PCR cloning by genome walking of a complete polygalacturonase-inhibiting protein gene from *Eucalyptus grandis*

R. Bhoora[a,c], D.K. Berger[b,c], M.J. Wingfield[c] and B.D. Wingfield[a,c*]

POLYGALACTURONASE-INHIBITING proteins (PGIPs), associated with the cell walls of many plant species, form part of the active defence of plants against phytopathogenic fungi. PGIPs are leucine-rich repeat proteins that are evolutionarily related to several plant resistance (*R*) genes, which participate in gene-for-gene interactions. The objective of this study was to clone and characterize the complete *Eucalyptus pgip* gene. Based on alignments between the pear *pgip* and partial *Eucalyptus pgip* isolated in a previous study,[8] gene-specific primers were designed and used to amplify a *pgip* fragment of 1020 base pairs from *Eucalyptus grandis* genomic DNA. Cloning and sequence analysis of the fragment revealed that it did not contain sequences upstream of the start codon or the 3′ termination codon. A genome-walking PCR strategy was applied to amplify the unknown regions of the *pgip* gene. Two positive clones designated as pCR-*Sca*IGW and pCR-*Dra*IGW were sequenced and the resulting sequences represented the unknown 5′ and 3′ regions, respectively. The complete *Eucalyptus pgip* gene contains a single open reading frame, which encodes a protein of 331 amino acids and contains features that are typical of PGIPs.

Introduction

Polygalacturonase-inhibiting proteins (PGIPs) are extracellular glycoproteins that are ionically bound to plant cell walls and are capable of inhibiting fungal, but not bacterial, endopolygalacturonases (PGs).[1] PGIPs form specific, reversible, saturable, high-affinity complexes with fungal endopolygalacturonases.[1] The current hypothesis is that the PGIP-PG interaction enhances the half-life of oligogalacturonide 'elicitors' from the plant cell wall, which are known to activate the plant defence responses.[1] They have been shown to be structurally related to several plant disease resistance genes, and they belong to a super-family of leucine-rich repeat proteins that are specialized for the recognition of non-self molecules and the rejection of pathogens.[2]

South African researchers have shown considerable interest in PGIPs in recent years with reports of studies from wheat,[3] grapevine,[4] cotton,[5] bean,[6] apple[7] and *Eucalyptus*.[8] PGIPs have been purified to homogeneity from several dicotyledonous plants, including bean,[9] soybean,[10] tomato,[11] pear,[12] apple,[13] raspberry,[14] potato[15] and, very recently, cotton[5] and guava.[16] They have also been identified in the pectin-rich monocots, onion and leek.[17]

The first gene encoding a PGIP was cloned from *Phaseolus vulgaris*.[18] *Pgip* genes have since been cloned from many other plant species, with the implementation of several polymerase chain reaction (PCR) techniques. Most *pgip* genes characterized contain a single open reading frame approximately 1 kilobase (kb) in length that is not interrupted by introns.[8,12,18] However, the raspberry (*Rubus idaeus*) *pgip* contains a predicted intron sequence.[19] A highly hydrophobic region corresponds to the N-terminal signal peptide, which targets PGIPs to the endomembrane system for export to the extracellular space.[12,18]

Several PCR techniques have been employed to isolate *pgip* genes from various plant sources. PCR is a powerful tool developed and used to amplify microgram amounts of a specific, targeted gene.[20] Typical PCR reactions use oligonucleotide primers that hybridize to opposite strands so that extension proceeds inwards between the two primers.[20] A limitation of the technique is the inability to amplify regions that lie outside the boundaries of known sequences without the use of several major molecular modifications that include cloning and ligation of adaptors to the ends of amplified sequences.[20] Several methods have since been developed for 'walking' from a known DNA region to an unknown region in both cloned and uncloned DNA. These include inverse PCR, randomly primed PCR, adaptor-ligation PCR and genome-walking PCR.[20]

The molecular cloning and characterization of a *Eucalyptus grandis pgip* gene is described here. In a previous study[8] a partial *E. grandis pgip* gene was isolated. Sequence analysis revealed that the predicted translation of this partial gene showed a high degree of amino acid identity with the corresponding part of the pear PGIP, although it lacked the N-terminus including the signal peptide and the C-terminus of the protein. Using the genome-walking technology, we chose to 'walk' upstream as well as downstream of the known *pgip* sequence, in order to determine the sequences of the 5′ and 3′ ends of the gene. Isolation of the complete *pgip* gene is the first step towards understanding the role of this gene's product in defence against fungal pathogens of *Eucalyptus*.

