Biotechnology Letters **21:** 395–399, 1999.

Cloning and characterization of a gene encoding the endopolygalacturonase of *Penicillium griseoroseum*

A.O.B. Ribon[1], J.L.C. Coelho[1], E.G. de Barros[2] & E.F. Araújo[1,*]
[1]*Departamento de Microbiologia, Universidade Federal de Viçosa, Viçosa, MG, Brazil*
[2]*Departamento de Biologia Geral, Universidade Federal de Viçosa, Viçosa, MG, 36571-000, Brazil*
[*]*Author for correspondence (Fax: +55 31 8992573; E-mail: ezfa@mail.ufv.br)*

Received 29 January 1999: Revisions requested 8 February 1999; Revisions received 8 March 1999; Accepted 9 March 1999

Key words: filamentous fungi, gene cloning, pectinase, sequence homology

Abstract

A conserved region of a polygalacturonase (PG) gene from *Penicillium griseoroseum* was PCR amplified and used to screen a genomic library from this fungus. The nucleotide sequence of the isolated clone (*pgg*I) consisted of 1497 bp, including a coding region of 1251 bp. This region potentially encodes a protein of 376 amino acids, and is interrupted by two introns. Extensive homology was observed between this protein and several fungal endopolygalacturonases. DNA hybridization analyses revealed that there is a low copy number of *pgg*I in the *P. griseoroseum* genome, probably one or two copies.

Introduction

Pectinases degrade the major constituent found in the middle lamela and primary cell walls of plant cells. Fungal pectinases used in the food industry for clarification of fruit juices and wines, liquefaction of fruit and vegetable pulps (Manachini *et al.* 1988, Aguilar & Huitron 1986, Grassin & Fauquembergue 1996). Researches have been conducted to obtain pectinase overproducing strains of filamentous fungi which can be used in the textile industry for degumming of natural fibers extracted from flax, jute, and ramie (Sharma 1987, Fernandes-Salomão *et al.* 1996). The enzymatic degumming of natural fibers is less polluting than the chemical process normally used, and it reduces the use of chemicals and energy (Sharma 1987).

The use of fungal pectinases in the textile industry requires the selection of good pectinase producers, showing low cellulase secretion to prevent the weakening of the natural fibers. Two species from the genus *Penicillium*, *P. griseoroseum* and *P. expansum*, have been selected in our laboratory based on their pectinase production (Baracat *et al.* 1989). These species have been extensively studied both physiologically (Picolli-Valle *et al.* 1995, Baracat-Pereira *et al.* 1997) and genetically (Fernandes-Salomão *et al.* 1996). At present, three isoforms can be detected during the partial purification of PG produced by *P. griseoroseum*. Preliminary molecular studies conducted in our laboratory have shown that PG genes from this fungus are differentially expressed depending on the nature of the carbon source present in the growth medium. One of our goals is to obtain *P. griseoroseum* strains that secrete large amounts of PG using alternative carbon sources as inducers, so that the enzymatic process can be economically attractive to the textile industry.

Initially, a heterologous transformation system was developed for genetic studies of *P. griseoroseum* (Queiroz *et al.* 1998). We are now cloning and characterizing genes encoding *P. griseoroseum* pectinases. Here we report the cloning and characterization of a polygalacturonase gene (*pgg*I) from *P. griseoroseum*.

Materials and methods

Isolation of genomic clones

Genomic DNA from *Penicillium griseoroseum* was extracted and partially cleaved with *Sau*3A and frag-

ments ranging from 15 to 20 kb were cloned into the λEMBL3 vector (Stratagene, USA). To screen the library, a 400-bp fragment from *P. griseoroseum* was used. This fragment, which is part of a conserved region present in genes encoding PG in different species, was PCR amplified using the genomic DNA from *P. griseoroseum* as template. The primers and PCR conditions were the same employed by Cary *et al.* (1995). Plaque lifts were prepared as described by Benton & Davis (1977). Hybridization was performed in 5 × SSC (1 × SSC is 0.15 M NaCl plus 0.015 M trisodium citrate, pH 7.6)/0.5% SDS/Denhardt's reagent and 100 μg salmon sperm/ml at 65 °C for 18 h. Filters were washed once in 4 × SSC/0.1% SDS, twice in 2 × SSC/0.1% SDS and once in 1 × SSC/0.1% SDS at 63 °C. The identities of the positives clones were confirmed by hybridization with plasmid pCC2 that harbors the endoPG gene from *Fusarium moniliforme* (Caprari *et al.* 1993). Recombinant phage DNA was extracted according to Sambrook *et al.* (1989).

