. Benoit et al. (1987) have deduced a pattern for mammalian pro-hormone cleavage at monobasic sites characterized by the facts that: (1) a second basic amino acid (Arg or sometimes His) is present three or five or seven amino acids before the arginine at the cleavage site; (2) a leucine (isoleucine) or several alanines are virtually always present in the two amino acids immediately preceding and the two amino acids following the arginine at the cleavage site. The pro-peptide of the glucose oxidase precursor shows this pattern, but the pro-peptide of the PGII precursor and the related pro-peptides do not. On the basis of this criterion, and due to the lack of homology between the pro-peptides of the PGII precursors, α-sarcin and CBHII, on the one hand, and of the glucose oxidase precursor, on the other, it seems reasonable to assume that these are representatives of two different classes of pro-peptides. In addition, the pro-peptides with dibasic cleavage sites represent a third class.

Acknowledgements. We thank Gerda Nicolai and Ditte Preker for their contributions to the expression experiments and Michel Flipphi for isolating the *A. tubigensis* gene.

References

Bennetzen JL, Hall BD (1982) J Biol Chem 257: 3026–3031
Benoit R, Ling N, Esch F (1987) Science 238: 1126–1129
Berka RM, Ward M, Wilson LJ, Hayenga KJ, Kodama KH, Carlomagno LP, Thompson SA (1990) Gene 86: 153–162
Boel E, Hansen MT, Hjort I, Hoegh I, Fiil NP (1984) EMBO J 3: 1581–1585
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
Cathala G, Savouret J-F, Mendez B, West BL, Karin M, Martial JA, Baxter JD (1983) DNA 2: 329–335
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, pp 385–392
Church GM, Gilbert W (1984) Proc Natl Acad Sci USA 81: 1991–1995
Collmer A, Keen NT (1986) Annu Rev Phytopathol 24: 383–409
Cooke RD, Ferber CEM, Kanagasabapathy L (1976) Biochim Biophys Acta 452: 440–451
Cooper RM (1984) The role of cell wall-degrading enzymes in infection and damage. In: Wood RKS, Jellis GJ (eds) Plant diseases: infection, damage and loss. Blackwell, Oxford, pp 13–27
Dean RA, Timberlake WE (1989) 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
Frederick KR, Tung J, Emerick RS, Masiarz FR, Chamberlain SH, Vasavada A, Rosenberg S, Chakraborty S, Schopter LM, Massey V (1990) J Biol Chem 265: 3793–3802
Goosen T, Bloemheuvel G, Gysler C, de Bie DA, van den Broek HWJ, Swart K (1987) Curr Genet 11: 499–503
Graaff L de, van den Broek H, Visser J (1988) Curr Genet 13: 315–321
Grierson D, Tucker GA, Keen J, Ray J, Bird CR, Schuch W (1986) Nucleic Acids Res 14: 8595–8603
Gurr SJ, Unkles SE, Kinghorn JR (1987) 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
Heijne G von (1986) Nucleic Acids Res 14: 4683–4690
Higgins DG, Sharp PM (1989) CABIOS 5: 151–153
Hinton JCD, Gill DR, Lalo D, Plastow GS, Salmond GPC (1990) Mol Microbiol 4: 1029–1036
Huang J, Schell MA (1990) J Bacteriol 172: 3879–3887
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
Kester HCM, Visser J (1990) Biotech Appl Biochem 12: 150–160
Khanh NQ, Albrecht H, Ruttkowski E, Löffler F, Gottschalk M, Jany KD (1990) Nucleic Acids Res 18: 4262
Knowles J, Lehtovaara P, Teeri T (1987) Trends Biotechnol 5: 255–261
Korman DR, Bayliss FT, Barnett CC, Carmona CL, Kodama KH, Royer TJ, Thompson SA, Ward M, Wilson LJ, Berka R (1990) Curr Genet 17: 203–212
Kusters-van Someren MA, Samson, RA, Visser J (1991) Curr Genet 19: 21–26
Lewin B (1987) Genes 3rd edn. Wiley, New York
Maniatis T, Fritsch EF, Sambrook J (1982) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York
May GS, Tsang MLS, Smith H, Fidel S, Morris NR (1987) Gene 55: 231–243
Misaghi IJ (1982) Physiology and biochemistry of plant-pathogen interactions. Plenum Press, New York, London
Nunberg JH, Meade JH, Cole G, Lawyer FC, McCabe P, Schweickart V, Tal R, Wittman VP, Flatgaard JE, Innes MA (1984) Mol Cell Biol 4: 2306–2315
Oka T, Natori Y, Tanaka S, Tsurugi K, Endo Y (1990) Nucleic Acids Res 18: 1897
Rambosek J, Leach J (1987) CRC Crit Rev Biotechnol 6: 357–393
Rexová-Benková L, Markovic O (1976) Adv Carbohydr Chem Biochem 33: 323–385
Rexová-Benková L, Mracková M (1978) Biochim Biophys Acta 523: 162–169
Rexová-Benková L, Slezárik A (1970) Coll Czech Chem Commun 35: 1255–1260
Ried JL, Collmer A (1985) Appl Environ Microbiol 50: 615–622
Rombouts FM, Pilnik W (1980) 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
Saarilahti HT, Heino P, Pakkanen R, Kalkkinen N, Palva I, Palva ET (1990) Mol Microbiol 4: 1037–1044
Teeri TT, Lehtovaara P, Kauppinen S, Salovuori I, Knowles J (1987) Gene 51: 43–52
Thomas PS (1983) Methods Enzymol 100: 255–266
Tien M, Tu CPD (1987) Nature 326: 520–523
Yanisch-Perron C, Vieira J, Messing J (1985) Gene 33: 103–119

Communicated by L. A. Grivell