Materials and methods

Genomic DNA was extracted from 100 mg of fresh, young leaves collected from a *E. grandis* clone (TAG5) using the DNeasy Plant Mini Kit (Qiagen, Germany) as recommended by the manufacturer's protocol. Gene-specific primers PC6 (5′-ACATCTCTCAGGCTCTCA ACC-3′) and SPPGIP2 (5′-GCAGTGTG GAGGGGTGCACCACACAGGCA-3′) designed from the conserved regions of pear[12] and *Eucalyptus*[8] *pgip* sequences, respectively, enabled the amplification of a *E. grandis pgip* fragment of approximately 1.0 kb. The PCR product was gel-purified using the QIAquick Gel Extraction Kit (Qiagen) and cloned into the polylinker region of the pGEM-T-Easy commercial cloning vector. T7 and SP6 primers were used to sequence the double-stranded DNA insert of a positive recombinant clone designated as pGEM-*Eucpgip*110B. Sequencing was achieved, using the BIG Dye terminator cycle sequencing kit, with an ABI Prism model 3100 automated sequencer (Perkin-Elmer, DNA sequencing facility, Department of Genetics, University of Pretoria).

A BLASTX search using the partial *E. grandis pgip* sequence from pGEM-*Eucpgip*110B was conducted to determine regions of homology to other published PGIP polypeptide sequences. Sequence comparisons revealed the absence of a 3′ downstream region on the *E.grandis pgip* fragment [CLUSTAL (EBI) software http://www2.ebi.uk]. Genome walking adaptors designed according to those described by Siebert *et al.*[20] were ligated to blunt-end digested *E. grandis* (clone TAG5) genomic DNA. Primer pairs, N-PGIP-L1 (5′-CCATGAGGCCAAGACAT AGG-3′) and PGIP-Inv-L2 (5′-GGACGG

Departments of [a]Genetics and [b]Botany, and [c]Forestry and Agricultural Biotechnology Institute (FABI), University of Pretoria, Pretoria 0002, South Africa.
*Author for correspondence.
E-mail: brenda.wingfield@fabi.up.ac.za

AGGAGAAGAGTAGG-3′) and N-PGIP-R1-1 (5′-CTGAACGTGAGCTACAA CAG-3′) and PGIP-Inv-R3 (5′-CCATAA CCGATGCCTGTGTG-3′) were designed in opposite orientation to normal PCR primers using the pGEM-*Eucpgip*110B sequence data. Primary PCR reactions were performed using genome-walking adaptor primer 1 (ADP1), designed as described by Siebert *et al.*[20] and N-PGIP-L1 for amplifying the upstream 5′ region of the gene, and primers ADP1 and N-PGIP-R1-1 were used for amplifying the downstream 3′ region. Secondary PCR reactions were performed using adaptor primer 2 (ADP2) and PGIP-Inv-L2 for the upstream region and primers ADP2 and PGIP-Inv-R3 for amplifying the downstream 3′ region. The PCR products were purified from the gel and ligated into the multiple cloning site of the pCR®2.1 vector provided in the TA® Cloning Kit (Invitrogen). The ligated samples were transformed into One Shot® cells (TOP10F′). Two positive clones designated pCR-*Sca*IGW-4 (upstream fragment) and pCR-*Dra*I-GW2 (downstream fragment) were sequenced using the M13 forward and M13 reverse primers.