Restriction mapping, subcloning and DNA sequencing

The clones were mapped using the strategy of single and multiple restriction enzyme digestions, followed by Southern blot analysis. Genomic fragments were subcloned into pUC18, according to standard procedures (Sambrook *et al.* 1989). The nucleotide sequence was determined by the dideoxynucleotide chain-termination method (Sanger *et al.* 1977) using the Thermo Sequenase dye terminator cycle sequencing pre-mix kit according to the recommendations of the supplier (Amersham Life Science, USA).

Southern blot analysis

Genomic DNA (6 μg) was digested with restriction enzymes and subjected to electrophoresis on a 0.8% agarose gel (Sambrook *et al.* 1989). DNA was transferred to a nylon membrane using 10 × SSC as transfer buffer. The membrane was probed with the 1.5 kb *Xba*I fragment, labeled with d-UTP fluorescein (Stratagene, USA). Hybridization and washes were carried at 65 °C. The filter was washed once in 2 × SSC/0.1% SDS, once in 1 × SSC/0.1% SDS and twice in 0.1 × SSC/0.1% SDS.

Results and discussion

The titer of the genomic library was 1.13×10^6 p.f.u./ml. About 60,000 plaques were analyzed and six positives clones were detected. Two clones (λPG25 and λPG31) were selected for further analyses based on their restriction patterns. Two *Xba*I fragments of 2.2 kb and 1.5 kb derived from λPG25 and λPG31, respectively, hybridized with the 400-bp probe, and were subcloned into pUC 18. The clones were nominated pPG15 and pPG21, respectively.

The complete sequence of the 1.5 kb *Xba*I fragment was determined (Figure 1). A putative TATA box (TATATAA) was found at position −96. The sequence surrounding the translation start codon (AGC**ATG**G) resembles the consensus sequence reported for higher eukaryotes normally found at this site (Kozak 1984). The coding region consists of 1251 bp potentially encoding a protein of 376 aa, after splicing of the putative introns. The existence of two introns (57 bp and 66 bp) was deduced based on sequence agreement with fungal 5′ and 3′ splice consensus sequences, GTANGT and PyAG, respectively (Gurr *et al.* 1987). The internal intron recognition sequence is also present in both introns (NPuCTPuAC) and the intron sizes coincide with those reported for other filamentous fungi (Rambosek & Leach 1987). Additionally, in-frame stop codons were observed in both introns.

The *pgg*I deduced amino acid sequence has a potential N-glycosylation site found at amino acid 307. This feature is also found in other sequenced PG genes (Ishida *et al.* 1997, Parenicová *et al.* 1998).

Comparison of deduced amino acid sequences of endoPG from *Penicillium griseoroseum*, *Colletothrichum lindemuthianum*, *Cochliobolus carbonum*, *Sclerotinia sclerotiorum*, *Penicillium janthinellum*, *Aspergillus flavus*, *A. parasiticus* and *A. niger* is shown (Figure 2). *Penicillium griseoroseum* PG shares 64% amino acid identity with PG from *P. janthinellum*, *C. carbonum* and *C. lindemuthianum*. Amino acid identity of 61% was found between *P. griseoroseum* PG and the *Aspergillus* species mentioned above. *P. griseoroseum* PG showed the highest degree of identity (67%) with PG from *S. sclerotiorum*. The specific sequence for fungal pectinases CXGGHGXSIGSVG (Centis *et al.* 1996) was also found in the amino acid sequence of *pgg*I.

