Chapter One

Literature Review

Fungal Endopolygalacturonases and Polygalacturonase-inhibitors in Disease Resistance

CONTENTS

Introduction

Elucidating the role of plant defence at the molecular level is one of the pressing issues facing modern plant biologists (Walton, 1997). Plants are constantly being exposed to variety of potential enemies, these include, bacteria, fungi, viruses and nematodes (Walton, 1994). Unlike vertebrates, however, plants lack an immune system and they therefore have to rely on alternative mechanisms to resist attack (De Lorenzo and Cervone, 1997). Oligomers released from the polysaccharide during pathogen penetration acts as recognition elicitors for the plant. Early detection of the pathogen is essential for plant survival and depends greatly on the speed at which pathogen signal molecules are perceived by the plant and how quickly appropriate biochemical reactions can begin (Agrios, 1988).

Following microbial attack, plants are able to induce a number of biochemical defences. These include the synthesis of antimicrobial phytoalexins, the reinforcement of plant cell walls by the deposition of lignins and the accumulation of pathogenesis-related proteins (Lamb *et al.,* 1989). In plant-pathogen interactions, two forms of reactions can occur between the host and a pathogen, which determine either resistance or susceptibility. In incompatible reactions, the pathogen fails to infect the host due to the presence of plant resistance (*R*) genes, which recognises the pathogen elicitors and rapidly induce biochemical responses (Agrios, 1988). In contrast, compatible reactions lead to the onset of disease symptoms (Collinge and Slusarenko, 1987). Disease resistance (*R*) genes control gene-for-gene resistance (Bent, 1996). The concept of gene-for-gene resistance implies that for every gene that confers virulence to a pathogen, there must be a corresponding gene in the host that confers resistance to the host (Agrios, 1988).

Prior to the development of genetic mapping and cloning technologies, classical breeding methods and recurrent selection of desirable traits were utilised for establishing disease resistant plants (Agrios, 1988). Although molecular cloning and the ability to construct transgenic organisms have only recently been introduced, it was previously hypothesised that some forms of resistance were due to single specific genes. With the advent of molecular biology, researchers have been able to identify

and clone these resistance genes in order to determine whether certain biochemical functions are indeed involved in physiological processes such as plant disease resistance (Keen, 1999).

Several plant resistance *(R)* genes have been cloned. These include the *Hm1* gene of maize, the *Pto*, *Prf*, *Cf-9 and Cf-2* genes of tomato, *RPM1* and *RPS2* genes of *Arabidopsis*, the *Xa21* gene of rice, the *N* gene of tobacco and the *L6* gene of flax (Jones *et al.*, 1994; Bent *et al.*, 1996; Dixon *et al.*, 1996). A number of these *R*-genes, particularly from dicotyledonous plants, encode proteins that are structurally similar to leucine-rich repeat (LRR) domains that are crucial for the recognition of non-self molecules (Warren *et al.*, 1998; Komjanc *et al.*, 1999).

The accumulation of pathogenesis-related proteins such as the leucine rich polygalacturonase-inhibiting proteins (PGIPs), which form part of the cell wall receptors that respond to fungal elicitors, is an important point in plant-pathogen interactions. In this chapter, the role of endopolygalacturonases in pathogenicity, biochemical and molecular characterisation of PGIPs and the role of PGIPs in disease resistance are discussed.

The Plant Cell Wall And Fungal Invasion Strategies

The external cuticle and the polysaccharide-rich cell wall are the major physical barriers to potential pathogens of plant tissues (Walton 1994; De Lorenzo *et al.*, 1997). The complexity of the polysaccharides provides cells with both mechanical support as well as protection against biotic stress factors that include infection by bacteria, fungi and viruses (Hahn *et al.*, 1989). The two types of polysaccharide matrices that make up the plant cell walls are the pectate network and the cellulose network (Herron *et al.*, 2000).

Fungal pathogens penetrate their hosts in order to establish infection. Like bacteria and viruses, fungal plant pathogens have evolved diverse strategies to ingress plant tissues. They can either enter through wounds, natural openings such as the stomata, or by penetrating the cell wall components directly (Hahn *et al.*, 1989; Schäfer, 1994). Direct penetration can be achieved either through mechanical forces, where the

deposition of melanins on the inner appressorial cell wall mediates the build-up of hydrostatic pressure that enables the fungus to puncture the plant epidermis, or by means of enzymatic degradation (Schäfer, 1994).

It is generally accepted that the enzymatic arsenal of fungal pathogens contributes, together with mechanical forces, to the degradation of both the cuticle as well as the polysaccharide-rich cell wall (De Lorenzo *et al.*, 1997). Pathogenic fungi need to degrade the plant cell wall not only for establishing infection, but also for releasing from the wall polysaccharides, nutrients necessary for their growth (De Lorenzo *et al.*, 1997). Studies on the enzymatic digestion of cell walls has led to the description of at least 20 different cell wall-degrading enzymes (CWDEs) (Walton, 1997). Most of these enzymes are extracellular, inducible glycoproteins that exist as multiple isoenzymes (Hahn *et al.*, 1989). Enzyme activities that have been identified include the pectic enzymes, cellulases, arabinase, xylanase, and galactanase (De Lorenzo *et al.*, 1997; Collmer and Keen, 1986; Hahn *et al.*, 1989; Walton, 1997; Ridley *et al.*, 2001; Fischer and Bennett, 1991).

Pectin Degrading Enzymes

Pectins belong to a group of complex polysaccharides that are often described in terms of 'smooth' and 'hairy' blocks (Fisher and Bennett, 1991). The 'smooth' blocks are defined by a linear copolymer of a- (1? 4)-linked galacturonic acid and its methyl ester, while the 'hairy' pectin blocks include rhamnogalacturonans I and II and are complex heteropolymers, comprising 12 different sugars (Fischer and Bennett 1991). These 'smooth' and 'hairy' pectin blocks cement the plant cells together and are thus responsible for maintaining plant tissue integrity (Varner and Lin, 1989; Ridley *et al.*, 2001).

The degradation of pectic polymers is one of the earliest functions performed by plant pathogenic fungi, making pectin-degrading enzymes considerably important for the plant infection process (Whitehead *et al.*, 1995; Caprari *et al.*, 1993a). Several types of pectinases are produced and these include the endo and exo-pectate lyases (PL), endo and exo-polygalacturonase (PG) and pectin methylesterase (PME) (Walton, 1994).

Pectin lyase, pectate lyase and exopolygalacturonate lyase cleave the glycosidic bonds of the pectin backbone by ß-elimination thus generating products with a 4,5-unsaturated galacturonosyl residue at the non-reducing end (Lima *et al.,* 2002; Collmer and Keen, 1986). The highly pectinolytic gram-negative bacteria, *Erwinia chrysanthemi* and *E. carotovora* are among the best known producers of pectate lyases (Davis *et al.,* 1984; Collmer and Keen, 1986). The role of pectate lyase in pathogenesis has been investigated by producing a mutant *E. chrysanthemi* strain in which four isozymes of pectate lyase was deleted. The mutant strain was not only able to grow on polygalacturonic acid as the sole carbon source, but it also caused significant plant tissue maceration, indicating that pectate lyase are not the only enzymes required for maceration (Hahn *et al.,* 1989; Collmer and Keen, 1986, Walton, 1994).

The cucumber pathogen *Cladosporium cucumerinum* produces pectate lyases that cause maceration in susceptible cucumber hypocotyls (Robertsen, 1989). The purified enzyme showed a pH optimum of about 9.7 as well as a dependance on Ca^{2+} ions for its activity (Robertsen, 1989). Similarly, the production of pectin lyase by *Colletotrichum lindemuthianum* in infected bean tissue, showed maximum activity at alkaline pH (Wijesundera *et al.,* 1984). However, unlike endopolygalacturonases, which are constitutively expressed, the lyases are largely regulated by induction and by cyclic AMP mediated catabolite repression (Keen *et al.,* 1984; Robertsen, 1989).

Polygalacturonases (PGs) are classified into endo-PGs and exo-PGs. They are more active in degrading demethylated than methylated pectin. Exo-polygalacturonases act on the non-reducing ends of the substrate chain, removing one or two sugars at a time (Walton 1994; Lima *et al.,* 2002). They do not cause plant tissue maceration, but have a role in degrading elicitor-active oligogalacturonides released by endoPGs and are generally not inhibited by polygalacturonase-inhibiting proteins (García-Maceira *et al.,* 2000). Endopolygalacturonases (Endo-PGs) are enzymes that catalyse the hydrolytic cleavage of a-(1? 4) galacturonan linkages of homogalacturonan and have been associated with tissue maceration and killing of plant cells (Fischer and Bennett, 1991; Hahn *et al.,* 1989).

Enzymatic secretion occurs in a time-specific manner depending on the accessibility of the cell wall polymers, nutrient suitability of the pathogen and the level of resistance to polymer degradation (Walton, 1994). The sequential occurrence of different classes of cell wall-degrading enzymes suggests that different signal molecules are responsible for inducing the specific enzymes (Dixon and Lamb, 1990).

Endopolygalacturonase activity appears to be crucial during the initial phase of fungal infection. At the onset of infection, the cell wall pH is low and the degradation is localised to the site of infection. As time progresses, the cell wall pH rises to above 6.5, at which time polygalacturonase activity is greatly reduced. Pectate lyase activity thus increases together with an increase in fungal growth (De Lorenzo *et al.*, 1997).

Role of Endopolygalacturonases in Pathogenicity

All plant pathogens secrete enzymes that are capable of degrading the polysaccharides of plant cell walls (Albersheim and Anderson, 1971). The fungal pathogen *C. lindemuthianum* has been shown to produce a wide range of cell wall-degrading enzymes that were secreted sequentially depending on the age of the culture (English *et al.,* 1972). A large variety of these polysaccharide-degrading enzymes have been isolated from plants, fungi and bacteria (Karr and Albersheim, 1970). Pectin degrading enzymes are invariably among the first of these enzymes to be secreted by pathogens grown in culture on host cell walls (Albersheim and Anderson, 1971; Scott-Craig *et al.,* 1990; Hahn *et al.,* 1989).

Endopolygalacturonases (Endo-PGs) are pectin-degrading enzymes. They are produced by numerous fungi and bacteria, higher plants and some plant-parasitic nematodes (Patino *et al.,* 1997). Endo-PGs, from a number of different fungi have been purified and characterised (Caprari *et al.,* 1993b). These chain-splitting glycoside hydrolases degrade the pectin polymers by hydrolysing the a-1,4 glycosidic linkages between galacturonic acid residues within the homolgalacturonan domain of pectin (Agrios, 1988; Herron *et al.,* 2000). Endopolygalacturonases are thus of particular importance due to the fact that they are produced in the largest amounts and are capable of macerating plant tissues and killing plant cells on their own (De Lorenzo and Cervone, 1997).

Polygalacturonases have several features, which indicate that these enzymes may have multiple roles during pathogenesis. Pre-treatment with endopolygalacturonases facilitates depolymerization of plant cell walls by other fungus-secreted cell wall-degrading enzymes (Karr and Albersheim, 1970; Cervone *et al.,* 1989). Products produced by enzymatic cleavage are utilized as nourishment by the fungus during the initial stages of colonization (Cervone *et al.,* 1989). In addition, several filamentous fungi produce multiple forms of secreted endopolygalacturonases and the redundancy of these enzymes may facilitate pathogenesis in various hosts and under a variety of conditions (Cooper, 1984; Caprari *et al.,* 1993a; De Lorenzo and Cervone, 1997).

According to Caprari *et al.* (1993b) analysis of chemically induced mutants provided the first genetic evidence for the involvement of endopolygalacturonases in the fungal diseases of plants. Scott-Craig *et al.* (1990) were the first to report the cloning and characterisation of an endopolygalacturonase gene designated as *PGN1* from the maize fungal pathogen, *Cochliobolus carbonum.* This enabled the construction of a precise gene disruption mutant in which only the PG gene is affected. The resulting mutant strain retained its ability to infect maize, indicating that endopolygalacturonases were not required in this disease interaction. In 1996, Caprari *et al.* provided the first molecular evidence for the role of endopolygalacturonases in cell wall degradation. A comparison between the endopolygalacturonase sequences from bacteria, fungi and plants showed the existence of highly conserved amino acid regions (Scott-Craig *et al.,* 1990). This conserved region contains a histidine residue that is essential for the activity of PGs (Caprari *et al.,* 1993b, 1996). Using site-directed mutagenesis on an endopolygalacturonase gene from *Fusarium moniliforme*, Caprari *et al.* (1996) were able to show that replacement of the His234 residue with a Lysine abolished enzymatic and macerating activity. Similarly, mutations of either the Ser234 or Ser240 residues to a Glycine, reduced enzymatic activity in *Fusarium moniliforme* PG.

Martel *et al.* (1998) reported the separation of 16 isoforms of polygalacturonases from the plant pathogen *Sclerotinia sclerotiorum.* They were able to identify seven genes constituting two subfamilies of endopolygalacturonases and hypothesised that the occurrence of multiple enzyme forms increases adaptive capacity and flexibility of

pathogens. Cook *et al.* (1999) went a step further and showed that different fungal PGs differ in their ability to degrade pectic substrates and that they exhibit different levels of susceptibility to inhibitory proteins. This implies that the underlying ability of a pathogen to degrade its host cell wall depends on whether the PGs produced are sufficiently active to initiate pathogenesis and also on the number of different inhibitory proteins present in the host.

One of the key features of fungal endoPGs is that they are elicitors of plant defence responses. This was initially observed in *Rhizopus stolonifer* endo-PGs, which were able to elicit phytoalexin accumulation in castor bean seedlings (Lee and West, 1981a). EndoPGs may, therefore, play two opposing roles during invasion, firstly as agents of fungal aggression and secondly as signal molecules (Cervone *et al.*, 1989). However, evidence suggests that endopolygalacturonases are not directly responsible for the induction of plant defence responses. Instead, the oligogalacturonide fragments released by the activity of endoPGs on plant cell walls are the true elicitors (De Lorenzo and Cervone, 1997).

Oligosaccharide Signals For Defence Response

Plants have evolved several mechanisms enabling ingress by a pathogen (Ridley *et al.*, 2001). These include the strengthening of the cell wall by deposition of lignin (Altamura *et al.*, 1998), the synthesis of phytoalexins and other antimicrobial compounds (Lee and West, 1981a; Walker-Simmons *et al.*, 1983; Jin and West, 1984; Davis *et al.*, 1986) and the production of several pathogenesis-related (PR) proteins (Cervone *et al.*, 1989). Fungal endopolygalacuronases have been implicated in the activation of a variety of these plant defence responses (Cervone *et al.*, 1989). Linear a-1,4-D-galactopyranosyluronic acid residues produced by the enzymatic cleavage of cell wall homogalacturonans by endopolygalacturonases, are biologically active carbohydrates. These carbohydrates are utilized by the pathogen as a carbon source and also function as signal molecules in the activation of defence responses (Ridley *et al.*, 2001; Cervone *et al.*, 1989).

The majority of elicitors of plant defence responses that have been characterised are derived from invading microorganisms (Nothnagel *et al.,* 1983). Lee and West (1981b) identified a glycoprotein polygalacturonase from *R. stolonifer* as an elicitor of casbene synthetase activity in castor beans. The elicitor activity was subsequently shown to be dependent on the catalytic activity of the polygalacturonase enzyme. Several lines of evidence support the idea that pectic fragments released from castor bean cell walls act as obligate intermediates in the elicitation process (Jin and West, 1984).

Hahn *et al.* (1981) described the release of endogenous elicitors from cell walls of soybeans by partial acid hydrolysis. The results of their study showed that the endogenous elicitor was a polysaccharide fragment as it contained the essential constituent, galacturonic acid. Studies to confirm and further characterize this endogenous elicitor, were later undertaken by Nothnagel *et al.* (1983). Consistent with results obtained by Hahn *et al.* (1981), treatment of purified pectin elicitors with highly purified endo-a-1,4-polygalacturonase, completely destroyed the elicitor activity.