The DNA sequences obtained were further analysed for homology to the original *Eucalyptus pgip* sequence and the potential presence of control regions. The theoretical molecular weight and isoelectric point for the mature *E. grandis* PGIP were determined using the algorithm from ExPASy's Compute pI/Mw program (http://scansite.mit.edu/calc_mw_pi.html). Databases used for *in silico* promoter prediction were: PlantCARE[21] (http://intra.psb.ugent.be:8080/PlantCARE/); PLACE[22] (http://www.dna.affrc.go.jp/htdocs/PLACE/signalscan.html); PlantProm DB (http://www.softberry.com/berry.phtml?topic=promoter); Neural Network Promoter Prediction[23] (http://www.fruitfly.org/seq_tools/promoter.html); the Markov Chain Promoter Prediction Server[24] (http://genes.mit.edu/McPromoter.html).

Results

Nucleotide sequence analysis of the recombinant pGEM-*Eucpgip*110B clone revealed that amplification with the gene-specific PC6 primer enabled determination of the start codon (ATG) and signal sequence of PGIP, which was not included in the partial *Eucalyptus* PGIP fragment sequenced previously.[8] The genome-walking strategy of Siebert *et al.*[20] was thus employed to amplify the *E. grandis pgip* upstream and downstream regions. Restriction enzyme digestion of TAG5 genomic DNA with *Sca*I enabled the amplification of the 5′ upstream region, using adaptor primer (ADP2) and PGIP-Inv-L2, while digestion with *Dra*I led to the amplification of the downstream 3′ region when ADP2 and PGIP-Inv-R3 were used.

Recombinant plasmids pCR-ScaIGW-4 and pCR-DraI-GW2 were selected and sequenced following PCR and restriction enzyme screening. Based on alignment of these sequences to the upstream and downstream regions of the pGEM-*Eucpgip*110B clone, we concluded that the inserts from pCR-*Sca*IGW-4 and pCR-*Dra*I-GW2 represented the 5′ and 3′ regions of the *E. grandis pgip* gene, respectively. A 162-bp sequence was determined from the pCR-*Sca*IGW-4 clone using the M13 forward and M13 reverse primers. Alignment with the partial *E. grandis pgip* sequence showed that the genome-walking clone shared an overlap of 91 bp at the 5′ region of the gene. Only one nucleotide difference, which results in a silent mutation, was observed within the signal sequence region.

DNA sequence information for 44 bp upstream of the *pgip* start codon was obtained from pCR-*Sca*IGW-4. This region was searched *in silico* for a predicted promoter and transcription initiation site. On account of genome sequencing projects, algorithms have been developed for prediction of promoters from raw sequence data.[24] Several of these have been made available through the Internet and some are based on databases of experimentally proven plant promoters (see Materials and methods). No RNA Polymerase II binding site (TATA-box) or transcription initiation site could be identified, since it is likely that the promoter is further upstream.

The pCR-*Dra*I-GW2 clone contained a 693-bp sequence in the downstream region of the partial *E. grandis pgip* gene. There were only 39 bp that shared an overlap with the pGEM-*Eucpgip*110B clone. The translation termination codon (TAG) was observed in the nucleotide sequence of pCR-*Dra*I-GW2.

The 5′ and 3′ DNA sequences obtained from genome-walking PCR were linked *in silico* to the partial *E. grandis pgip* sequence from pGEM-*Eucpgip*110B to give the composite *E. grandis* PGIP sequence, which has been deposited in Genbank (accession number AY445043).

Discussion and conclusion

Polygalacturonase-inhibiting proteins, associated with the cell wall of many plant species, form part of the active defence of plants against phytopathogenic fungi. The composite *E. grandis pgip* gene reported in this study comprises a single open reading frame of 996 bp that is uninterrupted by introns. It encodes a polypeptide of 331 amino acids. The *E. grandis* PGIP amino acid sequence showed 99%, 93%, 49% and 46% identity with sequences of pear, apple, bean and soybean PGIPs, respectively.[1] This was reflected in the conservation of several features, including the potential processing site for signal peptidase, 10 leucine-rich repeats, three cysteine residues at the C-terminus and several potential glycosylation sites.[1] The mature protein is predicted to be 304 amino acids in size with a calculated molecular mass of 34 kDa and an isoelectric point of 6.2 (see Materials and methods).