The *pgg*I gene was used to probe genomic DNA of *P. griseoroseum* cleaved with different enzymes (Figure 3). Restriction with five different enzymes

```
TCCAGTTCGTTTCAGGATGAACCGTAGAGGTATATAAGACTCCCCTCGTCTGCTCTTTGC  -73

TTGAAAATCATAAACAAACCATCACACTCTTTTAAAACCCCACATTCATTTCTATTTCAT  -13

ATTTTCTTTATCATGGCCTCCAGTCTCAAACTTGGCTTGATTGCCCTACTCGGTGCGACT   48
            M  A  S  S  L  K  L  G  L  I  A  L  L  G  A  T    16

GCTGTCAACGCAGCTCCCGCCGCTGAGCCCGTTCTTGGAACCTCTCTCCTTACTTCTCGT  108
 A  V  N  A  A  P  A  A  E  P  V  L  G  T  S  L  L  T  S  R    36

GCTTCTTGTACCTTCTCCGGCTCCAGCGGTGCTGCGGAAGCCATCAAGAGCAAGACCTCT  168
 A  S  C  T  F  S  G  S  S  G  A  A  E  A  I  K  S  K  T  S    56

TGCTCAACCATTACCCTGAGCAACGTCGAAGTCCCGGCCGGAACTACTCTGGACCTGACA  228
 C  S  T  I  T  L  S  N  V  E  V  P  A  G  T  T  L  D  L  T    76

GGACTCAAGTCCGGTACCACgtaagctcatcacgatctccaacctcaaatctccatacta  288
 G  L  K  S  G  T  T                                            83

aaccttatctcgcatagCGTCATCTTCGAAGGAACCACCACCTTCGGTTACAAGGAATGG  348
                  V  I  F  E  G  T  T  T  F  G  Y  K  E  W     97

GAAGGTCCTCTGGTCTCCGTCTCCGGAACCTCCATCACCGTCCAGGGTGCCTCAGGCGCG  408
 E  G  P  L  V  S  V  S  G  T  S  I  T  V  Q  G  A  S  G  A   117

CAGCTCAACGGTGATGGTGCCCGCTGGTGGGACGGCAAGGGAACCAACGGCGGCAAGACC  468
 Q  L  N  G  D  G  A  R  W  W  D  G  K  G  T  N  G  G  K  T   137

AAGCCCAAGTTCTTCTACGCTCACAGCTTGACCAACTCCAAGATCGAGAACATCTACATC  528
 K  P  K  F  F  Y  A  H  S  L  T  N  S  K  I  E  N  I  Y  I   157

AAGAACTCCCCCGTGCAGGTCTTCAGCATCAACGGCGCCAAGGAGCTGACTCTTAGCGGA  588
 K  N  S  P  V  Q  V  F  S  I  N  G  A  K  E  L  T  L  S  G   177

ATCACCGTCGACACCGCTGACGGCGATAGCAACGGCGGCCACAACACCGACGCTTTCGAC  648
 I  T  V  D  T  A  D  G  D  S  N  G  G  H  N  T  D  A  F  D   197

GTCGGCTCCAGCAACGGTGTCTACATCACCAGCCCTATCGTCCACAACCAGGATGACTGC  708
 V  G  S  S  N  G  V  Y  I  T  S  P  I  V  H  N  Q  D  D  C   217

CTGGCTGTCAACTCCGGCACTgtatgtacataccgacaaattgggacttatcttatttca  768
 L  A  V  N  S  G  T                                           224

tgaactttaactaactgatattcccagAACGTCCACTTCACTGGCGCTCAGTGCACTGGC  828
                           N  V  H  F  T  G  A  Q  C  T  G     235

GGCCACGGTATCTCCATCGGCTCCGTCGGTGGACGCTCCGACAACACCGTTGACGGTGTC  888
 G  H  G  I  S  I  G  S  V  G  G  R  S  D  N  T  V  D  G  V   255

ACCGTCGAGAGCTGCACCATCAAGGACTCCGACAACGGCGTCCGCATCAAGACCGTCTAC  948
 T  V  E  S  C  T  I  K  D  S  D  N  G  V  R  I  K  T  V  Y   275

GGCGCCACCGGTACAGTCCAAGGCGTCACCTACAAGGACATCACTCTTTCCGGAATTGCC 1008
 G  A  T  G  T  V  Q  G  V  T  Y  K  D  I  T  L  S  G  I  A   295

AAGTACGGTATCGTCATCGAGCAGGATTACGAGAACGGTAGCCCTACCGGAACCCCCACC 1068
 K  Y  G  I  V  I  E  Q  D  Y  E  N  G  S  P  T  G  T  P  T   315
                                    *

AGCGGTGTCCCCATTACCGACTTGACCCTCGACAACGTCCACGGAACTGTCGCCAGCAGT 1128
 S  G  V  P  I  T  D  L  T  L  D  N  V  H  G  T  V  A  S  S   335

GGTGTCGATACCTACATCCTGTGTGCTTCTGGTGCTTGCTCCGACTGGTCTTGGAGCGGC 1188
 G  V  D  T  Y  I  L  C  A  S  G  A  C  S  D  W  S  W  S  G   355

GTTAGCATCACCGGTGGTCAGACTAGCAAGAAGTGCAAGGGTATTCCCAGCGGTGCTAGC 1248
 V  S  I  T  G  G  Q  T  S  K  K  C  K  G  I  P  S  G  A  S   375

TGCTAAATTCTGCGAATTTTGAGTCGGACTGTGTAGGCCTTTTTAACTCTTTTAATTGGG 1308
 C                                                             376

GTTTTTGTTGTCTACTTGTTACCCACCCCGTTGTGTATACTATTTTCCCCACTGGCAA   1366
```