Structural characterisation of elicitors has shown that the most active elicitor is a dodeca-a-1,4-D-galacturonide, which is a homo-oligomer of galacturonic acid (Jin and West, 1984). Linear a-(1? 4)-linked oligogalacturonides with a degree of polymerisation between 10 and 13, have thus been shown to be the most active elicitors of plant defence responses. Oligogalacturonides with shorter than 10 glycosyl residues have either little or no elicitor activity (Hahn *et al.,* 1989; De Lorenzo and Cervone, 1997).

Cervone *et al.* (1989) considered whether the degradation of elicitor-active oligogalacturonides into smaller inactive oligomers, could be manipulated such that the accumulation of active oligomers is prolonged. They investigated the effect of bean (*Phaseolus vulgaris*) PGIP on the digestion of polygalacturonic acid by either *Aspergillus niger* or *F. moniliforme* endopolygalacturonases. In the presence of PGIP, the half-life of elicitor-active oligogalacturonides is increased as opposed to when the inhibitory protein is absent (Cervone *et al.,* 1989). In the absence of the protein, endopolygalacturonases rapidly depolymerise the active oligogalacturonides

into shorter, inactive molecules. Bergmann *et al.* (1994) later showed that PGIPs are induced in suspension-cultured bean cells when elicitor-active oligogalacturonides or fungal glucan are added to the medium.

According to De Lorenzo and Cervone (1997), endopolygalacturonases and PGIPs form very specific interactions that lead to the accumulation of high concentrations of oligogalacturonides, ultimately activating plant defence responses. It has thus been hypothesised that PGIPs have a general role in reducing the activities of fungal PGs such that active oligogalacturonides are present for longer periods of time at the infection sites thus enhancing the signalling properties of plants (Ryan and Farmer, 1991).

Polygalacturonase-Inhibiting Proteins (PGIPs)

Polygalacturonase-Inhibiting Proteins (PGIPs), have been defined by their ability to inhibit fungal polygalacturonases (PGs) (De Lorenzo and Cervone, 1997). PGIP activity was found to occur in both infected (Abu-Goukh and Labavitch, 1983) and uninfected plant tissues (Albersheim and Anderson, 1971; Abu-Goukh *et al.*, 1983a) suggesting that they form part of the plant's pre-existing defence mechanism. They were initially described by Weurman (1953), but Albersheim and Anderson (1971) were the first to report the presence of a protein in *P. vulgaris* cell walls that could inhibit endoPGs of *C. lindemuthianum* (Lafitte *et al.*, 1984; Esquerré-Tugaýe *et al.*, 2000; Boudart and Dumas, 2000). PGIPs have since been shown to be associated with the cell walls of several dicotyledonous plants (De Lorenzo and Cervone, 1997). They have subsequently been purified from several plant species (Table 1).

Table 1.1 Summary of PGIPs purified from several plant species.

	Plant Species	Reference
Pear	*Pyrus communis* L. cv. Bartlett	Abu-Goukh *et al.* (1983a); Stotz *et al.* (1993)
Bean	*Phaseolus vulgaris* L. cv. Cannellino	Albersheim and Anderson (1971); Cervone *et al.* (1987); Salvi *et al.* (1990)
Soybean	*Glycine max* L. Merr.cv.Canton	Favaron *et al.* (1994)
Raspberry	*Rubus idaeus* L. cv. Autumn Bliss	Johnston *et al.* (1993)
Apple	*Malus domestica* cv. Golden Delicious	Yao *et al.* (1995)
Tomato	*Lycopersicon esculentum* cv. Sunny	Stotz *et al.* (1994)
Onion*	*Allium cepa* L.	Favaron *et al.* (1993)
Leek*	*Allium porrum* L.	Favaron *et al.* (1993) (1997)
Potato	*Solanum tuberosum* L. cv. Spunta	Machiandiarena *et al.* (2001)
Cotton	*Gossypium hirsutum* L.	James and Dubery (2001)
Guava	*Psidium guajava* Linn	Deo and Shastri (2003)

* monocotyledonous plant species

Tissue-specific Expression of PGIP genes

The accumulation of pgip mRNA is spatially and temporally regulated during development and in response to several stress stimuli (Devoto *et al.*, 1997; De Lorenzo *et al.*, 2001). Bioassays of PGIP activity from various *P. vulgaris* tissues indicated that the protein is ubiquitously expressed throughout the plant (Salvi *et al.*, 1990). The lowest level of PGIP expression was observed in the roots, while the vegetative apex displayed the highest activity. PGIP levels also increased in the stem tissue during growth but no appreciable difference could be observed at the different developmental stages of the leaves (Salvi *et al.*, 1990; Devoto *et al.*, 1997; Jones and Jones, 1997).

Both mature and immature flowers of *P. vulgaris* displayed high levels of PGIP activity (Salvi *et al.*, 1990). Northern blot analysis revealed that the stamens and the style displayed the highest *pgip* transcript levels (Salvi *et al.*, 1990; Devoto *et al.*,

1997). Biochemical analysis of the purified proteins from both flowers and the hypocotyls showed a molecular mass difference between the two proteins. This suggested that the different bean PGIPs are regulated in a tissue-specific manner (Salvi *et al.*, 1990).

Extraction and bioassay of PGIP activity from various soybean (*Glycine max*) tissues indicate that the protein is expressed at low levels in the roots and hypocotyls, with higher expression occurring in the leaves and the cotyledons (Favaron *et al.*, 1994; Jones and Jones 1997). In fruits such as pear, apple, raspberry and grape, *pgip* mRNA levels decrease as the fruits mature (Stotz *et al.*, 1993; Johnston *et al.*, 1993; Yao *et al.*, 1999; De Lorenzo *et al.*, 2001). Northern blot analysis of pear (*Pyrus communis*) *pgip* mRNA indicated that the transcript level in fruits was approximately 200-fold higher than in flowers and 1400-fold higher than in leaves (Stotz *et al.*, 1993). In contrast, no PGIP activity could be detected in raspberry (*Rubus idaeus*) flowers and in extracts from tomato (*Lycopersicon esculentum*) leaves (Johnston *et al.*, 1993; Stotz *et al.*, 1994).

Several stress stimuli have been reported to induce PGIP activity (De Lorenzo *et al.*, 2001). It has been observed that the accumulation of *pgip* transcripts could be induced in bean suspension-cultured cells by adding elicitor-active oligogalacturonides and fungal glucan to the medium (Bergmann *et al.*, 1994). In hypocotyls, the accumulation of *pgip* transcripts occurred in response to wounding or salicylic acid treatment (Bergmann *et al.*, 1994). Immunoblotting analysis of potato (*Solanum tuberosum*) PGIP revealed that the protein is induced by wounding, salicylic acid or by infecting the leaves with *Phytophtora infestans* (Machiandiarena *et al.*, 2001). The induction of PGIP by salicylic acid, which is a systemic stress signal, strengthens the hypothesis that PGIPs play a role in the resistance of plants to fungal attack (Machiandiarena *et al.*, 2001).

Northern blot analysis was used to study the temporal location of *pgip* mRNA in two near-isogenic lines of *P. vulgaris*. The two lines, one resistant and one susceptible to infection by *C. lindemuthianum,* was infected with race ? of *C. lindemuthianum*. Rapid accumulation of mRNA correlated with the appearance of the hypersensitive response in incompatible interactions, while a more delayed increase in *pgip* mRNA

in compatible interactions, was coincident with the formation of lesions (Nuss *et al.*, 1996; De Lorenzo *et al.*, 2001). By means of in-situ hybridization, Devoto *et al.* (1997) investigated the spatial distribution of *pgip* transcripts in infected hypocotyls and leaves. They found that the accumulation of *pgip* mRNA is mainly localised at the site of infection in the incompatible interactions.

Biochemical And Molecular Characterisation of PGIPs

Biochemical characterisation of several plant PGIPs have revealed the occurrence of two forms of protein namely, wall-bound PGIPs and soluble PGIPs (Lafitte *et al.*, 1984). Most PGIPs that have been characterised are relatively heat stable glycoproteins that are ionically bound to the plant cell wall (Salvi *et al.*, 1990; Johnston *et al.*, 1994; Jones and Jones, 1997). Several observations have suggested that almost all PGIP activity is released from cell debris following extraction with a high salt buffer (Machiandiarena *et al.*, 2001). However, soluble extracytoplasmic PGIP forms have also been described. In *P. vulgaris* both soluble and cell wall bound forms of the inhibitor are present, but they were indistinguishable from each other based on their elution profiles during ion-exchange chromatography or by gel filtration. They thus appeared to be the same protein (Lafitte *et al.*, 1984). This is in contrast to the results obtained for *Pisum sativum* (pea) inhibitors, which appeared to be different proteins (Hoffman and Turner; 1984 Lafitte *et al.*, 1984).

Preparative iso-electric focusing (IEF) showed that PGIPs from several plants have basic pI values, with the exception of apple (*Malus domestica*) PGIP, which has a pI value of 4.6, pear PGIP, which has isoforms with pIs of 4.5, 6.6 and 7.7, and guava (*Psidium guajava*) PGIP, with pI values ranging between 4.2 and 4.4 (Stotz *et al.*, 1993; Yao *et al.*, 1995; Glinka and Protsenko, 1998; Deo and Shastri, 2003). PGIPs characterised from bean and raspberry have very high pI values thus facilitating the binding of these positively charged proteins to the negatively charged pectin backbone of the plant cell walls (Toubart *et al.*, 1992; Johnston *et al.*, 1993).

Table 1.2 Summary of biochemical characteristics of some PGIP proteins as determined by SDS-PAGE and Isoelectric focusing analysis.

Plant Species	kDa	pI Value	Reference
Glycine max	37-40	9-10	Favaron *et al.* (1994)
Pyrus communis	34	4.6	Abu-Goukh *et al.* (1983); Stotz *et al.* (1993)
		6.6	
		7.7	
Malus domestica	34	4.6	Yao *et al.* (1995)
Lycopersicon esculentum	34	9	Stotz *et al.* (1994)
Actinidia deliciosa	34	8.2	Simpson *et al.* (1995)
Solanum tuberosum	41	ND	Machiandiarena *et al.* (2001)
Psidium guajava	29.5	4.2-4.4	Deo and Shastri, (2003)
	32.5		
	35		

Table 1.3 Summary of biochemical characteristics of some PGIP proteins as predicated from the gene sequences.

Plant Species	kDa	pI Value	Reference
Phaseolus vulgaris	34	8.81	Toubart *et al.* (1992); Leckie *et al.* (1999)
Pyrus communis	36.5	6.2	Stotz *et al.* (1993)
Glycine max	34	8.21	Favaron *et al.* (1994)
Rubus ideaus	38.5[a]	+10.0	Johnston *et al.* (1993); Ramanathan *et al.* (1997)

Molecular mass determination by either gel filtration or SDS-PAGE provided molecular mass estimations for PGIPs ranging from 37 to 54 kDa (Glinka and Protsenko, 1998). Potato PGIP, which exists as a monomer, has a molecular mass of 41 kDa. This value is consistent with molecular mass estimations of bean and raspberry, but shows little similarity to pear, which consists of two isoforms of 91 kDa each (Abu-Goukh *et al.*, 1983a). Leek (*Allium porrum*) contains three PGIP types, with molecular weights ranging from 39kDa to 42kDa (Favaron *et al.*, 1997). The three guava PGIPs recently reported seem to show comparatively lower molecular weights as compared to those previously characterised (Deo and Shastri,

2003). Molecular mass differences between PGIP proteins may be due to different levels of glycosylation (Glinka and Protsenko, 1998).

Isolation and characterisation of the *pgip* gene from bean

Trypsin digestion of purified bean PGIP enabled the identification of the N-terminal amino acid sequence and four internal peptide sequences of the protein. Based on the amino acid sequence of *P. vulgaris* cv. Pinto PGIP three degenerate oligonucleotides designated as N-A, N-B and Int were synthesised. Amplification with primers N-A and Int lead to the production of a 788bp fragment designated as PGIP-1, whereas primers N-B and Int gave a PGIP-2 fragment of 758bp. Southern analysis using labelled PGIP-2 as a probe showed that both PGIP-1 and PGIP-2 hybridised with the same intensitiy. The 758bp PGIP-2 fragment was thus cloned, sequenced and subsequently used as a hybridization probe in the isolation of the full-length PGIP gene from a commercial genomic library of *P. vulgaris* cv. Saxa. The isolated genomic clone consisted of a single, continuous open reading frame of 1026 nucleotides (342 amino acids) with four potential glycosylation sites and two in-frame ATGs (Toubart *et al.*, 1992).

A cDNA library of *P. vulgaris* cv. Pinto was screened using the genomic *pgip* clone isolated from *P. vulgaris* cv. Saxa (Leckie *et al.*, 1999). Seventeen clones were selected and subjected to restriction enzyme digestion and southern blot analysis. The 10 longest inserts were subcloned and sequenced. Sequence analysis revealed that two members (*pgip1* and *pgip2*) of the *pgip* gene family of *P. vulgaris* cv. Pinto was isolated. The *pgip1* gene was identical to the genomic clone of *P. vulgaris* cv. Saxa , which contained the two in-frame ATGs, while *pgip2* started with an ATG corresponding to the second ATG of *pgip1* (Leckie *et al.*, 1999). The mature PGIP had a molecular mass of 34kDa, and the presence of several lysines and arginines confer a high isoelectric point to the protein (Toubart *et al.*, 1992).

Isolation and characterisation of a *pgip* gene from pear

The gene encoding the pear PGIP was isolated using a combination of degenerate PCR and RACE-PCR (Stotz *et al.,* 1993). Degenerate oligonucleotide primers, designed based on partial amino acid sequence of the purified pear PGIP protein, were used to amplify a 569bp product from the pear cDNA. Sequence analysis of the degenerate-oligo primed cDNA, enabled the design of gene-specific primers that were subsequently used in RACE-PCR to amplify the 5' and 3' ends of the PGIP transcript (Stotz *et al.* 1993). A second set of gene-specific primers based on the sequence of the RACE-PCR products, were designed and used to amplify the complete PGIP coding sequence of 1059 bp. The isolated *pgip* gene encodes a protein of 330 amino acids, including a signal sequence of 24 amino acids and seven potential N-Glycosylation sites, of which only two are shared with bean. Amino acid sequence comparisons between the pear PGIP and that of bean, shows only moderate sequence identity of 50%. However, all eight cystein residues of the field bean PGIP are conserved in pear (Stotz *et al.,* 1993).

Isolation and characterisation of a *pgip* gene from soybean

Based on the homology between the field bean and soybean PGIPs, oligonucleotide primers corresponding to the N and C-terminal regions of the field bean PGIP gene were synthesised and used to amplify a fragment of approximately 950bp (Favaron *et al.,* 1994). The PCR product was cloned and subjected to nucleotide sequencing, which revealed that the gene contained a single open reading frame that encodes a polypeptide of 313 amino acids. The resulting amino acid sequence showed 66.8% and 47% homology to that of the previously isolated field bean and pear PGIPs, respectively. Nine cysteine residues are located in the N-terminal and C-terminal regions of the protein, eight of which are conserved in the field bean and pear PGIPs, suggesting that they may be involved in stabilizing the tertiary structure of the protein (Stotz *et al.,* 1993; Favaron *et al.,* 1994). Three potential N-glycosylation sites (Asn-X-Ser/Thr) located in the N-terminal region are conserved in the soybean, field bean and pear PGIPs (Favaron *et al.,* 1994).

Mahalingam *et al.* (1999) went a step further in the study and amplified an expected 950bp *pgip* fragment using oligonucleotide primers designed by Favaron *et al.* (1994). They used genomic DNA and root cDNAs of two soybean genotypes, PI437654 and a commercial cultivar Essex, in which parasitism of the host plant occurs by a soybean cyst nematode (SCN) as templates for amplification. Sequence analysis of the isolated *pgip* genes from both genotypes showed that they were 100% identical to each other at the nucleotide level. When compared to other published PGIP sequences, the PGIP from roots showed 88% amino acid identity to that of the field bean PGIP and only 72% amino acid identity to the soybean PGIP sequence published by Favaron *et al.* (1994). Previous data showed that the PGIP obtained by Favaron *et al.* (1994) was obtained from a single genomic clone that showed only 67% amino acid identity to the field bean PGIP. Based on the results presented by Mahalingam *et al.* (1999), they were able to verify that the PGIP isolated from the root cDNA represented a true copy of the soybean PGIP.

