Molecular cloning, nucleotide sequence and expression of the gene encoding prepro-polygalacturonaseII of *Aspergillus niger*

Hendrik J.D. Bussink, Harry C.M. Kester and Jaap Visser

Department of Genetics, Agricultural University, Dreyenlaan 2, 6703 HA Wageningen, The Netherlands

Received 17 August 1990

PolygalacturonaseII of *Aspergillus niger* was fragmented using CNBr and the NH_2-terminal fragment and another fragment were partially sequenced. The polygalacturonaseII (*pga*II) gene was then isolated by using an oligonucleotide mixture based on the internal amino acid sequence as a probe. The nucleotide sequence of the *pga*II structural gene was determined. It was found that polygalacturonaseII is synthesized as a precursor having an NH_2-terminal prepro-sequence of 27 amino acids. The cloned gene was used to construct polygalacturonaseII over-producing *A. niger* strains. PolygalacturonaseII was isolated from one such strain and was determined to be correctly processed and to be fully active.

Prepro-peptide processing; Polygalacturonase; Transformation; Industrial enzyme; Secretion; *Aspergillus niger*

1. INTRODUCTION

Endopolygalacturonases (poly(1,4-α-D-galacturonide) glycanohydrolase, EC 3.2.1.15), reviewed in [1,2], hydrolyse pectate in a more or less random fashion. In nature, pectate is primarily found in its esterified form, i.e. as pectin, in the middle lamella and primary cell wall of higher plants. The microbial degradation of pectin is a complex process involving different pectinolytic activities, e.g. *A. niger* synthesizes endopolygalacturonase (endo-PG), exo-PG, endopectin lyase and pectinesterase.

The pectinolytic enzymes of *A. niger* are widely applied in the fruit juice technology as crude and usually ill-characterized mixtures. We have purified and characterized five different endo-PGs of *A. niger* from a single commercial preparation [3]. The importance of the individual endo-PGs is not understood and at the molecular level little is known with respect to their structure as well as to the structure and expression of the corresponding gene(s). In order to really improve our knowledge about fungal PGs, it is therefore necessary to extend the conventional repertoire for enzyme purification and characterization with the tools offered by modern molecular biology, and to develop an efficient expression system for the PG gene(s) in filamentous fungi [4].

Correspondence address: J. Visser, Department of Genetics, Agricultural University, Dreyenlaan 2, 6703 HA Wageningen, The Netherlands

Abbreviations: PG, polygalacturonase; *pga*II, the PGII gene; bp, base pair; DABITC, dimethylaminoazobenzene isothiocyanate

Here we report on the partial sequencing of an *A. niger* endo-PG, the cloning and sequencing of the corresponding gene and the expression of the cloned gene in transformed *A. niger* strains.

2. EXPERIMENTAL

2.1. *Purification, detection and sequencing of polygalacturonaseII*

Purification of polygalacturonaseII (PGII) and further protein analyses were performed as described (see [3] and references herein). Probing of Western blots with a PGII specific monoclonal antibody was done according to the relevant manual issued by Bio-Rad. Manual NH_2-terminal endgroup analysis was carried out according to [5]. Cyanogen bromide fragments were separated by SDS-polyacrylamide gel electrophoresis and subsequently electroblotted onto Immobilon-P (Millipore) polyvinylidene difluoride membranes [6]. The relevant pieces of membrane were recovered and then used for sequence analysis, using a gas phase sequencer equipped with a PTH analyzer as described [7].

2.2. *Manipulation of DNA*

Manipulations of DNA were performed according to procedures [8] and using *E. coli* strains (NM593, LE392, JM109, DH5αF') well known in the art. Filter hybridization with the oligonucleotide probe mixture was carried out at 47°C in a buffer containing 2×SSC, 10× Denhardt's solution and 0.1% SDS, followed by intermittent washes in 2×SSC at room temperature and a stringent wash at 63°C in 6×SSC, 0.05% sodium pyrophosphate for half-an-hour. pEMBL vectors [9] were used for routine subcloning. Sequencing of recombinant single stranded M13 phages and plasmid DNA using the ^{T7}Sequencing Kit was according to the recommendations of the supplier (Pharmacia LKB Biotechnology AB), employing three additional primers.