Fig. 1. Nucleotide and deduced amino acid sequence of the *Penicillium griseoroseum* endopolygalacturonase *pgg*1 gene. Putative TATA box is in bold characters and underlined. ATG initiation codon and TAA stop codon are boxed and in bold characters. Putative introns are in lower case. A putative N-glycosilation site (∗) is indicated. The nucleotide sequence data reported in this paper was deposited in the GenBank Database under accession No. AF085238.

```
Pgriseoros ELTLSGITVDTADGDSNGGHNTDAFDVGSSNGVYITSPIVHNQDDCLAVNSGTNVHFTGA 231
Pjanthinel NLIIDGVTIDNSDGDENGGHNTDGFDISESTGVTIRNAVVKNQDDCIAINSGQNIYFTGG 227
Clindemuth NLGVYDVSVDNSAGDSAGGHNTDAFDVGSSTGVYISGADVKNQDDCLAVNSGTNITFTGG 218
Ccarbonum  GLTLDRITIDNSAGDSAGAHNTDAFDIGSSSGITISNANIKNQDDCVAINSGSDIHVTNC 219
Ssclerotio GLTLSGITIDNSAGNSLG-HNTDAFDVGSSTDITISGANVQNQDDCLAINSGTGITFTGG 238
Aflavus    NLNITDVTIDNSAGTAEG-HNTDAFDVGSSTYINIDGATVYNQDDCLAINSGSHITFTNG 218
Aparasitic NLNITDVTIDNSAGTAEG-HNTDAFDIGSSTYINIDGATVYNQDDCLAINSGSHITFTNG 218
Aniger     YLTMEDITIDNTDGDDGEAANTDGFDIGDSTYITITGANVYNQDDCVAVNSGENIYFSGG 235
            * .  ...*.. *      *** **...*. . *  . . *****.*.***  .  ..

Pgriseoros QCTGGHGISIGSVGGRSDNTVDGVTVESCTIKDSDNGVRIKTVYGATGTVQGVTYKDITL 291
Pjanthinel TCSGGHGLSIGSVGGRDDNTVKNVTITDSTVTDSANGVRIKTVYDATGSVSDVTFSDITV 287
Clindemuth TCSGGHGLSIGSVGGRKDNVVKSVSITNSKIINSDNGVRIKTVAGATGPVSDITYSGITL 278
Ccarbonum  QCSGGHGVSIGSVGGRKDNTVKGVVVSGTTIANSDNGVRIKTISGATGSVSDITYENITL 279
Ssclerotio TCSGGHGLSIGSVGGRSDNVVSDVIIESSTVKNSANGVRIKTVLGATGSVSGVTYKDITL 298
Aflavus    YCDGGHGLSIGSVGGRSDNTVEDVTISNSKVVNSQNGVRIKTVYDATGTVSNVKFEDITL 278
Aparasitic YCDGGHGLSIGSVGGRSDNTVEDVTISNSKVVNSQNGVRIKTVYGATGTVSNVKFEDITL 278
Aniger     VCSGGHGLSIGSVGGRSDNTVKNVTFYDSDIKSSQNGVRIKTIYGDTGSVSEVTYKEITL 295
            * ****.********.**.*..*      . .* *******. . **.*. .... **.

Pgriseoros SGIAKYGIVIEQDYENGSPTGTPTSGVPITDLTLDNVHGTVASSGVDTYILCASGACSDW 351
Pjanthinel SGITDYGIVIEQDYENGSPTGTPTSGVPITDLTVKGITGSVESDAVEVYILCGDDACSDW 347
Clindemuth SNIAKYGIVIEQDYENGSPTGKPTSGVPISGLTLSKISGSVSSSATPVYILCA--SCTNW 336
Ccarbonum  KNIAKYGIVIEQDYLNGGPTGKPTTGVPITGVTLKNVAGSVTGSGTEIYVLCGKGSCSGW 339
Ssclerotio SGITSYGVVIEQDYENGSPTGKPTSGVPITGVTLSNVHGTVSSSATNVYVLCA--KCSGW 358
Aflavus    SGITKYGLIVEQDYENGSPTGTPTNGIKVSDITFDKVTGTVESDATDIYILCGSGSCTDW 338
Aparasitic SGITKYGLVVEQDYENGSPTGTPTNGITVSGITFEKVTGTVESDATDIYILCGSGSCTDW 338
Aniger     SDITDYGIVVEQNYDDTSES--PTDGITIEDFVLDNVQGSVESSGTNIYIVCGSDSCTDW 353
            ..*..**...**.* .   .  ** *. . . .   . *.* ....  *..*.   *..*
```