The complete polypeptide sequence of the *E. grandis* PGIP differs from the published pear polypeptide sequence by only three amino acids (one within the signal peptide and two at the C-terminus). The *E. grandis* sequence obtained in this study contained the amino acids TSID at position 191 of the mature polypeptide, which are identical to those found in both pear PGIP and apple PGIP1. In contrast, the partial *Eucalyptus pgip* sequence reported previously had four different amino acids (GKHR) at this position.[8] There are another eight amino acids that differ from the partial *Eucalyptus* sequence.[8]

Several explanations can be offered to explain differences in amino acid sequence between the partial *Eucalyptus* PGIP[8] and the complete *E. grandis* PGIP obtained in this study. The most obvious reason could be that different *E. grandis* clones were used in each study, which could possibly result in a difference in the polypeptide sequence. Alternatively, the differences observed could be attributed to the fact that a different *PGIP* allele or *pgip* gene family member was sequenced in the previous study.[8] The type of polymerase used to amplify the *pgip* gene also plays a significant role in whether mutations are incorporated in the gene sequence. In the present study, the DNA sequence electrophoretograms were analysed carefully in this region. Furthermore, the sequence of this region was confirmed by direct sequencing from genomic *E. grandis* (TAG5) DNA. It is possible that in the previous study,[8] this region may have been difficult to read by the automated sequence prediction software. This would have resulted in a frameshift over this region by the insertion or deletion of bases before and after this region. However,

Chimwamurombe *et al.*[8] obtained the same predicted sequence (GKHR) for *pgip*s isolated from four other *Eucalyptus species* , thus adding weight to the view that different *PGIP* alleles or *pgip* gene family members were previously sequenced.

It is interesting that the PGIP from *Eucalyptus* , a member of the Myrtaceae, shows a high degree of sequence similarity to pear and apple PGIPs, both plants in the Rosaceae. The PGIPs from *Eucalyptus* that have been characterized are the only PGIPs that have been described for members of the Myrtaceae.

Determination of the complete sequence of a *E. grandis pgip* gene in this study paves the way for the characterization of a single PGIP from this plant. This could be achieved by transformation and expression of the PGIP in a test plant such as tobacco. The specificity of this PGIP against PGs from different fungal pathogens of *Eucalyptus* could then be tested. This would overcome the limitations of studies with PGIP extracts from *Eucalyptus*, in which there may be mixtures of different PGIPs. If PGIP activity can be demonstrated, the gene could then be used in genetic modification strategies to enhance resistance of *Eucalyptus* against targeted fungi.

This work was supported by the National Research Foundation, the Forestry Molecular Biology Cooperative Programme, the Tree Pathology Cooperative Programme and THRIP.