Isolation and characterisation of a *pgip* gene from tomato

Degenerate oligonucleotide primers have been synthesised from the amino acid sequence of purified tomato PGIP A (Stotz *et al.*, 1994). PCR reactions performed on genomic DNA as well as tomato fruit cDNA, yielded identical products of 417 bp, which were subsequently used as probes to screen a tomato genomic library. A full-length *PGIP* gene with an uninterrupted open reading frame of 984 nucleotides was identified. The gene encodes a protein of 328 amino acids and contains a putative signal sequence of 20 amino acids.

The predicted molecular mass (34.3 kDa) as well as the isoelectric point (8.4) of the mature polypeptide, corresponds to the values obtained for tomato PGIP A (Stotz *et al.*, 1994). Comparisons of the deduced tomato PGIP to that of pear and bean, showed that the mature polypeptide is more closely related to pear PGIP (68% identity) than to bean PGIP (50 % identity), and that eight cysteine residues are conserved between these species. Tomato PGIP contains seven potential N-glycosylation sites, two of which are conserved between the tomato, pear and bean

PGIPs, and three additional glycosylation sites that are shared with the pear PGIP (Stotz *et al.*, 1994).

Isolation and characterisation of a *pgip* gene from apple

The apple pgip gene has been isolated and characterised by two research groups. Arendse *et al.* (1999) employed a degenerate and inverse PCR technique while Yao *et al.* (1999) used degenerate PCR to isolate the full length apple *pgip* gene. The N-terminal amino acid sequence of the purified apple inhibitor was determined by protein sequencing (Yao *et al.*, 1995). Comparison of the apple PGIP proteins sequence to GenBank proteins sequences revealed that the apple PGIP shared 96% amino acid identity to the published pear PGIP protein sequence (Yao *et al.*, 1995). Three degenerate oligonucleotide primers based on the homology between pear and apple PGIP amino acid sequences could thus be designed. (Yao *et al.*, 1995; Yao *et al.*, 1999). One primer was designed from the N-terminal apple PGIP sequence (PGIP-N1), while the other two were designed from the C-terminus of pear PGIP (PGIP-C1 and PGIP-C2) (Yao *et al.*, 1999). Amplification of apple genomic DNA with primers PGIP-N1 and PGIP-C1 resulted in a fragment of 896bp, which was used as a probe to screen an apple cDNA library. A full-length *pgip* with a nucleotide sequence of 1162bp was identified. It contained a single open reading frame encoding a predicted protein of 330 amino acids, the first 24 amino acids being the signal peptide sequence.

Degenerate primers were designed to conserved internal regions between the bean (Toubart *et al.*, 1992) and pear (Stotz *et al.*, 1993) PGIP sequences (Arendse *et al.*, 1999). Amplification of apple genomic DNA gave an internal fragment that was 351bp in length. The internal apple PGIP fragment showed 94% identity on the DNA level and 97% similarity on the amino acid level to the corresponding pear *pgip* region (Arendse *et al.*, 1999). Gene-specific primers based on the internal fragment sequence were designed and used to amplify the flanking regions of the *pgip* by means of inverse PCR. Apple genomic DNA digested with *Bgl*II was used as a template in the inverse PCR reaction, and yielded a product of 820bp, which represented the 5' end of he gene. Based on the sequence similarity to the internal

PGIP fragment, the two sequences were joined to give a single open reading frame encoding 274 amino acids. A second set of inverse PCR primers were designed to amplify the 3' end of the gene. The complete sequence of the composite apple *pgip* gene was identical to the full-length *pgip* gene from Golden Delicious apples (Arendse *et al.*, 1999; Yao *et al.*, 1999).

The predicted molecular mass of the mature peptide is 34 kDa, which is in accordance with the estimated value for the native protein. However, the isoelectric point is 7, while that of the native protein was measured to be within the pH range of 3.0-5.9. This difference in pI values was attributed to the different levels of glycosylation of the seven potential N-glycosylation sites on the same protein (Yao *et al.*, 1999). When compared to several other PGIP sequences, apple PGIP showed 98, 83,82, 81, 67 and 63% identities at the amino acid sequence level with those from pear, orange, kiwifruit, tomato, bean and soybean, respectively (Yao *et al.*, 1999).

Isolation and characterisation of a *pgip* gene from raspberry

Primers designed to the conserved regions of bean (Toubart *et al.*, 1992) and pear (Stotz *et al.*, 1993), were used to amplify an expected 806bp *pgip* gene from raspberry genomic DNA. An unexpected PCR product of 1050bp was obtained instead. Sequence analysis of the PCR product showed a high degree of sequence identity to pear PGIP at both ends of the insert. However, the 990bp insert also contained an internal sequence region that showed no similarity to previously isolated *pgip* genes. Raspberry-specific primers were used to amplify a 260bp probe that was subsequently labelled and used to screen a cDNA library prepared from immature raspberry fruit RNA (Ramanathan *et al.*, 1997). Two of nine positive clones were sequenced, and a full-length *pgip* cDNA of 1325bp was identified (Ramanathan *et al.*, 1997). The cDNA clone was identical to the genomic clone, with the exception of the 293bp internal intron present in the genomic clone.

The raspberry *pgip* cDNA contains an open reading frame that is predicted to encode a 331 amino acid protein. The polypeptide contains a 22 amino acid N-terminal signal sequence, and eight conserved cysteine residues (Ramanathan *et al.*, 1997).

The raspberry PGIP contains four potential N-glycosylation sites and the deduced amino acid sequence shows a high degree of similarity to the predicted amino acid sequences from pear, tomato and kiwifruit (Ramanathan *et al.,* 1997).

Isolation and characterisation of a partial *pgip* gene from *Eucalyptus spp.*

Degenerate oligonucleotide primers, one designed from the pear PGIP sequence (Stotz *et al.,* 1993) and the other designed by inspection of the C-termini of aligned PGIP polypeptides (Chimwamurombe *et al.*,2001), were used to amplify a 909bp fragment from genomic DNA of *E. grandis*. The partial *Eucalyptus pgip* gene isolated encodes a protein of 298 amino acid residues that is uninterrupted by intron sequences. The partial *pgip* gene shows relatively high sequence similarity to apple (95%), pear (94%), kiwifruit (69%) and tomato (65%) *pgip* genes (Chimwamurombe *et al.,* 2001). The *E. grandis* PGIP polypeptide contains seven potential N-glycosylation sites that are comparable to those found in pear PGIP (Stotz *et al.,* 1993). In contrast to other reported PGIPs, however, the partial mature peptide of *E. grandis* contains only seven cysteine residues located in the N- and C-terminal regions of the polypeptide (Chimwamurombe *et al.,* 2001). Analysis of the amino acid composition of several cloned PGIPs, indicates the presence of a high content of leucine residues (De Lorenzo and Cervone, 1997; Glinka and Protsenko, 1998). *Eucalyptus* PGIPs also fall into the category of the leucine-rich repeat class of proteins (Chimwamurome *et al.,* 2001). A modified 24 amino acid leucine-rich repeat motif is conserved in all *pgip* genes sequenced thus far (De Lorenzo *et al.,* 1994). In plants, LRR proteins are responsible for many protein-protein interactions that are relevant in both plant development and defence (De Lorenzo and Cervone, 1997; Leckie *et al.,* 1999).

Leucine-Rich Repeat Proteins (LRRs)

Leucine-Rich Repeats (LRRs) were initially discovered in the leucine-rich human serum protein, a2-glycoprotein (Kobe and Deisenhofer, 1993). More than a hundred different LRR proteins with differing functions and cellular locations have now been

described (Kobe and Deisenhofer, 1994; Leckie *et al.,* 1999). LRRs are usually present in tandem, and are distinguished by a consensus sequence consisting predominantly of leucine residues. They are commonly 24 residues in length, but LRRs between 20 and 30 residues have been identified (Kobe and Deisenhofer, 1994). LRRs belong to a superfamily of proteins that appear to be involved in protein-protein interactions, and to some extent, some may even be involved in signal transduction pathways (Kobe and Deisenhofer, 1994; Jones and Jones, 1997).

The existence of LRRs in animals and fungi has been known for some time, but LRR proteins have only recently been described in plants (Jones and Jones, 1997). The first plant LRR proteins identified were the Polygalacturonase-Inhibiting Proteins (PGIPs) and receptor kinases (Jones and Jones, 1997). Many plant *R*-genes have LRR domains (Dixon *et al.,* 1998; Marino *et al.,* 2000). Proteins containing LRRs differ greatly in structure and function and many are part of the cell wall receptors that respond to elicitors (Ramanathan *et al.,* 1997; Marino *et al.,* 2000). Plant *R*-genes confer gene-for-gene resistance to a wide array of natural enemies (Stahl and Bishop, 2000). The *R*-gene products thus serve as receptors for pathogen-encoded avirulence (avr) proteins and they lead to the activation of downstream defence response pathways (Leckie *et al.,* 1999). The first disease resistance (*R*) gene cloned was a *Pto* gene, which confers resistance in tomato to a bacterial spot disease caused by *Pseudomonas syringae* (Dixon *et al.,* 1998). Several other genes, including the tomato *Cf9* gene, which confers resistance to races of *Cladosporium fulvum;* and the rice gene *Xa21*, which confers resistance to the bacterium *Xanthomonas oryzae* pv. *oryzae* have since been cloned (Jones and Jones, 1997; Hammond-Kosack *et al.,* 1998).

Kobe and Deisonhofer (1993) determined the three-dimensional structure of porcine ribonuclease-inhibitor protein (PRI) as a model for other LRR proteins. They were able to show that the LRR of the RI molecule consists of a-helixes that are approximately parallel to short ß-sheets. This gives rise to a non-globular, horseshoe shaped molecule. The binding of PRI to the ribonuclease involved multiple amino acids, which were probably located on the exposed surface of the parallel ß-sheet. It was thus concluded that like the PRIs, leucine-rich repeats of PGIPs could also form three-dimensional structures capable of inhibiting fungal polygalacturonases by

extensive binding to exposed amino acids (Kobe and Deisonhofer, 1993; Yao *et al.*, 1999).

Each of the PGIPs that have been isolated consist primarily of 10 imperfect LRR, with an average of 24 amino acids. The consensus sequence, LxxLxxLxxLxLxxNxLxGxIPxx, remains conserved in all PGIPs currently sequenced and it also shows significant homology to several disease resistance (*R*) genes (Leckie *et al.*, 1999; Jones and Jones, 1997).

PGIPs are the only LRR's for which a ligand has been demonstrated. Models for determining the specificity with which they are able to differentially inhibit PGs from different fungal sources have thus been proposed (Stotz *et al.*, 2000; Leckie *et al.*, 1999). Leckie *et al.* (1999) hypothesised that sequence variations with the LRR domains could influence recognition specificity. The interaction of plant PGIPs and fungal PGs at the molecular level was thus investigated.

Like most *R*-genes, PGIPs consist of multigene families with different inhibitory specificities. In *P. vulgaris* cv. Pinto, two members of the *pgip* gene family have been isolated and shown to have only eight amino acid differences in their mature peptides (Leckie *et al.*, 1999). Using site-directed mutagenesis and surface plasmon resonance (SPR), Leckie *et al.* (1999), showed that single amino acid mutations of residue Q253 within the LRR domain of the *pgip-2* gene, into corresponding residues of the *pgip-1* gene, resulted in the loss of affinity for *F. moniliforme* PG. Q253 therefore, plays an essential role in the binding capacity of the complex. To test this hypothesis, residue K253 of the *pgip-1* gene was mutated to a glutamine, resulting in the protein acquiring new recognition specificities (Leckie *et al.*, 1999). Their results clearly proved that PGIP specificity and affinity for fungal PGs are determined by amino acids residing within the solvent-exposed ß-sheet/ß-turn structure. Non-synonymous variations within the LRR domains explained the ability of PGIPs from different plant sources to recognize specific ligands. This, therefore, gives a plausable explanation for the observed differential inhibition of PGIP to fungal PGs.

The Role of PGIPs in Disease Resistance

PGIP proteins purified from a vast array of plant species show differential inhibition towards pathogenic fungi (Leckie *et al.,* 1999). Several plant PGIPs are encoded by a family of genes and it is possible that each member of the PGIP family encodes a protein that differs from other members in terms of regulation and specificity for fungal PGs (Desiderio *et al.,* 1997; De Lorenzo *et al.,* 2001). The inhibition patterns from various plant sources were thus established when tested against several fungal PGs.

Abu-Goukh and Labavitch (1983) demonstrated that purified pear PG inhibitors are effective against PGs from *Apergillus niger, Botrytis cinerea* and *Dothiorella gregaria*, slightly inhibitory to PGs from *Penicillim expansum,* but that it does not affect PG activity from *Fusarium oxysporum.* Tomato PGIPs were reported to completely inhibit PG activity from *Glomerella cingulata*, but they were incapable of inhibiting *Botrytis cinerea* PGs (Brown and Adikaram, 1983). Contradictory to these results, Stotz *et al.* (1994) could identify two isoforms of the tomato PGIPs that were able to inhibit PGs from *B. cinerea,* even though their inhibition was less effective than that of the pear PGIP genes.

Purified bean PGIP was shown to have inhibitory activity against PGs produced by *C. lindemuthianum* (Lafitte *et al.,* 1983). Cervone *et al.* (1989) later showed that purified bean PGIP slows down the degradation of sodium polypectate by purified *A. niger* or *F. moniliforme* PGs, leading to the production of longer oligogalacturonides that were able to induce plant defence responses. They thus hypothesised that in addition to preventing degradation of plant cell walls by fungal PGs, PGIPs also have a role in cell signalling.

Experiments were conducted by Cook *et al.* (1999) to determine whether PGs from various fungal sources, could be inhibited by PGIPs isolated from bean, tomato and pear, which were previously reported to differ in their ability to inhibit PGs. Their results indicated that PGIPs from bean have a broad inhibition range, whereas those isolated from tomato and pear, show differential inhibition to fungal PGs. These results were confirmed by experiments conducted by Stotz *et al.* (2000), who directly

compared the effectiveness of PGIPs among these different plant species. *B. cinerea* PGs were inhibited with similar efficiency by bean, tomato and pear PGIPs, but only tomato and bean PGIPs could inhibit *A. niger* PGs. *F. moniliforme* PG in contrast was inhibited only by bean PGIPs. Similarly, the recently isolated potato PGIP also has a broad spectrum of interactions and has been shown to inhibit PGs from *A. niger, F. moniliforme* and *F. solani.* (Machinandiarena *et al.,* 2001).

PGIPs from a single plant source have also been reported to show differential inhibition towards PGs from different fungi (De Lorenzo and Cervone, 1997). At least five PGIP genes are present in *P. vulgaris*, and the two that have been characterized, differ in their abilities to inhibit PGs from *F. moniliforme* and from *Aspergillus niger* (Stotz *et al.,* 2000). Contrary to the results obtained by Cook *et al.* (1999), Di Pietro and Roncero (1996) were able to show that the tomato PGIP effectively inhibited PGs from *F. moniliforme.* Therefore, in addition to plants having multiple PGIP genes, most plant pathogens also produce multiple forms of PGs (Sharrock and Labavitch, 1994).

As single plant species may contain multiple forms of PGIP, each with nearly identical biochemical properties. It is therefore, necessary to express cloned genes in heterologous systems in order to overcome the problem of purifying single PGIPs from the bulk PGIP (De Lorenzo *et al.,* 2001). Although several expression systems such as yeast and other fungi have been used, they proved to be inadequate. Plants, however, have been shown to successfully express functional PGIPs via plant transformation or transient expression using a PVX vector (De Lorenzo *et al.,* 2001).

In an attempt to demonstrate the influence of inhibitor activity in disease development, Desiderio *et al.* (1997), expressed high levels of bean PGIP-1 in transgenic tomato plants. PGIP-1 from *P. vulgaris* showed limited ability to inhibit PGs from *F. oxysporum* f.sp *lycopersici, B. cinerea* and *Alternaria solani.* In comparison to untransformed plants, transgenic tomato plants over expressing the PGIP-1 gene did not exhibit enhanced resistance to these pathogenic fungi. This further strengthened the idea that multiple PGIP genes exist in *P. vulgaris,* each with its own PG specificity.