Fig. 2. Comparison of the deduced amino acid partial sequences of endopolyglacturonases produced by several fungi. The conserved domain of eight polygalacturonases is shown in bold characters. *Penicillium griseoroseum, P. janthinellum* (Ishida *et al.* 1997), *Colletotrichum lindemuthianum* (Centis *et al.* 1996), *Cochliobolus carbonum* (Scott-Craig *et al.* 1990), *Sclerotinia sclerotiorum* (Reymond *et al.* 1994), *Aspergillus flavus* (Whitehead *et al.* 1995), *A. parasiticus* (Cary *et al.* 1995), and *A. niger* (Parenicová *et al.* 1998). Identical (∗) and conserved (·) aa substitutions are indicated. Alignments were performed using the CLUSTAL software program (IntelliGenetics).

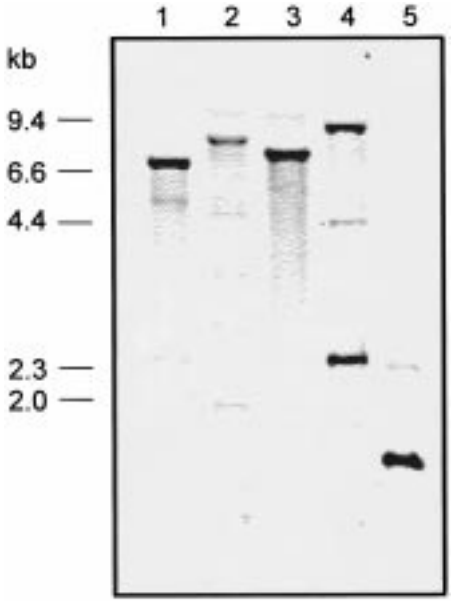

Fig. 3. DNA hybridization analysis of genomic DNA of *Penicillium griseoroseum* probed with *pgg*I gene, labeled with d-UTP fluorescein (Stratagene, USA). Genomic DNA (6 μg) was digested with *Eco*RI (1), *Kpn*I (2), *Sac*I (3), *Sal*I (4) and *Xba*I (5).