2.3. *Transformation and cultivation of* A. niger

Aspergillus niger N400 (CBS 120.49) and its derivatives N402 (*csp*A) and N593 (*csp*A, *pyr*A) have been described [10]. Transformation of *A. niger* N593 was performed essentially as described [10], using pGW635 [11] as the selective vector. In order to ensure a high co-transformation frequency, the co-transforming plasmid was added in

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/90/$3.50

a 20-fold excess. The medium used to induce PG synthesis consists of trace elements [12], 1.5 g/l KH_2PO_4, 0.5 g/l KCl, 0.5 g/l $MgSO_4$, 4.0 g/l NH_4Cl, 10 g/l sugar beet pulp, ground in a Waring blender, and 10 g/l pectin (d.e. 61.2%; Obipektin, Bischofszell, Switzerland)

3. RESULTS

PGII, which is identical to endo-II as described by Kester and Visser [3], was selected for further analysis. Of the five endo-PGs isolated from a commercial enzyme preparation, it is the most abundant enzyme and it has the highest specific activity.

Aspartic acid was identified as the NH_2-terminal amino acid, using the DABITC method [5]. Next, PGII was fragmented using CNBr and partial amino acid sequences of a 5 kDa and a 17 kDa fragment were subsequently determined. The NH_2-terminal amino acid sequence of the 5 kDa fragment is as follows: Asp-Ser-X-Thr-Phe-Thr-Thr-Ala-Ala-Ala-Ala-Lys-Ala. At position 3 no amino acid residue was detected. 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. The sequence of the 17 kDa fragment is shown in Fig. 1. It does not start with aspartic acid, must therefore be internally located and will thus be preceded by a methionine. This sequence was used to design a specific 29-mer oligonucleotide probe mixture of 32 components (Fig. 1). The number of oligonucleotides in the mixture was reduced by introducing deoxyinosine at the third position of four-fold degenerated codons [13]. Initially two mixtures were synthesized with either TCI or AGI representing the serine codon, but these mixtures were combined before use.

The PGII-specific oligonucleotide probe was used to screen a genomic DNA library of *A. niger* N400 in the lambda replacement vector EMBL4. Six positive signals were obtained among the about 12 000 plaques screened. After a rescreening step to purify the positive phages, their DNA inserts were characterized with restriction enzymes. At least five of these were found to overlap. The hybridizing 4.1 kbp and 2.4 kbp *Xba*I-*Eco*RI fragments of two different phages, respectively, were inserted into pEMBL vectors, resulting in pGW1800 and pGW1803, respectively. The physical map of pGW1800 (Fig. 2) is colinear with the map of part of the *A. niger* DNA, as determined by Southern blot analysis of restricted chromosomal *A. niger* N400 DNA, using the 1.45 kbp as well as the 2.05 kbp *Eco*RV-*Bgl*II fragment of pGW1800 as a probe (data not shown). pGW1803 was found to have a shorter, but otherwise identical, insert as pGW1800, extending from the *Xba*I site at position 1 to just before the *Bgl*II site at position 2400. Using Southern blot analysis, the specific sequence which hybridizes with the oligonucleotide probe mixture was located in between the unique *Nco*I site and the *Hin*cII site at position 2000 (Fig. 2).

```
CB-17        AlaPheSerValGlnAlaAsnAspIleThrPhe
                . . . . . .
          5' ATGGCITTYTCIGTICARGCIAAYGAYAT      3'
probe              **
          5' ATGGCITTYAGIGTICARGCIAAYGAYAT      3'
                . . . . . .
gene      5' ATGGCGTTTAGTGTCCAGGCGAATGACATTACGTTT 3'
```

Fig. 1. The NH_2-terminal amino acid sequence of the 17 kDa CNBr fragment (CB-17), the oligonucleotide probe mixture as based on this sequence (probe) and the nucleotide sequence of the *pga*II gene matching the probe (gene). The asterisks indicate the differences between the two oligonucleotide mixtures before they were combined. Y = C, T; R = A, G; I = inosine.