1. De Lorenzo G., D'Ovidio R. and Cervone F. (2001). The role of polygalacturonase-inhibiting proteins (PGIPs) in defense against pathogenic fungi. *Annu. Rev. Phytopathol.* **39**, 313–335.
2. Jones D.A. and Jones J.D.G. (1997). The role of leucine-rich repeats in plant defences. *Adv. Bot. Res.* **24**, 89–167.
3. Kemp G. (2001). *Isolation and characterization of a possible polygalacturonase-inhibiting protein from wheat.* Ph.D. thesis, University of the Free State, Bloemfontein.
4. De Ascensao A. (2001). *Isolation and characterisation of polygalacturonase inhibiting protein and its encoding gene from* Vitis vinifera L. Ph.D. thesis, Stellenbosch University, South Africa.
5. James J.T. and Dubery I.A. (2001). Inhibition of polygalacturonase from *Verticillium dahliae* by a polygalacturonase inhibiting protein from cotton. *Phytochemistry* **57**, 149–156.
6. Berger D.K., Oelofse D., Arendse M.S., Du Plessis E. and Dubery I.A. (2000). Bean polygalacturonase inhibitor protein-1 (PGIP-1) inhibits polygalacturonases from *Stenocarpella maydis*. *Physiol. Molec. Plant Pathol.* **57**, 5–14.
7. Arendse M.S., Dubery I.A. and Berger D.K. (1999). Isolation by PCR-based methods of a plant antifungal polygalacturonase-inhibiting protein gene. *Electronic J. Biotech.* **2**, 152–159.
8. Chimwamurombe P.M., Botha A-M., Wingfield M.J. and Wingfield B.D. (2001). Molecular relatedness of the polygalacturonase-inhibiting protein genes in *Eucalyptus* species. *Theoret. Appl. Genet.* **102**, 645–650.
9. Salvi G., Giarrizzo F., De Lorenzo G. and Cervone F. (1990). A polygalacturonase-inhibiting protein in the flowers of *Phaseolus vulgaris* L. *J. Plant Physiol.* **136**, 513–518.
10. Favaron F., D'Ovidio R., Porceddu E. and Alghisi P. (1994). Purification and molecular characterisation of a soybean polygalacturonase-inhibiting protein. *Planta* **195**, 80–87.
11. Stotz H.U., Contos J.J.A., Powell A.L.T., Bennett A.B. and Labavitch J.M. (1994). Structure and expression of an inhibitor of fungal polygalacturonases from tomato. *Plant Molec. Biol.* **25**, 607–617.
12. Stotz H.U., Powell A.T., Damon S.E., Greve L.C., Bennett A.B. and Labavitch J.M. (1993). Molecular characterization of a polygalacturonase inhibitor from *Pyrus communis* L. cv Bartlett. *Plant Physiol.* **102**, 133–138.
13. Yao C., Conway W.S. and Sams C.E. (1995). Purification and characterization of a polygalacturonase-inhibiting protein from apple fruit. *Phytopathology* **85**, 1373–1377.
14. Johnston D.J., Ramanathan V. and Williamson B. (1993). A protein from immature raspberry fruits which inhibits endopolygalacturonases from *Botrytis cinerea* and other micro-organisms. *J. Exp. Bot.* **44**, 971–976.
15. Machiandiarena M.F., Olivieri F.P., Daleo G.R. and Oliva C.R. (2001). Isolation and characterisation of a polygalacturonase-inhibiting protein from potato leaves. Accumulation in response to salicylic acid, wounding and infection. *Plant Physiol. Biochem.* **39**, 129–136.
16. Deo A. and Shastri N.V. (2003). Purification and characterization of polygalacturonase-inhibitory proteins from *Psidium guajava* Linn. (guava) fruit. *Plant Sci.* **164**, 147–156.
17. Favaron F., Castiglioni C., D'Ovidio R. and Alghisi P. (1997). Polygalacturonase inhibiting proteins from *Allium porrum* L. and their role in plant tissue against fungal endopolygalacturonases. *Physiol. Molec. Plant Pathol.* **30**, 403–417.
18. Toubart P., Desiderio A., Salvi G., Cervone F. and De Lorenzo G. (1992). Cloning and characterisation of the gene encoding the endopolygalacturonase-inhibiting protein (PGIP) of *Phaseolus vulgaris* L. *The Plant Journal* **2**, 367–373.
19. Ramanathan V., Simpson C.G., Thow G., Iannetta P.P.M., McNicol R.J. and Williamson B. (1997). cDNA cloning and expression of polygalacturonase-inhibiting proteins (PGIPs) from red raspberry (*Rubus idaeus*). *J. Exp. Bot.* **48**, 1185–1193.
20. Siebert P.D., Chenchik A., Kellogg D.E., Lukyanov K.A. and Lukyanov S.A. (1995). An improved PCR method for walking in uncloned genomic DNA. *Nucleic Acids Res.* **23**, 1087–1088.
21. Lescot M., Déhais P., Thijs G., Marchal K., Moreau Y., Van de Peer Y., Rouzé P. and Rombauts S. (2002). PlantCARE, a database of plant *cis*-acting regulatory elements and a portal to tools for *in silico* analysis of promoter sequences. *Nucleic Acids Res.* **30**, 325–327.
22. Higo K., Ugawa Y., Iwamoto M. and Korenaga T. (1999). Plant *cis*-acting regulatory DNA elements (PLACE) database. *Nucleic Acids Res.* **27**, 297–300.
23. Reese M.G. (2001). Application of a time-delay neural network to promoter annotation in the *Drosophila melanogaster* genome. *Comput. Chem.* **26**, 51–56.
24. Ohler U. and Niemann H. (2001). Identification and analysis of eukaryotic promoters: recent computational approaches. *Trends Genet.* **17**: 56–60.