Powell *et al.* (2000) hypothesised that any alteration of internal plant PGIPs should reduce plant tissue maceration and necrosis by fungal endopolygalacturonases. A heterologous pear PGIP was thus expressed in tomato plants in order to evaluate whether resistance to *B. cinerea* could be increased. Their results showed that transgenic plants expressing the pear PGIP had attenuated disease symptoms in comparison to the control plants (Powell *et al.*, 2000). They further showed that over expression of PGIPs does not necessarily prevent the initial plant-pathogen interaction, but it controls subsequent development of the fungus in the plant tissues. It was thus concluded that the use of heterologous defence factors in certain host plants might effectively reduce its susceptibility to a pathogen.

The fungal pathogen of maize, *Stenocarpella maydis* causes epidemics of both ear and stalk rot in certain regions of South Africa. It has previously been reported that *F. moniliforme,* another pathogen of maize, could be inhibited by bean PGIP-2 (Leckie *et al.*, 1999). Berger *et al.* (2000) thus investigated whether the PGs produced by *S. maydis* could also be inhibited by bean PGIP. Recombinant DNA technology was used to clone the PGIP-1 gene into an *Agrobacterium*-based binary vector. Tomato cotyledons were transformed by means of *Agrobacterium*-mediated transformation and transgenic plants selected on kanamycin, were subsequently hardened off in the greenhouse. Reducing sugar assays verified that the transgenic tomato tissue extracts that do not exhibit high levels of endogenous PGIPs, expressed the bean PGIP-1 gene and biochemical assays showed that PGs from both *A. niger* as well as *S. maydis* were inhibited. However, pathogenicity tests were not conducted, as tomato is not a host for the maize pathogen *S. maydis* (Berger *et al.*, 2000).

Conclusion

This review attempted to highlight the importance of polygalacturonase-inhibiting proteins in plant defence responses. Several lines of evidence show that PGIPs are cell wall associated proteins that function to limit fungal invasion. They prevent the degradation of the pectin-rich plant cell walls by fungal polygalacturonases, by forming a stable complex with the PG thus blocking its enzymatic activity and allowing the release of oligogalacturonides that act as elicitors of defence responses.

PGIPs from various plant sources have been purified to homogeneity and several genes have also been cloned, sequenced and biochemically characterized. They exhibit characteristics similar to several reported pathogenesis related proteins and have LRR motifs that are typical of several disease resistance (*R*) genes. Multiple PGIP genes have been reported in some plant species, each having different inhibition specificities to various fungal polygalacturonases.

Several plant PGIPs have thus been cloned and expressed in heterologous plant species in order to establish the role of these proteins in resistance. The aims of the studies presented in this thesis were, therefore, to clone and characterise the full-length *pgip* gene from *E. grandis* (Chapter 2) and subsequently to transform *Nicotiana tabacum* with the *Eucalyptus pgip* gene (Chapter 3). Transgenic tobacco plants expressing the *Eucalyptus grandis pgip* gene were assessed to determine whether they could inhibit PGs produced by the *Eucalyptus* pathogen *Coniothyrium zuluense* (Chapter 3).

Chapter Two

Cloning and Characterisation of the *Eucalyptus grandis pgip* gene

CONTENTS

Abstract

Polygalacturonase-inhibiting Proteins (PGIPs), associated with the cell wall of many plant species form part of the plant's active defence against phytopathogenic fungi. PGIPs are leucine-rich repeat proteins that are evolutionarily related to several plant resistance (*R*) genes, involved in gene-for-gene reactions. The objective of this study was to clone and characterize the complete *Eucalyptus pgip* gene. Based on the alignments between the pear *pgip* and partial *Eucalyptus pgip* isolated in a previous study, gene-specific primers were designed and used to amplify a *pgip* fragment of 1020bp 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 downstream of the 3' termination codon. Genome-walking PCR was, therefore, 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 domains 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) (De Lorenzo and Ferrari, 2002). PGIPs form specific, reversible, saturable, high-affinity complexes with fungal endopolygalacturonases, thereby activating plant defense responses (Cervone *et al.*, 1989). 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 (Jones and Jones, 1997).

PGIPs have been purified to homogeneity from several dicotyledonous plants, including bean, soybean, tomato, pear, apple, raspberry, potato, and very recently, cotton and guava (Salvi *et al.*, 1990; Abu-Goukh *et al.*, 1993; Johnston *et al.*, 1993; Stotz *et al.*, 1993, 1994; Favaron *et al.*, 1994; Yao *et al.*, 1995; Machiandiarena *et al.*, 2001; James and Dubery, 2001; Deo and Shastri, 2003). They have also been identified in the pectin-rich monocotyledonous plants such as, onion and leek (Favaron *et al.*, 1993, 1997).

The first gene encoding a PGIP was cloned from *Phaseolus vulgaris* (Toubart *et al.*, 1992). *Pgip* genes have since been cloned from many other plant species, with the implementation of several PCR techniques. Most *pgip* genes that have been characterized contain a single open reading frame, approximately 1kb in length that is not interrupted by introns (Toubart *et al.*, 1992; Stotz *et al.*, 1993; Chimwamurombe *et al.*, 2001). However, the raspberry (*Rubus idaeus*) PGIP, contains an internal sequence region that showed no similarity to previously isolated PGIP clones (Ramanathan *et al.*, 1997). A highly hydrophobic region corresponds to the N-terminal signal peptide, which targets PGIPs to the endomembrane system for export to the extracellular space (Toubart *et al.*, 1992; Stotz *et al.*, 1993; Powell *et al.*, 1994; Ramanathan *et al.*, 1997). The positions of eight cysteine residues are conserved in all *pgip* genes (Toubart *et al.*, 1992; Stotz *et al.*, 1993, 1994; Yao *et al.*, 1995).

Several PCR techniques have been employed to isolate *pgip* genes from plants. PCR is a powerful tool developed and used to amplify microgram amounts of a specific, targeted gene (Parker *et al.,* 1991). Typical PCR reactions utilize oligonucleotide primers that hybridize to opposite strands, so that extension proceeds inwards between the two primers (Ochman *et al.,* 1988). A limitation to the technology is the inability to amplify regions that lie outside the boundaries of known sequences. In this case, several major molecular modifications are required, which include cloning and ligation of adaptors to the ends of amplified sequences (Parker *et al.,* 1991). Several methods have 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 (Siebert *et al.,* 1995).

Siebert *et al.* (1995) have developed an improved gene walking technique for walking in uncloned genomic DNA regions (Figure 2.1). The method is based on the ligation of a special adaptor to genomic DNA, which has been digested with restriction enzymes that generate blunt-ends. The sequences of the genome-walking adaptor and adaptor primers are shown in Table 1. The blunt-end of the adaptor ligates to any blunt-end of the genomic DNA. The desired genomic region is amplified with an adaptor primer, which is specific to the outer part of the adaptor and a second gene-specific primer. The adaptor primers are much shorter than the adaptors, thus allowing the amplification of only the desired fragments. PCR products that contain double-stranded adaptor sequences at both ends, form panhandle structures (PS-effect) that prevent the exponential amplification of the desired product.

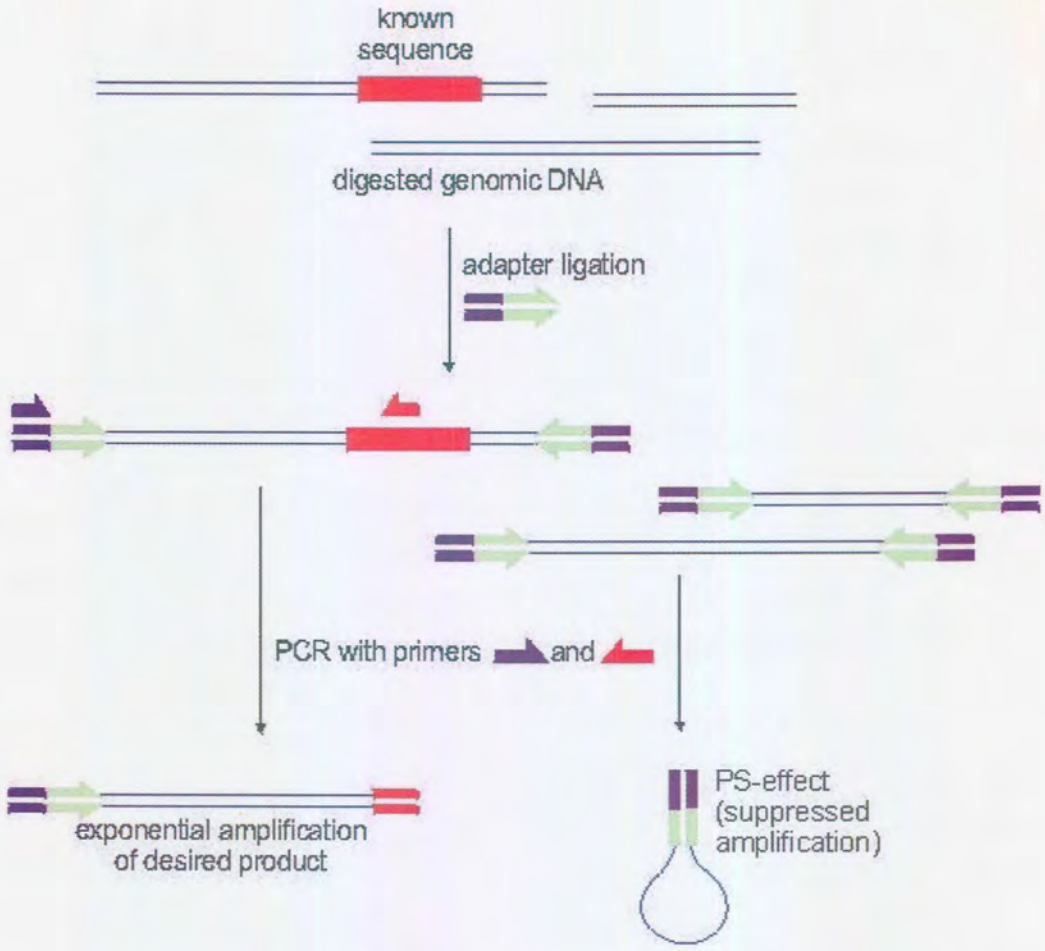

Figure 2.1 Schematic representation of the amplification of unknown genomic DNA segments by means of Genome Walking adapted from www.euregene.com.

In this chapter, a study focused on cloning and characterisation of a *Eucalyptus pgip* gene is presented. In a previous study by Chimwamurombe *et al.* (2001) a partial *Eucalyptus pgip* gene was isolated. Sequence analysis revealed the absence of the putative signal peptide and the downstream termination codon. The isolated PGIP also showed a high degree of amino acid identity to the pear *pgip* gene. Using the genome-walking technology, I chose to 'walk' upstream as well as downstream of the known PGIP sequence, in order to determine the signal peptide sequence and thus the complete *Eucalyptus* PGIP sequence. In addition, I was interested to know whether the same degree of amino acid identity is maintained between the complete *Eucalyptus* and the pear *pgip* gene sequences.

Materials and Methods

Genomic DNA Extraction

One hundred milligrams of fresh, young leaves were collected from a *Eucalyptus grandis* TAG5 clone. Genomic DNA was extracted from the leaves using the DNeasy Plant Mini Kit (QIAGEN) as recommended by the manufacturer's protocols. The plant material was ground to a fine powder in liquid nitrogen followed by the addition of a lysis buffer containing RNaseA. The mixture was incubated at 65°C to allow the reaction to proceed. Proteins, polysaccharides and detergents were subsequently salt precipitated. The cell debris were removed by a brief spin through QIAshredder ™ filtration unit. The cleared lysate was transferred to a new tube to which a binding buffer and ethanol were added. The mixture was applied to a DNeasy mini spin column placed in a 2ml collection tube, which was then centrifuged enabling the binding of the DNA to the DNeasy membrane. The DNeasy spin column was placed into a new 2ml collection tube and the membrane was washed twice allowing contaminants such as proteins and polysaccharides to be efficiently removed. Following the wash steps, the DNeasy spin column was placed into a sterile 1,5ml tube and purified DNA was eluted in 100µl of TE buffer.

DNA Amplification

Gene specific primers, PC6 (5'-ACATCTCTCAGGCTCTCAACC-3', Stotz *et al.,* 1993) and SPPGIP2 (5'-GCAGTGTGGAGGGGTGCACCACACAGGCA-3', Chimwamurombe *et al.,* 2001) were designed from the conserved regions of pear and *Eucalyptus* PGIP sequences respectively. Polymerase chain reactions were performed in 0,2ml thin walled tubes in the GeneAmp PCR system 9700 (Applied Biosystems). The following reaction mixture was used: oligonucleotide primer SPPGIP2 (0.2µM), oligonucleotide primer PC6 (0.2µM), 40ng *Eucalyptus* genomic DNA, 10 X PCR buffer, 0.25mM dNTPs, 1.5mM $MgCl_2$, 5 units *Taq* polymerase (Roche Diagnostics, Germany) in 50µl. PCR was performed for 30 cycles (one cycle = 1 min at 94ºC, 2min at 55ºC, 2min at 72ºC). The

reaction had an initial denaturation step of 2min at 94°C and a final elongation step of 7min at 72°C. Control PCR amplification was performed using 1ng of clone P1 plasmid DNA containing a partial *Eucalyptus* PGIP fragment (obtained from P. Chimwamurombe). The PCR products were separated along with a λIII marker and a 100bp DNA ladder (Promega), on a 1% (w/v) agarose/TAE (pH 8.0) gel stained with ethidium bromide (Sambrook *et al.,* 1989) and visualised under an ultraviolet light.

Cloning of the *Eucalyptus* PGIP fragment

Ligation into pGEM-T-Easy

An approximately 1.2kb *Eucalyptus* PGIP PCR fragment was purified from the gel, using the QIAquick Gel Extraction Kit Protocol (Qiagen, Germany). Ligation was performed with a 3:1 insert-to-vector molar ratio. A 10μl ligation reaction was prepared containing 50ng pGEM-T-Easy vector (Promega), 60ng PCR product, 2X Rapid Ligation Buffer and 3U T4 DNA ligase. A positive control using the control insert DNA (Promega) provided was included to assess transformation efficiency. A ligation reaction in which the insert DNA was omitted was included as a background control. The reactions were incubated for 16 hours at 4°C.

Transformation and Plasmid DNA isolation

The ligated plasmids were transformed into competent *E.coli* (JM109) cells (Promega). 25μl of competent cells were added to 2μl of each ligation reaction and the mixture was incubated on ice for 20 min. This was followed by a heat-shock treatment for 45 sec at 42°C. A vol. of 975μl SOC medium containing 50mM glucose and 1M Mg^{2+} was added to each reaction and incubated at 37°C for one hour. Transformants were screened on LB (Luria Bertani) plates containing 100μg/ml ampicillin, 0,5mM isopropyl-β-D-thiogalactopyranoside (IPTG)(Boehringer Mannheim) and 80μg/ml 5-bromo-4-chloro-3-indolyl-β-D-galactoside (X-gal). The plates were incubated at 37°C overnight.

Isolation of plasmid DNA was achieved using a modification of the alkaline lysis method of Sambrook *et al.* (1989). Recombinant white colonies were selected and individually inoculated into 5ml LB medium supplemented with 50μg/ml ampicillin. Following an overnight incubation at 37°C, the bacterial cells were collected by centrifugation and re-suspended in 100μl of Solution 1 (Appendix B). Cells were lysed by the addition of 200μl of Solution 2 (Appendix B) and incubation on ice for a period of 5min. The sample was then neutralized by adding 150μl of Solution 3 (Appendix B) and the resulting cell debris were removed by centrifuging the sample for 15 min at 4°C. The clear supernatant solution was precipitated with 2.5 vol.s of cold, absolute ethanol and the DNA was then pelleted at 4°C. Excess salts were washed from the sample with 70% ethanol and the pellet was dried and the plasmid DNA was re-suspended in TE buffer (pH 8). An RNaseA (10mg/ml) digestion was performed for each sample at 37°C for 30 min. The purified plasmid DNA products were separated on a 1% (w/v) agarose gel.