The nucleotide sequence of the *pga*II structural gene was determined over both strands of DNA (Fig. 3). The coding sequence is interrupted by a 52 bp intron which changes the reading frame and which contains translational stop codons in all possible reading frames. The *pga*II gene encodes a protein of 362 amino acids. It is now clear that the NH_2-terminal sequence of the 5 kDa CNBr fragment of PGII is identical to the NH_2-terminal amino acid sequence of mature PGII. The deduced amino acid sequence has an NH_2-terminal extension of 27 amino acids before the start of the mature protein, and this sequence has all the characteristics of a secretory signal sequence [14]. The calculated molecular mass of the deduced mature PGII is 35 kDa, which agrees well with experimental data [3]. An alignment of the sequences of the mature PGII and the mature tomato PG [15,16] indicates a low (27%) but significant homology and also confirms the presence and location of the intron in the *pga*II gene (not shown).

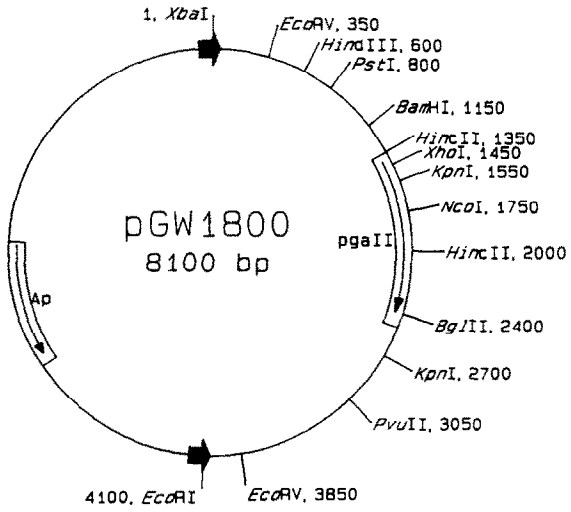

Fig. 2. Restriction map of the *pga*II gene. Plasmid pGW1800, which is the 4.1 kbp *Xba*I-*Eco*RI fragment (thick arrows) containing the *pga*II gene inserted into pEMBL18, is shown. The ampicillin resistance gene (Ap) is indicated, but restriction sites in the vector are not shown.

```
ATGCACTCGTTTGCTTCTCTTCTCGCCTACGGCCTGGTCGCCGGCGCCACCTTCGCTTCTGCCTCTCCTATC   72
MetHisSerPheAlaSerLeuLeuAlaTyrGlyLeuValAlaGlyAlaThrPheAlaSerAlaSerProIle   24

GAAGCTCGAGACAGCTGCACGTTCACCACCGCTGCCGCTGCTAAAGCGGGCAAGGCGAAATGCTCTACTATC  144
          ************************************
GluAlaArgAspSerCysThrPheThrThrAlaAlaAlaAlaLysAlaGlyLysAlaLysCysSerThrIle   48

ACCCTTAACAACATCGAAGTTCCAGCTGGAACCACCCTCGACCTGACCGGTCTCACCAGCGGTACCAAGGTC  216
ThrLeuAsnAsnIleGluValProAlaGlyThrThrLeuAspLeuThrGlyLeuThrSerGlyThrLysVal   72

ATCTTCGAGGGCACCACGACCTTCCAGTACGAAGAATGGGCAGGCCCCTTGATCTCCATGAGTGGCGAACAT  288
IlePheGluGlyThrThrThrPheGlnTyrGluGluTrpAlaGlyProLeuIleSerMetSerGlyGluHis   96

ATCACCGTCACTGGTGCCTCCGGCCACCTCATCAATTGCGATGGTGCGCGCTGGTGGGATGGCAAGGGAACC  360