✎ New Books

The following books have been just published, reissued or newly issued in paperback.

General Science

The Man Who Changed Everything: The Life of James Clerk Maxwell. By Basil Mahon. Pp. 254. Wiley. $27.95.

80 Not Out: My Autobiography. By Patrick Moore. Contender Books. £17.99. Britain's best-known amateur astronomer reminisces.

Inspiring Science: Jim Watson and the Age of DNA. Edited by John Inglis, Joseph Sambrook and Jan Witkowski. Pp. 503. Cold Spring Harbor Laboratory Press. $35.

The Third Man of the Double Helix: The Autobiography of Maurice Wilkins. By Maurice Wilkins. Pp. 274. Oxford University Press. $27.50.

The Maverick and His Machine. By Kevin Maney. Wiley. $29.95. The biography of Thomas J. Watson, the creator of IBM.

Measure for Measure: The Story of Imperial, Metric, and Other Units. By Alex Hebra. Pp. 232. Johns Hopkins University Press. $24.95.

I Have Landed. By Stephen Jay Gould. Vintage. £7.99 (pbk).

Don't Worry [It's Safe to Eat]. The True Story of GM Food, BSE and Foot and Mouth. By Andrew Rowell. Pp. 268. Earthscan. £16.99.

Science, Not Art: Ten Scientists' Diaries. Edited by Jon Turney. Pp.160. Calouste Gulbenkian Foundation. £8.50

Hitler's Scientists: Science, War and the Devil's Pact. By John Cornwell. Pp. 544. Viking. $29.95.

Prof: The Life of Frederick Lindemann. By Adrian Fort. Pp. 374. Jonathan Cape. £18.99. The absorbing account of the man who put Oxford's Clarendon Laboratory on the scientific map, and was Churchill's science adviser during the 1930s and the war that followed.

Politicizing Science: The Alchemy of Policymaking. Edited by Michael Gough. Pp. 313. Hoover Institution. $15. Essays that illustrate what can happen when scientific advice conflicts with the political interests of those in power.

Science, Money and Politics. By Dan Greenberg. University of Chicago Press. $20. The latest observations by the long-time Washington-watcher.

Megaprojects and Risk. By Bent Flyvbjerg, Nils Bruzelius and Werner Rothergatter. Cambridge University Press. £14.95. Why large-scale engineering projects quite often go expensively wrong.

Kepler's Conjecture: How Some of the Greatest Minds in History Helped Solve One of the Oldest Math Problems in the World. By George G. Szpiro. Pp. 304. Wiley. $24.95. The problem is to prove how identical spheres can be packed most densely (in a face-centred cubic lattice, in fact).

Biological Sciences

Spider Webs and Silks: Tracing Evolution from Molecules to Genes to Phenotypes. By Catherine L. Craig. Pp. 256. Oxford University Press. $59.95.

George Beadle, an Uncommon Farmer: The Emergence of Genetics in the 20th Century. By Paul Berg and Maxine Singer. Pp. 383. Cold Spring Harbor Laboratory Press. $35.

Complex Population Dynamics: A Theoretical/Empirical Synthesis. By Peter Turchin. Pp. 456. Princeton University Press. $75 (hbk); $29.95 (pbk).