Screening of Positive Transformants

The putative recombinant clones were screened by restriction enzyme digestion. A 20μl reaction was set up containing 200ng (10μl) plasmid DNA, 2μl restriction enzyme Buffer H and 10 U *Eco*RI (Roche). Digestion was allowed to proceed for 3 hours at 37°C followed by separation on a 1% (w/v) agarose gel. The isolated plasmid DNA was used as a template in PCR reactions and a 25μl PCR reaction was set up as described in Chapter 3. Gene specific primes EN-PGIP-FWD and REV-PB-PGIP were designed in this study and used to amplify the pgip fragment from the putative recombinant clones. Thirty PCR cycles were used with one cycle of 94°C for 1 min, 62°C for 1 min and 72°C for 1 min, an initial denaturation step of 94°C for 2 min and a final elongation step of 72°C for 7 min. PCR products were analysed by electrophoresis through a 1% (w/v) agarose gel in 1 X TAE buffer (pH 8.0) (Appendix B).

Nucleotide sequencing of the *pgip* fragment

A positive recombinant clone designated as pGEM-*Eucpgip*110B was selected and sequenced using the BIG Dye terminator cycle sequencing kit with an ABI Prism 3100 sequencer (ABI Advanced Biotechnological Institute, Perkin-Elmer Corporation, Foster City, USA). T7 and SP6 promoter primers were used to sequence the double-stranded DNA insert. The sequences were analysed using the ABI Prism® SeqScape Software Version 1.0 (Applied Biosystems) program and a BLASTX search was done to determine regions of homology to other published PGIP polypeptide sequences (http://www.ncbi.nlm.nih.gov). The partial *Eucalyptus* peptide sequence was aligned to the pear PGIP peptide using the CLUSTAL (EBI) software (http://www2.ebi.uk).

Genome Walking

Restriction Enzyme Digestion

Eighty nanograms of *Eucalyptus grandis* TAG5 genomic DNA was digested at 37°C with 10U/µl blunt end restriction enzymes that included *Hae*III, *Dra*I, *Sca*I and *Sma*I, in a total vol. of 50µl. An aliquot was analysed on a 1% (w/v) agarose gel in 1 x TAE buffer (pH 8.0) to determine whether complete digestion had been achieved. The total digested DNA was precipitated by adding 1/10 th vol. 3M NaOAc (pH 5.3) and 2.5 vol. of 96% EtOH. The samples were incubated on ice for 30 min followed by centrifugation at 15000 rpm in a microcentrifuge for 15 min. The pellets were washed with 70% EtOH and immediately centrifuged for 5 min at 15000 rpm. After air-drying, the pellets were re-suspended in 50µl of 1 x TE buffer.

Adaptor Ligation

The adaptors and adaptor primers used for genome walking in this study were designed as described by Siebert *et al.* (1995). A 5µM genome walking adaptor (Siebert *et al.*, 1995; Table 1) was ligated to 8µl of digested DNA overnight at 16°C with 5 U T4 DNA ligase in

a total vol. of 30µl. The ligation reaction was terminated by incubation of the samples at 70°C for 5 min. The samples were then diluted to a final vol. of 100µl with sterile distilled water.

Table 2.1 Nucleotide sequence of the genome walking adaptor and adaptor primers used in this study.

Primer	Nucleotide Sequence
Adaptor	5'CTAATACGACTCACTATAGGGCTCGAGCGGCCGCCCGGGCAGGT 3'
	3' H_2NCCCGTCCA 5'
ADP1	5'GGATCCTAATACGACTCACTATAGGGC 3'
ADP2	5'AATAGGGCTCGAGCGGC 3'

Genome Walking PCR

Genome walking primer pairs, N-PGIP-L1 and PGIP-Inv-L2 and N-PGIP-R1-1 and PGIP-Inv-R3 were designed in opposite orientation to normal PCR primers using the pGEM-*Eucpgip*110B sequencing data. Characteristics of the adaptor primers used for genome walking are shown in Table 2.1. All PCR reactions were performed in 0.2ml thin walled tubes in the GeneAmp PCR system 9700 (Applied Biosystems). Primary PCR reactions were performed using ADP1 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. PCR conditions were as follows: 0.5µl of ligated and diluted DNA, 0.45µM adaptor primer ADP1; 0.45µM N-PGIP-L2 or N-PGIP-R1-1; 0.2µM dNTPs; 1U of Expand *Taq* polymerase (Roche). The cycle parameters included an initial denaturation at 96°C for 1 min followed by 35 cycles of denaturation at 94°C for 30 sec and annealing/extension at 68°C for 6 min, and a final elongation at 68°C for 15 min. Secondary PCR reactions were performed using ADP2 and PGIP-Inv-L2 for the upstream region and primers ADP2 and

PGIP-Inv-R3 for amplifying the downstream 3' region. PCR conditions were the same as for the primary reactions, but the cycle parameters differed slightly. Twenty cycles of 94°C for 20sec, 62°C for 30 sec and 68°C for 1 min with an initial denaturation step at 96°C for 2 min and a final elongation at 68°C for 1 min, was performed. A negative control containing all PCR reagents, except template DNA was also included in both primary and secondary PCR reactions. PCR products were analysed on a 1% (w/v) agarose gel in 1 x TAE buffer (pH 8.0).

Cloning of the genome walking PCR products

Ligation into pCR®2.1

The QIAquick gel extraction kit protocol (Qiagen, Germany) was used to purify the PCR products from a 1% (w/v) agarose gel. The PCR fragments were ligated into the multiple cloning site of the pCR®2.1 vector provided in the TA® Cloning Kit (Invitrogen). A 10µl ligation reaction was prepared containing 50 ng PCR product, 25ng pCR®2.1 and 4 Weiss units T4 DNA Ligase. Ligation reactions were incubated overnight at 14°C.

Transformation

The ligated samples were transformed into One Shot®cells (TOP10F'). Two microlitres of each ligation reaction were added to 50 µl of One Shot®cells and gently mixed with the pipette tip. The vials were incubated for 30 min on ice, followed heat shocking the samples for exactly 30 sec in a 42°C water bath. SOC medium was added to each sample, which was then incubated at 37°C for 1h on a shaking incubator at 225rpm. A hundred microlitres of each transformation reaction was then spread on LB agar plates supplemented with 100µg/ml ampicillin, 0.5µM IPTG (Boehringer Mannheim) and 80µg/ml X-gal. The plates were incubated overnight at 37°C.

Plasmid DNA isolation and screening of putative transformants

Thirty, white recombinant colonies were selected and subjected to plasmid DNA isolation. The colonies were inoculated individually in 5ml LB medium supplemented with 50μg/ml ampicillin and grown overnight at 37°C. Plasmid DNA was isolated from the bacterial cultures using the Wizard®Plus SV Miniprep DNA Purification System (Promega). The putative recombinants were analysed by restriction enzyme digestion as described by Sambrook *et al.* (1989) as well as PCR analysis using M13 forward and M13 reverse primers. Two positive clones designated as pCR-*ScaI*GW-4 (Upstream fragment) and pCR-*DraI*-GW2 (Downstream fragment) were sequenced as described in Chapter 3, using the M13 forward and M13 reverse primers. Sequences were analysed using the ABI Prism ® SeqScape Software Version 1.0 (Applied Biosystems) program.

Table 2.2 Characteristics of all oligonucleotide primers used in this study

Primer	Size (bp)	Sequence 5'-3'	Tm	%GC
PC6	21	ACATCTCTCAGGCTCTCAACC	59	52
SPPGIP2	29	GCAGTGTGGAGGGGTGCACCACACAGGCA	74[a]	65
N-PGIP-L1	20	CCATGAGGCCAAGACATAGG	61	55
N-PGIP-R1-1	20	CTGAACGTGAGCTACAACAG	60	50
PGIP-Inv-L2	20	GGACGGAGGAGAAGAGTAGG	62	60
PGIP-Inv-R3	20	CCATAACCGATGCCTGTGTG	62	55
ADP1	27	GGATCCTAATACGACTCACTATAGGGC	66[a]	48
ADP2	17	AATAGGGCTCGAGCGGC	56	65
T7	20	ATTATGCTGAGTGATATCCC	56	40
SP6	21	ATTCTATAGTGTCACCTAAAT	54	30
M13 F	18	GGTTTTCCCAGTCACGAC	56	55
M13 R	20	AGCGGATAACAATTTCACAC	56	40

[a] Tm calculated using the website : http://www.basic.nwv.edu/biotools/oligocalc.html.

The melting temperature (Tm) was calculated using the Wallace formula :

Tm = 4 (G + C) + 2 (A + T)

G, C, A, T – number of respective nucleotides in the primer.

Results

Isolation of the partial *Eucalyptus grandis pgip* gene

In a previous study, Chimwamurombe *et al.* (2001) used degenerate oligonucleotide primers PC1 (5' GGAATTCAAYCCNGAYGAYAARGT 3') (Stotz *et al.*, 1993) and Per 3 (5'RCANWSNGGNARNGGNGCNCCRCANARRCA 3') (Chimwamurombe *et al.*, 2001) to amplify a partial *Eucalyptus pgip* gene of 909bp (Figure 2.2). The predicted 298 amino acid polypeptide lacked both the putative signal sequence as well as the downstream translation termination signal. It showed 95% amino acid sequence identity to the published pear PGIP, thus allowing the design of gene-specific primers corresponding to the N-terminal region of pear pgip (PC6) and the C-terminal region of the partial *Eucalyptus* pgip (SPPGIP2). The SPPGIP2 primer was designed from the sequence of the degenerate primer, Per3 of pear PGIP (Stotz *et al.*, 1993) and the partial *Eucalyptus pgip* sequence obtained by Chimamurombe *et al.* (2001). The gene-specific primer PC6 was designed from a region upstream of the first ATG and it thus lead to the amplification of the signal peptide sequence.

```
                                                                   PC6 ⇒
Pear DNA                                                       ACA TCT CTC AGG CTC TCA ACC AAA ACC AAA ACA    [33]

                                                                                                        ▼
Pear AA     M   E   L   K   F   S   T   F   L   S   L   T   L   L   F   S   S   V   L   N   P   A   L   S   D     [25]
Pear DNA   ATG GAA CTC AAG TTC TCC ACC TTC CTC TCC CTA ACC CTA CTC TTC TCC TCC GTC CTA AAC CCC GCT CTC TCC GAT   [108]

         G GAA TTC AAY CCN GAY GAY AAR AAR GT  PC1 ⇒
Euc DNA    GAA TTC AAT CCG GAC GAC AAG AAG GTC CTC CTA CAA ATC AAG AAA GCC TTC GGC GAC CCC TAT GTC TTG GCC TCA
Euc AA                                      V   L   L   Q   I   K   K   A   F   G   D   P   Y   V   L   A   S

Pear AA     L   C   N   P   D   D   K   K   V   L   L   Q   I   K   K   A   F   G   D   P   Y   V   L   A   S     [50]
Pear DNA   CTC TGC AAC CCC GAC GAC AAA AAA GTC CTC CTA CAA ATC AAG AAA GCC TTC GGC GAC CCC TAC GTC TTG GCC TCA   [183]

Euc DNA    TGG AAA TCA GAC ACC GAC TGT TGT GAT TGG TAC TGC GTC ACC TGT GAC TCA ACC ACA AAC CGC ATC AAC TCC CTC
Euc AA      W   K   S   D   T   D   C   C   D   W   Y   C   V   T   C   D   S   T   T   N   R   I   N   S   L

Pear AA     W   K   S   D   T   D   C   C   D   W   Y   C   V   T   C   D   S   T   T   N   R   I   N   S   L     [75]
Pear DNA   TGG AAA TCA GAC ACT GAC TGC TGC GAT TGG TAC TGC GTC ACC TGT GAC TCC ACC ACA AAC CGC ATT AAC TCC CTC   [258]

Euc DNA    ACC ATC TTT GCC GGC CAG GTA TCC GGC CAA ATC CCC GCC CTA GTT GGA GAC TTG CCG TAC CTT GAA ACC CTT GAA
Euc AA      T   I   F   A   G   Q   V   S   G   Q   I   P   A   L   V   G   D   L   P   Y   L   E   T   L   E

Pear AA     T   I   F   A   G   Q   V   S   G   Q   I   P   A   L   V   G   D   L   P   Y   L   E   T   L   E    [100]
Pear DNA   ACC ATC TTT GCC GGC CAG GTG TCA GGC CAA ATC CCC GCC CTA GTA GGA GAC TTG CCA TAC CTT GAA ACC CTT GAA   [333]

Euc DNA    TTC CAC AAG CAA CCC AAT CTC ACT GGC CCA ATC CAA CCC GCC ATT GCC AAG CTC AAA GGA CTC AAG TTT CTC AGG
Euc AA      F   H   K   Q   P   N   L   T   G   P   I   Q   P   A   I   A   K   L   K   G   L   K   F   L   R

Pear AA     F   H   K   Q   P   N   L   T   G   P   I   Q   P   A   I   A   K   L   K   G   L   K   S   L   R    [125]
Pear DNA   TTC CAT AAG CAA CCC AAT CTC ACT GGC CCA ATC CAA CCC GCC ATT GCC AAG CTC AAA GGA CTC AAG TCT CTC AGG   [408]
```

```
Euc DNA   CTC AGC TGG ACC AAC CTC TCA GGC TCT GTC CCT GAC TTC CTC AGC CAA CTC AAG AAC CTC ACA TTC CTC GAC CTC
Euc AA     L   S   W   T   N   L   S   G   S   V   P   D   F   L   S   Q   L   K   N   L   T   F   L   D   L

Pear AA    L   S   W   T   N   L   S   G   S   V   P   D   F   L   S   Q   L   K   N   L   T   F   L   D   L   [150]
Pear DNA  CTC AGC TGG ACC AAC CTC TCA GGC TCT GTC CCT GAC TTC CTC AGC CAA CTC AAG AAC CTC ACA TTC CTC GAC CTC   [483]


Euc DNA   TCC TTC AAC AAC CTC ACC GGC GCC ATC CCC AGC TCG CTT TCT CAG CTC CCA AAC CTC AAC GCT CTT CAT CTA GAC
Euc AA     S   F   N   N   L   T   G   A   I   P   S   S   L   S   Q   L   P   N   L   N   A   L   H   L   D

Pear AA    S   F   N   N   L   T   G   A   I   P   S   S   L   S   E   L   P   N   L   G   A   L   R   L   D   [175]
Pear DNA  TCC TTC AAC AAC CTC ACC GGT GCC ATC CCC AGC TCG CTT TCT GAG CTC CCA AAC CTC GGC GCT CTT CGT CTA GAC   [558]


Euc DNA   CGC AAT AAG CTC ACA GGT CAT ATT CCG AAA TCG TTT GGG CAG TTC ATT GGC AAC GTT CCA GAC CTG TAT CTC TCC
Euc AA     R   N   K   L   T   G   H   I   P   K   S   F   G   Q   F   I   G   N   V   P   D   L   Y   L   S

Pear AA    R   N   K   L   T   G   H   I   P   I   S   F   G   Q   F   I   G   N   V   P   D   L   Y   L   S   [200]
Pear DNA  CGC AAT AAG CTC ACA GGT CAT ATT CCG ATA TCG TTT GGG CAG TTC ATT GGC AAC GTT CCA GAC CTG TAT CTC TCC   [633]


Euc DNA   CAC AAC CAG CTC TCG GGC AAC ATT CCA ACC TCA TTT GCC CAG ATG GAC TTC GGC AAG CAT AGA CTA TCA CGG AAC
Euc AA     H   N   Q   L   S   G   N   I   P   T   S   F   A   Q   M   D   F   G   K   H   R   L   S   R   N

Pear AA    H   N   Q   L   S   G   N   I   P   T   S   F   A   Q   M   D   F   T   S   I   D   L   S   R   N   [225]
Pear DNA  CAC AAC CAG CTT TCT GGT AAC ATT CCA ACC TCA TTC GCT CAG ATG GAC TTC ACC AGC ATA GAC TTA TCA CGG AAC   [708]


Euc DNA   AAG CTC GAG  -  GAC GCA TCA GTG ATA TTT GGG CTG AAC AAG ACA ACC CAG ATT GTG GAC CTA TCC AGG AAC TTG
Euc AA     K   L   E   -   D   A   S   V   I   F   G   L   N   K   T   T   Q   I   V   D   L   S   R   N   L

Pear AA    K   L   E   G   D   A   S   V   I   F   G   L   N   K   T   T   Q   I   V   D   L   S   R   N   L   [250]
Pear DNA  AAG CTC GAA GGT GAC GCA TCC GTG ATA TTT GGG CTG AAC AAG ACA ACC CAG ATT GTG GAC CTG TCC AGG AAC TTG   [783]


Euc DNA   CTG GAA TTT AAT CTG TCA AAG GTG GAG TTT CCG ACA AGC TTG ACC TCA CTG GAT GTA AAC CAC AAT AAG ATC TAC
Euc AA     L   E   F   N   L   S   K   V   E   F   P   T   S   L   T   S   L   D   V   N   H   N   K   I   Y

Pear AA    L   E   F   N   L   S   K   V   E   F   P   T   S   L   T   S   L   D   I   N   H   N   K   I   Y   [275]
Pear DNA  CTG GAA TTT AAT CTG TCA AAG GTG GAG TTT CCG ACA AGC TTG ACC TCG CTG GAT ATC AAC CAC AAT AAG ATC TAC   [858]
```

```
Euc DNA   GGG AGT ATC CCA GTG GAG TTT ACC CAA CTG AAT TTC CAG TTC CTG AAC GTG AGC TAC AAC AGG CTG TGT GGT CAG
Euc AA     G   S   I   P   V   E   F   T   Q   L   N   F   Q   F   L   N   V   S   Y   N   R   L   C   G   Q

Pear AA    G   S   I   P   V   E   F   T   Q   L   N   F   Q   F   L   N   V   S   Y   N   R   L   C   G   Q   [300]
Pear DNA  GGG AGT ATC CCA GTG GAG TTT ACG CAA CTG AAT TTC CAG TTC CTG AAC GTG AGC TAC AAC AGG CTG TGT GGT CAG   [933]


Euc DNA   ATT CCA GTG GGC GGA AAG TTG CAA AGC TTC AAC GAG TAT TCT TAT TTC CAT AAC CGA
Euc AA     I   P   V   G   G   K   L   Q   S   F   N   E   Y   S   Y   F   H   N   R

Pear AA    I   P   V   G   G   K   L   Q   S   F   D   E   Y   S   Y   F   H   N   R                       [319]
Pear DNA  ATT CCT GTG GGT GGA AAG TTG CAG AGC TTC GAC GAG TAT TCT   TAT TTC CAT AAC CGA                       [990]

Euc DNA   TGC CTG TGT GGT GCA CCC CTC CAC ACT GC
       3'ACR RAN ACR CNC GNG GNR ANG GNS WNA CR-5' ⇐ Per3

Pear AA    C   L   C   G   A   P   L   P   S   C   K   *                                                   [330]
Pear DNA  TGC TTG TGC GGT GCT CCA CTC CCA AGC TGC AAG TAA                                                   [1026]
       3' ACG GAC ACA CCA CGT GG- GGA GGT GTG ACG -5' ⇐ SPPGIP2
```