IleThrValThrGlyAlaSerGlyHisLeuIleAsnCysAspGlyAlaArgTrpTrpAspGlyLysGlyThr  120

AGCGGAAAGAAGAAGCCCAAGTTCTTTTACGCCCATGGCCTTGACTCCTCGTCTATTACTGGATTAAACATC  432
SerGlyLysLysLysProLysPhePheTyrAlaHisGlyLeuAspSerSerSerIleThrGlyLeuAsnIle  144

AAAAACACCCCCCTTATGGCGTTTAGTGTCCAGGCGAATGACATTACGTTTACCGATGTTACCATCAATAAT  504
          ************************************
LysAsnThrProLeuMetAlaPheSerValGlnAlaAsnAspIleThrPheThrAspValThrIleAsnAsn  168

GCGGATGGCGACACCCAGGGTGGACACAACACTGATGCGTTCGATGTTGGCAACTCGGTCGGGGGTGAATATC  576
AlaAspGlyAspThrGlnGlyGlyHisAsnThrAspAlaPheAspValGlyAsnSerValGlyValAsnIle  192

ATTAAGCCTTGGGTCCATAACCAGGATGACTGTCTTGCGGTTAACTCTGGCGAGGTAAGCAGCTCTGCATAT  648
IleLysProTrpValHisAsnGlnAspAspCysLeuAlaValAsnSerGlyGlu<----------------

ATGCTTGATTCGTAATTATATTGATATTCTATAGAACATCTGGTTCACCGGCGGCACCTGCATTGGCGGCCAC  721
----intron----------------------->AsnIleTrpPheThrGlyGlyThrCysIleGlyGlyHis  223

GGTCTCTCCATCGGCTCTGTCGGCGACCGCTCCAACAACGTCGTCAAGAACGTCACCATCGAACACTCCACC  793
GlyLeuSerIleGlySerValGlyAspArgSerAsnAsnValValLysAsnValThrIleGluHisSerThr  247

GTGAGCAATTCCGAAAACGCCGTCCGAATTAAGACCATCTCTGGCGCCACTGGCTCCGTGTCCGAGATTACG  865
ValSerAsnSerGluAsnAlaValArgIleLysThrIleSerGlyAlaThrGlySerValSerGluIleThr  271

TACTCCAACATCGTCATGTCTGGCATCTCCGATTACGGCGTGGTCATTCAGCAGGATTACGAAGACGGCAAG  937
TyrSerAsnIleValMetSerGlyIleSerAspTyrGlyValValIleGlnGlnAspTyrGluAspGlyLys  295

CCTACGGGTAAGCCGACGAACGGTGTCACTATTCAGGATGTTAAGCTGGAGAGCGTGACTGGTAGCGTGGAT 1009
ProThrGlyLysProThrAsnGlyValThrIleGlnAspValLysLeuGluSerValThrGlySerValAsp  319

AGTGGGGCTACTGAGATCTATCTTCTTTGCGGGTCTGGTAGCTGCTCGGACTGGACCTGGGACGATGTGAAA 1081
SerGlyAlaThrGluIleTyrLeuLeuCysGlySerGlySerCysSerAspTrpThrTrpAspAspValLys  343

GTTACCGGGGGGAAGAAGTCCACCGCTTGCAAGAACTTCCCTTCGGTGGCCTCTTGTTAG             1141
ValThrGlyGlyLysLysSerThrAlaCysLysAsnPheProSerValAlaSerCysEnd              362
```

Fig. 3. The nucleotide sequence of the *pga*II structural gene. The deduced amino acid sequence is also shown. The NH_2-terminal amino acid sequence of the mature over-produced PGII is underlined and the NH_2-terminal amino acid sequences of the 5 kDa and 17 kDa CNBr fragments of PGII are indicated by asterisks.