produced one or two hybridizing fragments each, suggesting that there is only one copy of the *pgg*I gene in the fungal genome. Restriction sites for *Kpn*I and *Sal*I in the coding region of *pgg*I (Figure 3, lanes 2 and 4, respectively) was seen by restriction mapping as well as sequencing. Weak hybridizing signals present in all lanes (Figure 3) were assumed be due to the presence of other putative PG gene(s). To confirm this hypothesis, the same membrane was probed with subclone pPG15 demonstrating that these signals become considerably stronger. Partial sequencing analysis of a 2.2 kb *Xba*I fragment derived from subclone pPG15 revealed extensive homology with several fungal endoPG. However, it is certainly not identical to the *pgg*I gene (data not shown).

Acknowledgments

We thank Dr Giulia de Lorenzo for providing plasmid pCC2. This research was supported by grants from CNPq, FINEP, FAPEMIG and BID.

References

Aguilar G, Huitron C (1986) *Enz. Microbiol. Technol*. **2**: 541–545.
Baracat MC, Valentin C, Muchovej JJ, Silva DO (1989) *Biotechnol. Lett*. **11**: 899–902.
Baracat-Pereira MC, Minussi RC, Coelho JLC, Silva DO (1997) *J. Ind. Microbiol. Biotechnol*. **18**: 308–311.
Benton, Davis (1977) *Screening of Xgt Recombinant Clones*.
Caprari C, Richter A, Bergmann C, Lo Cicero S, Salvi G, Cervone F, De Lorenzo G (1993) *Mycol. Res*. **97**: 497–505.
Cary JW, Brown R, Cleveland TE, Whitehead M, Dean RA (1995) *Gene* **153**: 129–133.
Centis S, Dumas B, Fournier J, Marolda M, Tugayé MTE (1996) *Gene* **170**: 125–129.
Fernandes-Salomão TM, Amorim ACR, Chaves-Alves VM, Coelho JLC, Silva DA, Araújo EF (1996) *Rev. Microbiol*. **27**: 15–18.
Grassin C, Fauquembergue P (1996) Application of pectinases in beverages. In: Visser J, Voragen AGJ, eds. Amsterdam: Elsevier Science B.V., pp. 453–462.
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*. Oxford: IRL Press, pp. 93–139.
Ishida Y, Kakibuchi K, Hirao Y, Izumori K (1997) *J. Ferment. Bioeng*. **84**: 257–260.
Kozak M (1984) *Nucleic Acids Res*. **9**: 5233–5256.
Manachini PL, Parini C, Fortina MG (1988) *Enzyme Microbiol. Technol*. **10**: 682–685.
Parenicová L, Benen JAE, Kester HCM, Visser J (1998) *Eur. J. Biochem*. **251**: 72–80.
Piccoli-Valle RH, Baracat-Pereira MC, Silva DO (1995) *J. Basic Microbiol*. **35**: 189–193.
Queiroz MV, Barros AO, Barros EG, Guimarães W, Araújo EF (1998) *Can. J. Microbiol*. **44**: 487–489.
Rambosek J, Leach J (1987) *CRC Crit. Rev. Biotechnol*. **6**: 357–393.
Reymond P, Deléage G, Rascle C, Fèvre M (1994) *Gene* **146**: 233–237.
Sambrook J, Fritsch EF, Maniatis T (1989) *Molecular Cloning: a Laboratory Manual*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.
Sanger F, Nicklen S, Coulson AR (1977) *Proc. Natl. Acad. Sci. USA* **74**: 5463–5467.
Scott-Craig JS, Panaccione DG, Cervone F, Walton JD (1990) *Plant Cell* **2**: 1191–1200.
Sharma HSS (1987) *Int. Biodeterioration* **23**: 181–186.
Whitehead MP, Shieh MY, Clevland TE (1995) *Appl. Environ. Microbiol*. **61**: 3316–3322.