Figure 2.2 Sequence information used for PCR primer design to amplify the complete *Eucalyptus grandis pgip* gene. The nucleotide and amino acid sequences of the partial *Eucalypus pgip* gene (Chimwamurombe *et al.*,2001) and the published pear *pgip* gene (Stotz *et al.*, 1993) that were available at the start of the study are shown.

Eucalyptus grandis (TAG5) genomic DNA was used in a PCR reaction with gene-specific primers PC6 and SPPGIP2 to amplify a 1020bp fragment. An annealing temperature of 55°C enabled the exponential amplification of an expected fragment of approximately 1000bp. Amplification products are illustrated in figure 2.3.

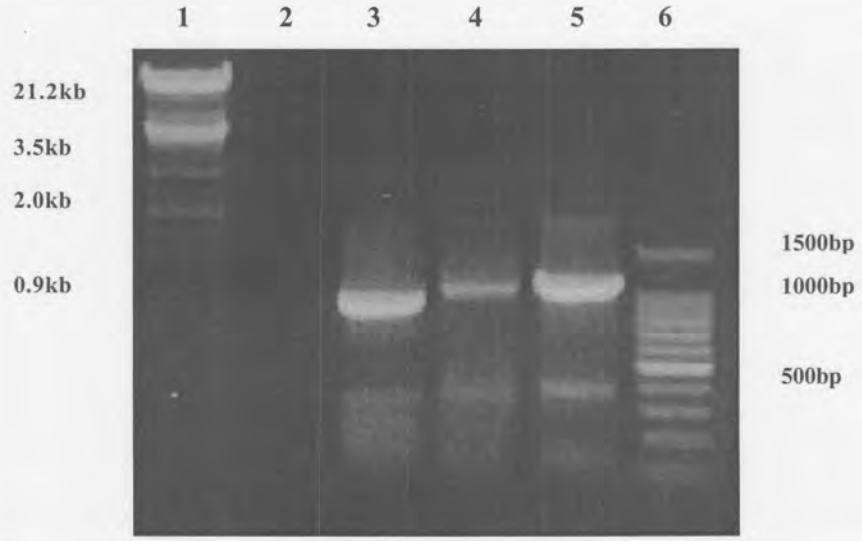

Figure 2.3 PCR amplification of the *pgip* gene from *Eucalyptus grandis* genomic DNA. Lane 3 contains the product of amplification with primers PC1 and Per3 using the plasmid P1 containing the partial *Eucalyptus* pgip (Chimwamurombe *et al.*, 2001). Lanes 4 and 5 contain PCR fragments amplified from genomic DNA using gene-specific primers PC6 and SPPGIP2. The fragments obtained are approximately 100bp larger than those obtained for the positive control. Lane 2 represents the negative control in which no template DNA was included. Lanes 1 and 6 are the molecular markers. Lane 1 contains λ DNA digested with *Hind*III and *Eco*RI, while lane 6 contains the commercial 100bp ladder (Promega).

The 1020bp pgip fragment was cloned into the commercial pGEM-T-Easy cloning vector (Promega) and twelve putative recombinant clones designated as pGEM-*Eucpgip*110A to 110L were subjected to restriction analysis using *Eco*RI. The restriction profile is

illustrated in figure 2.4. Five of the 12 clones contained the *Eucalyptus* pgip insert, which was cut into two fragments indicating the occurrence of an *Eco*RI restriction site within the *pgip* gene.

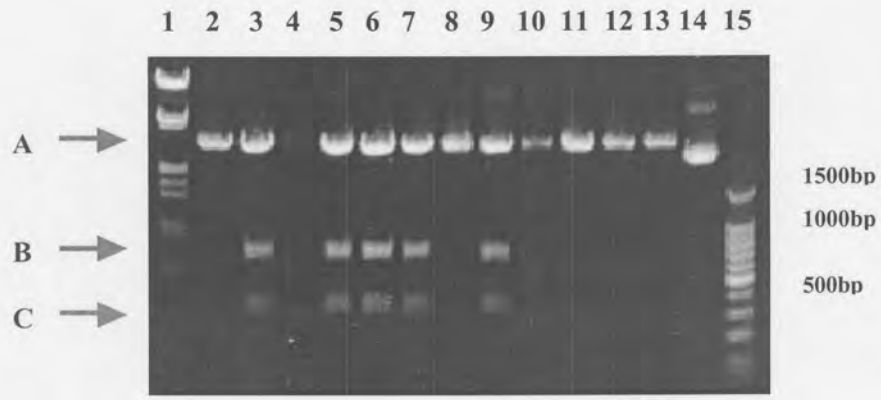

Figure 2.4 Restriction enzyme digest of putative recombinant clones with *Eco*RI. Lanes 2-13 represents clones pGEM-*Eucpgip*110A to 110L digested with *Eco*RI. Lane 14 represents an undigested clone and Lanes 1 and 15 contain the λ *Hind*III/*Eco*R1 marker and 100bp DNA ladder, respectively. Digestion with *Eco*RI gave three fragments in the recombinant clones. These are represented as, **(A)** 3015bp linearized pGEM-T-Easy vector, **(B)** 695bp fragment and **(C)** 337bp fragment.

The recombinant clones were screened by PCR analysis using a second set of gene-specific primers designed from the nucleotide sequences of the pGEM-*Eucpgip* clone and the published pear *pgip* gene sequence. The EN-FWD-PGIP primer was designed from the 5' region of the partial eucalyptus *pgip* gene and contains an *Nco*I restriction site. The reverse primer, REV-PB-PGIP was designed from the 3' region of the pear *pgip* gene and contains a *Bam*HI site. The design and use of these primers are discussed in detail in chapter 3. Clones containing the inserts, as established from restriction analysis, were selected and used as templates in PCR reactions with the EN-FWD-PGIP and REV-PB-PGIP primers. A fragment of approximately 1kb was amplified from each of the clones screened (Figure

2.5). The plasmid map of the constructed recombinant pGEM-*Eucpgip*110B clone is indicated in figure 2.6.

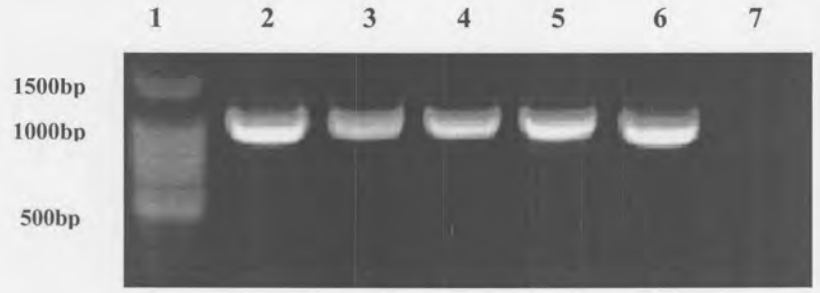

Figure 2.5 PCR amplification of recombinant pGEM-*Eucpgip* clones. Lane 1 contains the 100bp DNA ladder (Promega). Lanes 2-6 contains amplification products of the recombinant pGEM-*Eucpgip* clones 110B, 110D, 110E, 110F and 110H, respectively, using gene-specific primes EN-PGIP-FWD and REV-PB-REV. Lane 7 contains the negative ddH_2O control.

Sequence Analysis of the partial *Eucalyptus* PGIP

A recombinant PCR clone pGEM-*Eucpgip*110B was selected and sequenced with T7 and SP6 primers. Direct sequencing using gene-specific primers EN-FWD-PGIP and REV-PB-PGIP was also performed on the PCR product amplified from genomic DNA (TAG5), yielding the same internal nucleotide sequence as that of the pGEM-*Eucpgip*110B clone (Figure 2.7). Amplification with the gene-specific PC6 primer, designed from the N-terminal region of the pear PGIP, enabled determination of the start codon (ATG) and signal sequence of PGIP, which was not included in the partial *Eucalyptus* PGIP fragment sequenced by Chimwamurombe *et al.* (2001).

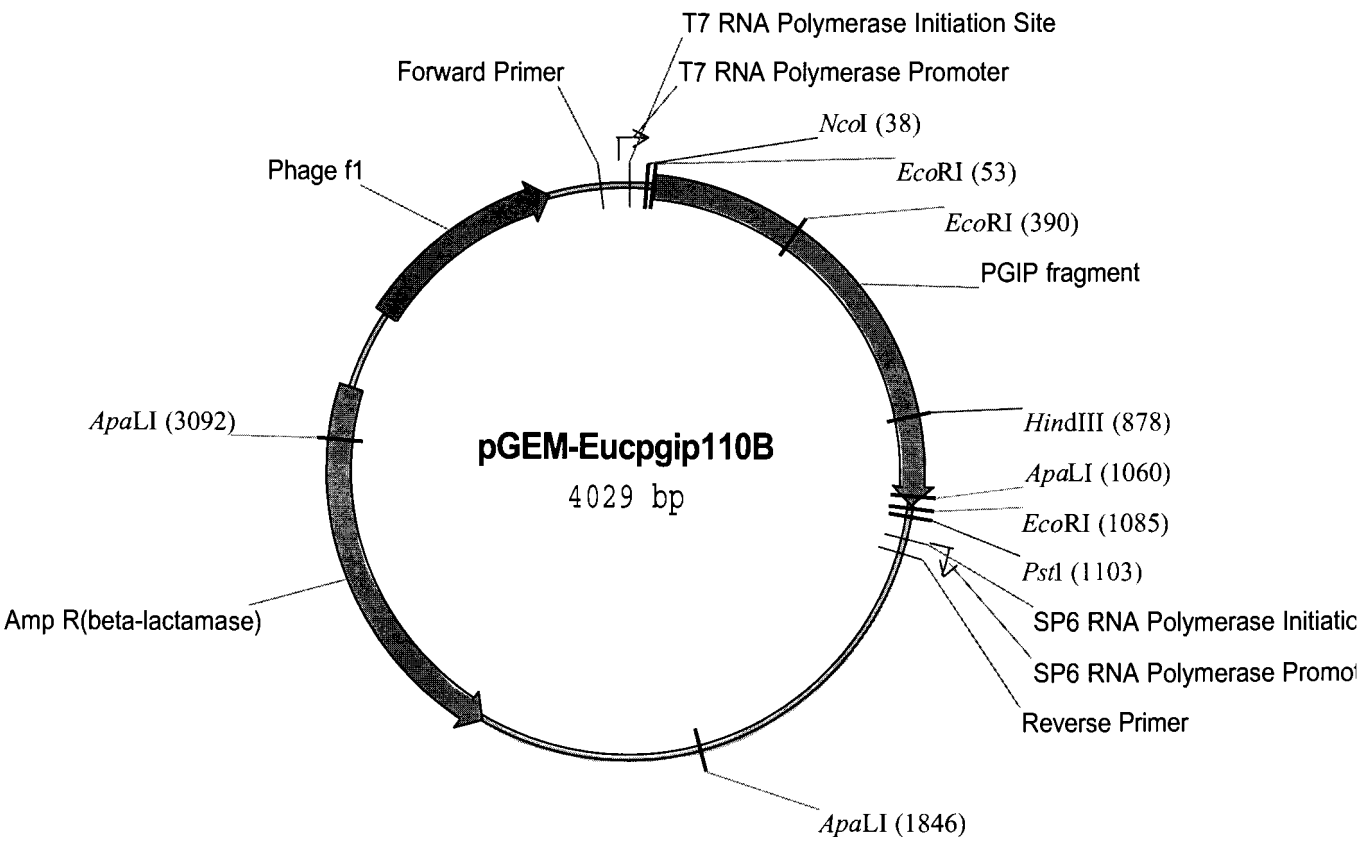

Figure 2.6 Plasmid map of the constructed pGEM-*Eucpgip*110B clone containing the *Eucalyptus pgip* fragment.