In order to prove that pGW1800 contains the complete and functional *pga*II gene of *A. niger*, we introduced pGW1800 in the *A. niger* genome, since this would allow analysis of the expression of the cloned *pga*II gene. To this end, pGW1800 was used as the unselected co-transforming plasmid in an experiment in which the auxotrophic *A. niger* strain N593 was transformed to uridine prototrophy, using a plasmid containing the functional *A. niger pyr*A^+ gene to complement the auxotrophic mutation [10]. The resulting transformants were grown in a medium favouring PG synthesis and subsequently the amount of PGII in the culture medium was assayed. Previous results showed that our laboratory strain *A. niger* N400 synthesizes multiple active forms of PG (to be published elsewhere) and, therefore, that conventional activity measurements per se might not be sufficient to monitor the expression of the *pga*II gene. However, a monoclonal antibody (SC24) has become available, which is specific for PGII (Visser and coworkers, unpublished results). This monoclonal antibody was used to probe the culture filtrates of the transformed strains for PGII by Western blotting. A typical result clearly showing overproduction of PGII in the transformed strain T27 is given in Fig. 4. *A. niger* N402 also secretes PGII reactive with the monoclonal antibody within 48 hours after inoculation (data not shown). The transformant T27 appears to have integrated multiple copies of pGW1800 in its genome, as indicated by Southern blot analysis (data not shown). The PGII secreted by strain T27 has been purified. The specific activity of the purified recombinant enzyme is identical to the specific activity of the previously isolated enzyme from the commercial enzyme preparation [3] and its isoelectric

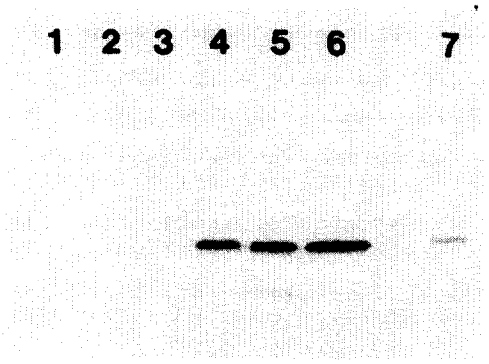

Fig. 4. Western blot analysis of culture filtrates obtained from *A. niger* N402 (lanes 1–3) and the transformant T27 (lanes 4–6) with the PGII specific monoclonal antibody. The samples were taken at 24 h (lanes 1 and 4), 46 h (lanes 2 and 5) and 69 h (lanes 3 and 6) after inoculation. The previously purified PGII was applied in lane 7.

point is 5.2. The NH_2-terminal amino acid sequence of the recombinant polygalacturonase was also determined and found to be identical to the amino acid sequence Asp-28 – Ile-48 as deduced from the nucleotide sequence (Fig. 3). No signal was observed at positions 3 and 18, corresponding to Cys-30 and Cys-45 in the deduced sequence, respectively, due to the fact that the protein had not been *S*-pyridylethylated. An unambiguous sequence was obtained, which indicates the absence of heterogeneity at the NH2-terminus of the over-expressed PGII.

4. DISCUSSION

The *pga*II gene of *A. niger* has been isolated, using a single specific oligonucleotide mixture. We have provided evidence for this by demonstrating that the cloned DNA encodes the NH_2-terminal amino acid sequence of the mature PGII, in addition to the amino acid sequence which had been used to design the probe. Furthermore, we have shown that the cloned gene is expressed in transformed *A. niger* strains whereas the resulting gene product is fully active, correctly processed and reactive with PGII specific monoclonal antibodies. This opens new perspectives to study PG gene expression and the relationship between PG structure and function. The *pga*II gene also provides a DNA probe that should be useful to address the question whether the distinct *A. niger* PGs observed [3] are encoded by different genes, and, if so, to isolate these genes for further characterization. Since there is evidence that polygalacturonase structural features have been conserved among different fungi [17], it is probably also possible to use this gene as a heterologous probe for the isolation of PG genes from phytopathogenic fungi.