PC6 ⇒

pGEM-*Eucpgip*110B ACATCTCTCA GGCTCTCAAC CAAAACCCAA AACAATGGAG CTCAAGTTCT TCACCTTCCT CTCCCTAACC CTACTCTTCT CCTCCGTCCT AAACCCCGCT
Eucalyptus ....CCCCGCT
Pear ACATCTCTCA GGCTCTCAAC CAAAACC.AA AACAATGGAA CTCAAGTTCT CCACCTTCCT CTCCCTAACC CTACTCTTCT CCTCCGTCCT AAACCCCGCT
Pear Amino Acid M E L K F S T F L S L T L L F S S V L N P A

pGEM-*Eucpgip*110B CTCTCCGATC TCTGCAACCC CGACGACAAA AAAGTCCTCC TACAAATCAA GAAAGCCTTC GGCGACCCCT ATGTCTTGGC CTCATGGAAA TCAGACACCG
Eucalyptus CTCTCCGATC TCCGCAACCC CGACGACAAA AAAGTCCTCC TACAAATCAA GAAAGCCTTC GGCGACCCCT ATGTCTTGGC CTCATGGAAA TCAGACACCG
Pear CTCTCCGATC TCTGCAACCC CGACGACAAA AAAGTCCTCC TACAAATCAA GAAAGCCTTC GGCGACCCCT ACGTCTTGGC CTCATGGAAA TCAGACACTG
Pear Amino Acid L S D L C N P D D K K V L L Q I K K A F G D P Y V L A S W K S D T

pGEM-*Eucpgip*110B ACTGTTGTGA TTGGTACTGC GTCACCTGTG ACTCAACCAC AAACCGCATC AACTCCCTCA CCATCTTTGC CGGCCAGGTA TCCGGCCAAA TCCCCGCCCT
Eucalyptus ACTGTTGTGA TTGGTACTGC GTCACCTGTG ACTCAACCAC AAACCGCATC AACTCCCTCA CCATCTTTGC CGGCCAGGTA TCCGGCCAAA TCCCCGCCCT
Pear ACTGCTGCGA TTGGTACTGC GTCACCTGTG ACTCCACCAC AAACCGCATT AACTCCCTCA CCATCTTTGC CGGCCAGGTG TCAGGCCAAA TCCCCGCCCT
Pear Amino Acid D C C D W Y C V T C D S T T N R I N S L T I F A G Q V S G Q I P A L

pGEM-*Eucpgip*110B AGTTGGAGAC TTGCCGTACC TTGAAACCCT TGAATTCCAT AAGCAACCCA ATCTCACTGG CCCAATCCAA CCCGCCATTG CCAAGCTCAA AGGACTCAAG
Eucalyptus AGTTGGAGAC TTGCCGTACC TTGAAACCCT TGAATTCCAT AAGCAACCCA ATCTCACTGG CCCAATCCAA CCCGCCATTG CCAAGCTCAA AGGACTCAAG
Pear AGTAGGAGAC TTGCCATACC TTGAAACCCT TGAATTCCAT AAGCAACCCA ATCTCACTGG CCCAATCCAA CCCGCCATTG CCAAGCTCAA AGGACTCAAG
Pear Amino Acid V G D L P Y L E T L E F H K Q P N L T G P I Q P A I A K L K G L K

pGEM-*Eucpgip*110B TCTCTCAGGC TCAGCTGGAC CAACCTCTCA GGCTCTGTCC CTGACTTCCT CAGCCAACTC AAGAACCTCA CATTCCTCGA CCTCTCCTTC AACAACCTCA
Eucalyptus TCTCTCAGGC TCAGCTGGAC CAACCTCTCA GGCTCTGTCC CTGACTTCCT CAGCCAACTC AAGAACCTCA CATTCCTCGA CCTCTCCTTC AACAACCTCA
Pear TCTCTCAGGC TCAGCTGGAC CAACCTCTCA GGCTCTGTCC CTGACTTCCT CAGCCAACTC AAGAACCTCA CATTCCTCGA CCTCTCCTTC AACAACCTCA
Pear Amino Acid S L R L S W T N L S G S V P D F L S Q L K N L T F L D L S F N N L

pGEM-*Eucpgip*110B CCGGTGCCAT CCCCAGCTCG CTTTCTGAGC TCCCAAACCT CGGCGCTCTT CGTCTAGACC GCAATAAGCT CACAGGTCAT ATTCCGATAT CGTTTGGGCA
Eucalyptus CCGGTGCCAT CCCCAGCTCG CTTTCTGAGC TCCCAAACCT CGGCGCTCTT CGTCTAGACC GCAATAAGCT CACAGGTCAT ATTCCGATAT CGTTTGGGCA
Pear CCGGTGCCAT CCCCAGCTCG CTTTCTGAGC TCCCAAACCT CGGCGCTCTT CGTCTAGACC GCAATAAGCT CACAGGTCAT ATTCCGATAT CGTTTGGGCA
Pear Amino Acid T G A I P S S L S E L P N L G A L R L D R N K L T G H I P I S F G Q

pGEM-*Eucpgip*110B GTTCATTGGC AACGTTCCAG ACCTGTATCT CTCCCACAAC CAGCTTTCTG GTAACATTCC AACCTCATTC GCTCAGATGG ACTTCACTAG CATAGACTTA
Eucalyptus GTTCATTGGC AACGTTCCAG ACCTGTATCT CTCCCACAAC CAGCTTTCTG GTAACATTCC AACCTCATTC GCTCAGATGG ACTTCACTAG CATAGACTTA
Pear GTTCATTGGC AACGTTCCAG ACCTGTATCT CTCCCACAAC CAGCTTTCTG GTAACATTCC AACCTCATTC GCTCAGATGG ACTTCACCAG CATAGACTTA
Pear Amino Acid F I G T V P D L Y L S H N Q L S G N I P T S F A Q M D F T S I D L

pGEM-*Eucpgip*110B TCACGGAACA AGCTCGAAGG TGACGCATCC GTGATATTTG GGCTGAACAA GACAACCCAG ATTGTGGACC TGTCCAGGAA CTTGCTGGAA TTTAATCTGT
Eucalyptus TCACGGAACA AGCTCGAAGG TGACGCATCC GTGATATTTG GGCTGAACAA GACAACCCAG ATTGTGGACC TGTCCAGGAA CTTGCTGGAA TTTAATCTGT
Pear TCACGGAACA AGCTCGAAGG TGACGCATCC GTGATATTTG GGCTGAACAA GACAACCCAG ATTGTGGACC TGTCCAGGAA CTTGCTGGAA TTTAATCTGT
Pear Amino Acid S R N K L E G D A S V I F G L N K T T Q I V D L S R N L L E F N L

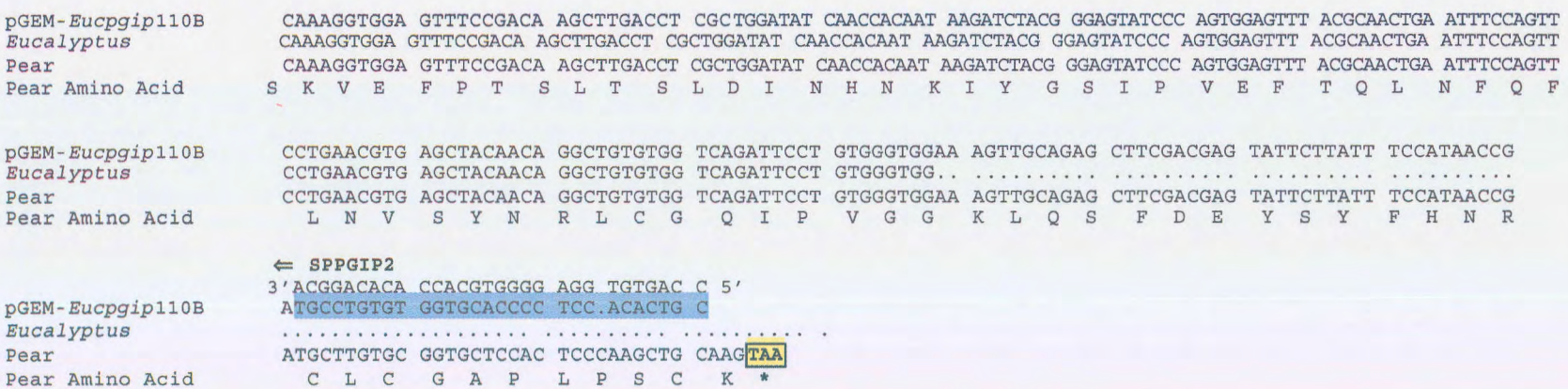

Figure 2.7 Nucleotide sequence alignments between amplified genomic DNA (*Eucalyptus*), the pGEM-*Eucpgip*110B clone and the published pear *pgip* sequence (Stotz *et al.*, 1993). The genomic amplified *pgip* gene (*Eucalyptus)* was subjected to direct sequencing using the gene-specific primers EN-FWD-PGIP and REV-PB-PGIP. T7 and SP6 primers enabled the sequencing of the entire cloned pgip fragment from the pGEM-*Eucpgip*110B clone. The regions used for synthesizing the PC6 and SPPGIP2 primers are indicated in blue. The pear PGIP stop codon (TAA) is indicated in yellow and is absent from the nucleotide sequence of the *Eucalyptus pgip* gene. Nucleotide differences between the amplified genomic DNA (*Eucalyptus*), the pGEM-*Eucpgip*110B clone and the published pear *pgip* sequence are highlighted in yellow.

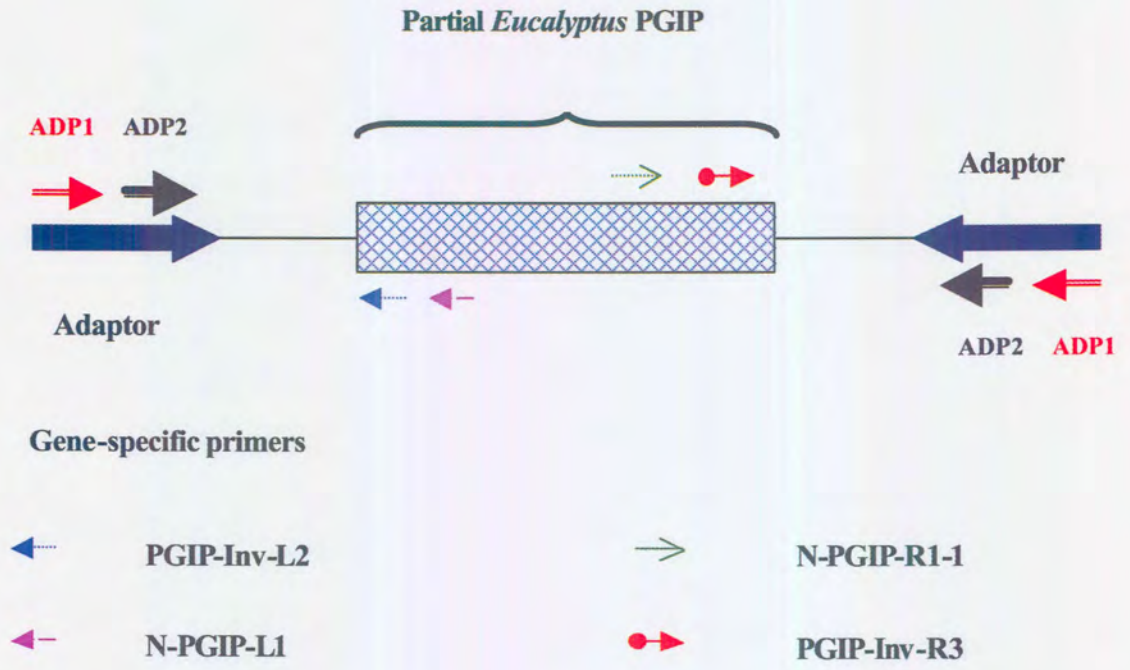

Figure 2.8 Schematic representation of the Genome Walking strategy used to clone the unknown 5' and 3' ends of the *Eucalyptus pgip* gene. Two sets of gene-specific primers were designed and used in conjunction with adaptor primers (ADP1 and ADP2) in the primary and secondary PCR walk reactions. In the primary PCR reactions, ADP1 and N-PGIP-L1 were used for walking upstream, while ADP1 and N-PGIP-R1-1 were used for walking towards the 3' region. The PCR products from the primary walk reactions were used in a secondary PCR, where ADP2 and PGIP-Inv-L2 were used for walking towards the 5' region and ADP2 and PGIP-Inv-R3, were used for the 3' region.

Isolation of the 5' and 3' ends of the *Eucalyptus pgip* gene by Genome Walking

The general procedure for genome walking has been illustrated in figure 1 and the genome walking cloning strategy for the *Eucalyptus pgip* gene is outlined in figure 8. *E. grandis* (EUCALYPTUS) genomic DNA was digested with a selection of blunt-end restriction enzymes. The restriction profile is represented in figure 2.9.

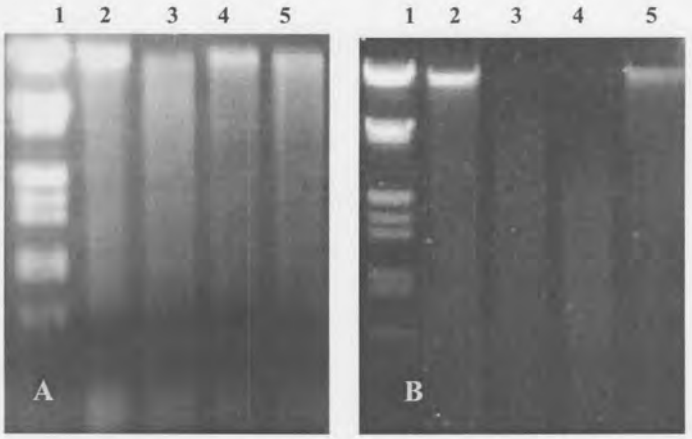

Figure 2.9 Restriction enzyme digest of TAG5 genomic DNA. **(A)** Lane 1 contains the λIII marker. Lane 2 represents undigested TAG5 genomic DNA and lanes 3, 4, and 5 contain genomic DNA digested with *Nco*I, *Pst*I and *Sca*I respectively. **(B)** Lane 1 contains the λIII marker and lane 2 contains the undigested genomic DNA. Lanes 3, 4 and 5 contain *Dra*I, *Hae*III and *Sma*I digested genomic DNA, respectively.

Restriction enzyme digestion with *Sca*I enabled the amplification of the 5' upsteam region, using adaptor primer (ADP2) and PGIP-Inv-L2, while digestion with *Dra*I lead to the amplification of the downstream 3' region when ADP2 and PGIP-Inv-R3 were used. Secondary PCR reactions were carried out at 62∘C, allowing the amplification of two distinct PCR fragments for the upstream region and a fragment of approximately 600bp, which represented the downstream region of the gene (Figure 2.10).

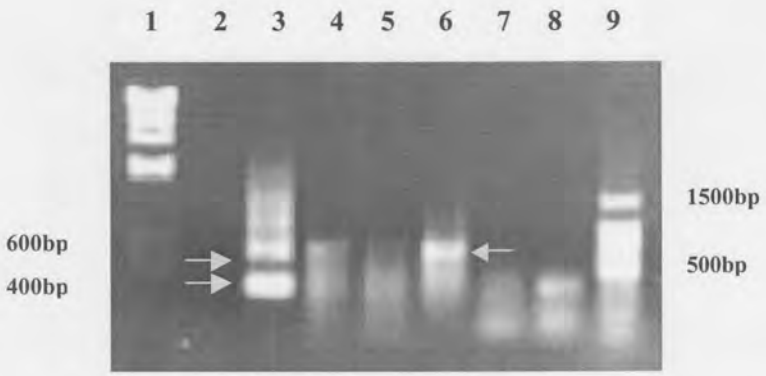

Figure 2.10 Genome Walking PCR analysis of genomic DNA using ADP2 and PGIP-Inv-L2 to walk upstream (Lanes 3, 7 and 8) and ADP2 and PGIP-Inv-R3 to walk downstream (Lanes 4,5 and 6). Lanes 3 and 4 represent amplification using genomic DNA digested with *Sca*I. In lanes 5 and 7 the template DNA was digested with *Hae*III and in lanes 6 and 7, *Dra*I digested DNA was used. Molecular weight marker λIII is shown in lane 1 and lane 9 contains a 100bp ladder (Promega). The arrows indicate PCR products that were cloned and sequenced.

Each fragment was individually cloned into the pCR®2.1 vector (TA®Cloning Kit, Invitrogen) and subjected to PCR screening and restriction enzyme analysis. Recombinant clones were digested with *Eco*RI to release the inserts, which are represented in figure 2.11. Recombinant plasmids containing the 700bp *Sca*I fragment were designated as pCR-*Sca*IGW-1 and pCR-*Sca*IGW-2, while plasmids containing the 400bp *Sca*I fragment were designated as pCR-*Sca*IGW-3 and pCR-*Sca*IGW-4. Plasmids pCR-*Dra*I-GW1 and pCR-*Dra*I-GW2 contain the 700bp downstream PCR fragment. The recombinant plasmids pCR-*Sca*IGW-3 and pCR-*Sca*IGW-4 show inserts that are approximately 280bp in length, indicating that the cloned PCR product possibly contains internal *Eco*RI sites. Restriction digestion of the pCR-*Sca*IGW-1 plasmid did not result in the release of the insert, while digestion of the pCR-*Sca*IGW-2 plasmid lead to the release of a 700bp insert.