As expected for a secreted protein, the PGII is synthesized as a precursor with an N-terminal extension of, in this case, 27 amino acids. However, since the charged amino acids Glu-25 and Arg-27 severely violate the -1,-3 rule for signal peptidase cleavage sites [14], it is very unlikely that processing by a signal peptidase occurs at Arg-27. We therefore hypothesize that the removal of the leader sequence is a two step process, in which the signal peptide is first cleaved by a signal peptidase, possibly at Ala-19 or Ala-21. The remaining short propeptide can subsequently be eliminated by the action of a trypsin-like protease. Such a mechanism has already been described for the heterologous processing of the *A. awamori* glucoamylase precursor in *S. cerevisiae* [18]. The leader sequence of tomato PG, which is much longer than the prepro-sequence of the fungal PGII and which does not show sequence homology herewith, has been suggested to play a role in protein targeting [15,16]. It is now possible to construct specific mutant *pga*II genes in vitro and to introduce and express these genes in *A. niger*. This makes it feasible to study the possible function of the prosequence of PGII during secretion.

Acknowledgement: This work is part of a joint programme between the Biotechnology Department of Ciba-Geigy in Basel and the Agricultural University, Wageningen.

REFERENCES

[1] Rombouts, F.M. and Pilnik, W. (1980) in: Microbial Enzymes and Bioconversions, Economic Microbiology, vol. 5 (Rose, A.H. ed) pp. 227- 282, Academic Press, London.
[2] Rexova-Benkova, L. and Markovic, O. (1976) Adv. Carbohydr. Chem. Biochem. 33, 323–385.
[3] Kester, H.C.M. and Visser, J. (1990) Biotechnol. Appl. Biochem. 12, 150- 160.
[4] Berka, R.M. and Barnett, C.C. (1989) Biotechnol. Adv. 7, 127–154.
[5] Chang, J.Y. (1983) Methods Enzymol. 91, 79–84.
[6] Matsudaira, P. (1987) J. Biol. Chem. 262, 10035–10038.
[7] Amons, R. (1987) FEBS Lett. 212, 68–72.
[8] Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular Cloning, A Laboratory Manual, Cold Spring Harbour Laboratory Press, Cold Spring Harbour, NY.
[9] Dente, L. and Cortese, R. (1987) Methods Enzymol. 155, 111–119.
[10] Goosen, T., Bloemheuvel, G., Gysler, C., De Bie, D.A., Van den Broek, H.W.J. and Swart, K. (1987) Curr. Genet. 11, 499–503.
[11] Goosen, T., Van Engelenburg, F., Debets, F., Swart, K., Bos, K. and Van den Broek, H. (1989) Mol. Gen. Genet. 219, 282–288.
[12] Vishniac, W. and Santer, M. (1957) Bacteriol. Rev. 21, 195–213.
[13] Ohtsuka, E., Matsuki, S., Ikehara, M., Takahashi, Y. and Matsubara, K. (1985) J. Biol. Chem. 260, 2605–2608.
[14] Von Heyne, G. (1986) Nucleic Acids Res. 14, 4683–4690.
[15] Grierson, D., Tucker, G.A., Keen, J., Ray, J., Bird, C.R. and Schuch, W. (1986) Nucleic Acids Res. 14, 8595–8603.
[16] Sheehy, R.E., Pearson, J., Brady, C.J. and Hiatt, W.R. (1987) Mol. Gen. Genet. 208, 30–36
[17] Cervone, F., De Lorenzo, G., Salvi, G. and Camardella, L. (1986) in: Biology and Molecular Biology of Plant-Pathogen Interactions, NATO ASI Series, vol. H1 (Baily, J. ed) pp. 385–392, Springer Verlag, Berlin, Heidelberg.
[18] Innes, M.A., Holland, M.J., McGabe, P.C., Cole, G.E., Wittman, V.P., Tal, R., Watt, K.W.K., Gelfand, D.H., Holland, J.P. and Meade, J.H. (1985) Science 228, 21–26.