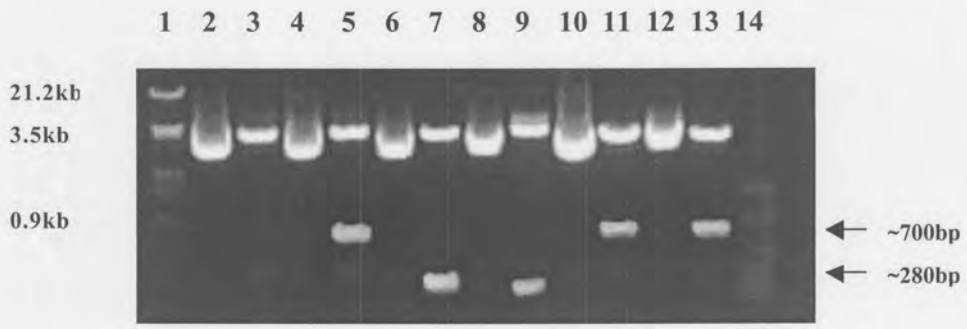

Figure 2.11 Restriction analysis of recombinant plasmids with *Eco*RI. Lanes 1 and 14 contain the molecular marker λIII and 100bp ladder (Promega) respectively. Lanes 2, 4, 6 and 8 contain the undigested plasmids pCR-*Sca*IGW-1 to pCR-*Sca*IGW-4 respectively, while lanes 3, 5, 7 and 9 contain plasmids pCR-*Sca*IGW-1 to pCR-*Sca*IGW-4 digested with *Eco*RI. Similarly, lanes 10 and 12 contain the undigested plasmids pCR-*Dra*I-GW1 and pCR-*Dra*I-GW2 respectively and lanes 11 and 13 represent *Eco*RI digested pCR-*Dra*I-GW1 and pCR-*Dra*I-GW2 respectively.

Each plasmid was used as a template in a PCR reaction with M13 forward and M13 reverse oligonucleotide primers. The PCR reaction was carried out at an annealing temperature of 64°C and the resulting PCR products are represented in figure 2.14. Inserts could not amplified from plasmids pCR-*Sca*IGW-1 and pCR-*Sca*IGW-2, however, a product of approximately 600bp could be amplified from plasmids pCR-*Sca*IGW-3 and pCR-*Sca*IGW-4. Recombinant plasmids pCR-*Dra*I-GW1 and pCR-*Dra*I-GW2 lead to the amplification of a fragment of approximately 900bp.

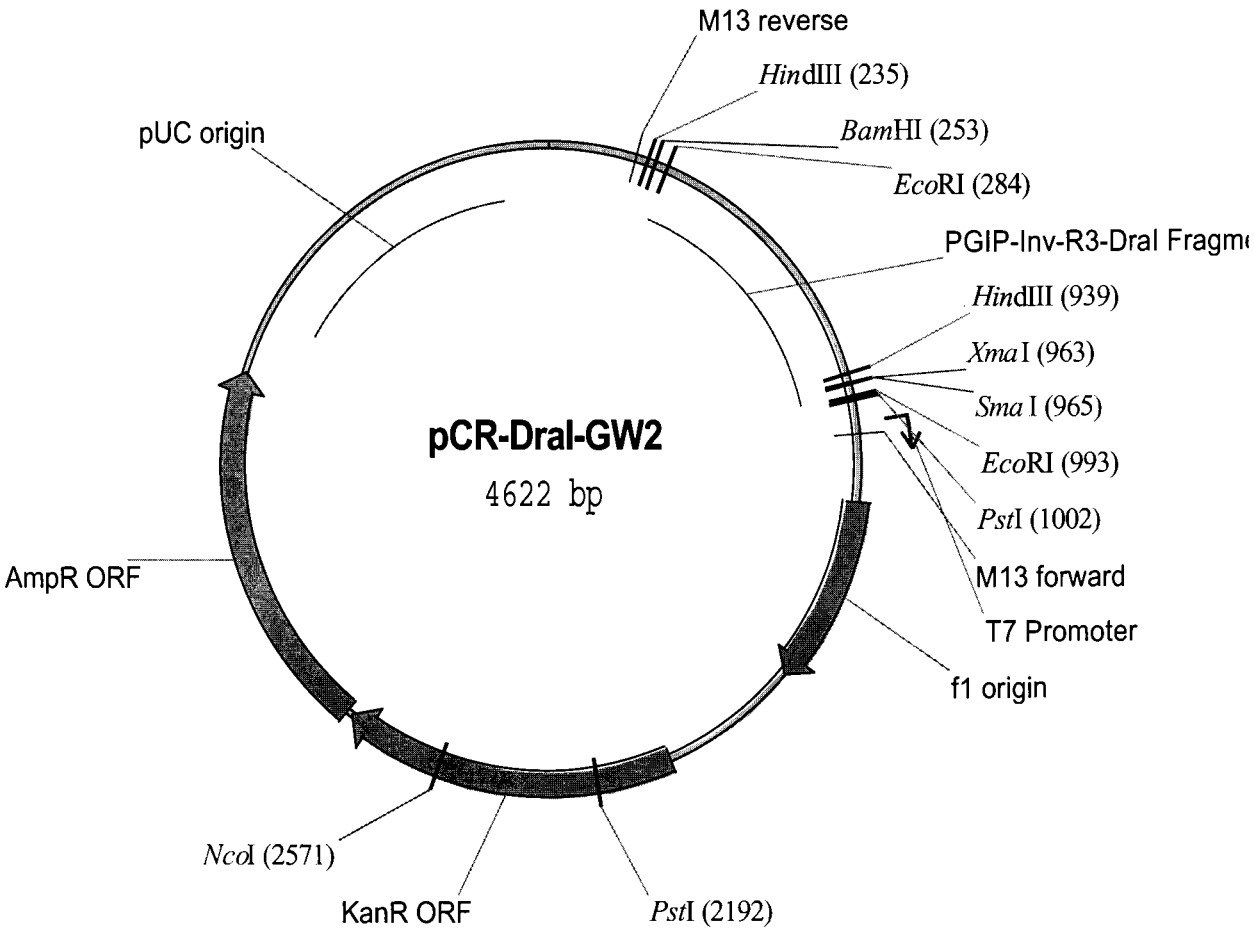

Figure 2.12 Plasmid map of the genome walking pCR-*Dra*I-GW2 construct containing the 3' downstream *Eucalyptus pgip* fragment.

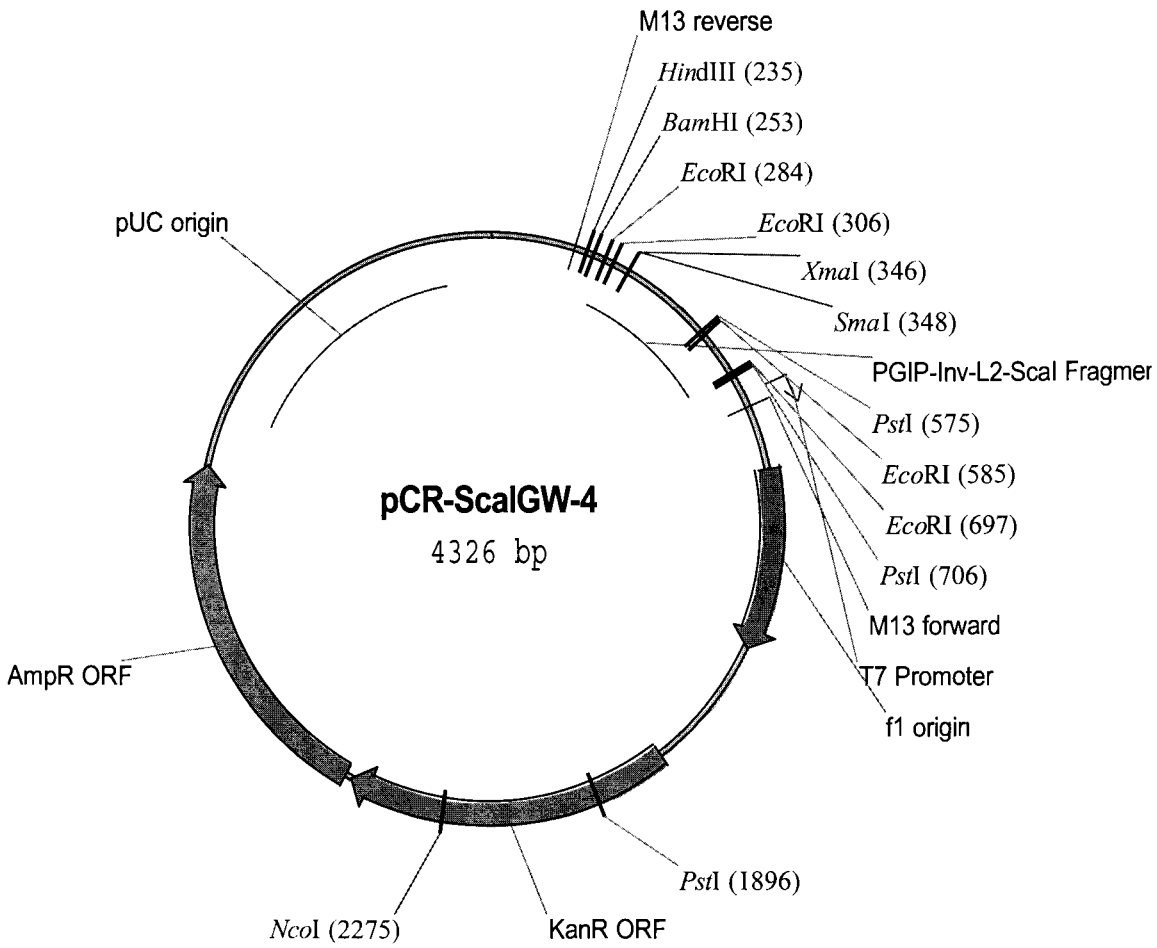

Figure 2.13 Plasmid map of the genome walking pCR-*ScaI*GW-4 construct containing the 5' upstream *Eucalyptus pgip* fragment.

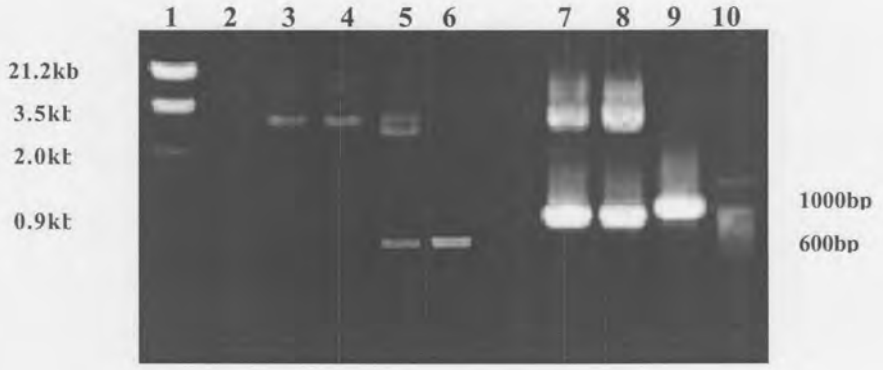

Figure 2.14 PCR analysis of genome walking plasmids using M13 forward and M13 reverse primers. Lanes 1 and 11 contain the molecular markers λIII and a 100bp DNA ladder (Promega) respectively. Lane 2 contains the negative water control and lane 9 contains a positive control that represents the amplification of the pgip fragment from the pGEM-*Eucpgip*110B clone with EN-FWD-PGIP and REV-PB-PGIP primers. Lanes 3 and 4 contain reactions in which pCR-*Sca*IGW-1 and pCR-*Sca*IGW-2 clones were used as templates, respectively. Lanes 5 represents amplification using pCR-*Sca*IGW-3 and lane 6 contains PCR products amplified from pCR-*Sca*IGW-4. Lanes 7 and 8 contain PCR products amplified from pCR-*Dra*I-GW1 and pCR-*Dra*I-GW2, respectively.

Clones pCR-*Sca*IGW-4, which represented the 5' upstream sequence and pCR-*Dra*I-GW2, which represents the 3' downstream sequence, were selected and subjected to nucleotide sequencing. The clones were sequenced in both directions using the M13 forward and M13 reverse primers. Plasmid maps of the pCR-*Dra*I-GW2 and pCR-*Sca*IGW-4 constructs are represented in figures 2.12 and 2.13, respectively.

Sequence Analysis of the complete *Eucalyptus pgip* gene

Based on inspection of the sequences to the upstream and downstream regions of the pGEM-*Eucpgip*110B clone it was concluded that the inserts from pCR-*Sca*IGW-4 and pCR-*Dra*I-GW2 represented the 5' and 3' regions of the *Eucalyptus pgip* gene respectively.

A 397bp fragment could be sequenced from the pCR-*Sca*IGW-4 clone using the M13 forward and M13 reverse primers (Figure 2.15). Alignment with the partial *Eucalyptus* sequence showed that the genome-walking clone shared an overlap of 91bp with the *Eucalyptus* sequence at the 5' region of the gene. Only one nucleotide difference, which results in a silent mutation could be observed within the signal sequence region. Further analysis of the additional 360bp of the pCR-*Sca*IGW-4 clone revealed that the upstream, untranslated region did not contain any transcription signals or promoter sequences.

The pCR-*Dra*I-GW2 clone contained a 693bp sequence of which only 39bp shared an overlap with the pGEM-*Eucpgip*110B clone (Figure 2.16). The overlap occurred in the downstream region of the partial eucalyptus *pgip* gene. The translation termination codon (TAG) could be observed in the nucleotide sequence of the genome-walking pCR-*Dra*I-GW2 clone. The 5' and 3' DNA sequences obtained from genome-walking PCR were linked to the partial *Eucalyptus pgip* sequence to give the composite *Eucalyptus* PGIP sequence. The complete nucleotide and deduced amino acid sequences of the composite *Eucalyptus pgip* gene are illustrated in figure 2.17.

```
                                                                ADP2 ⇒                ?
pCR-ScaI-GW4                                                    AATAGGGC TCGAGCGGCC GCCCGGGCAC TGAGCGCAAC GCAATTAATG  [80]
pGEM-Eucpgip110B      ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------

pCR-ScaI-GW4          TGAGTTAGCT CACTCATTAG GCACCCCAGG CTTTACACTT TATGCTTCCG GCTCGTATGT TGTGTGGAAT TGTGAGCGGA [160]
pGEM-Eucpgip110B      ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------

pCR-ScaI-GW4          TAACAATTTC ACACAGGAAA CAGCTATGAC CATGATTACG CCAAGCTATT TAGGTGACAC TATAGAATAC TCAAGCTATG [240]
pGEM-Eucpgip110B      ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------

pCR-ScaI-GW4          CATCCAACGC GTTGGGAGCT CTCCCATATG GTCGACCTGC AGGCGGCCGC GAATTCACTA GTGATTACTT CTCTCAGGCT [320]
pGEM-Eucpgip110B      ---------- ---------- ---------- ---------- ---------- ---------- ------ACAT CTCTCAGGCT [14]
                                                                                              ↑

pCR-ScaI-GW4          CTCAACCAAA ACCCAAAACA ATGGAGCTCA AGTTCTCCAC CTTCCTCTCC CTAACCCTAC TCTTCTCCTC CGTCCAA--- [397]
pGEM-Eucpgip110B      CTCAACCAAA ACCCAAAACA ATGGAGCTCA AGTTCTTCAC CTTCCTCTCC CTAACCCTAC TCTTCTCCTC CGTCCTAAAC [94]
                                                                                GGATG AGAAGAGGAG GCAGG
                                                                              ⇐ PGIP-Inv-L2

pCR-ScaI-GW4          ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
pGEM-Eucpgip110B      CCCGCTCTCT CCGATCTCTG CAACCCCGAC GACAAAAAAG TCCTCCTACA AATCAAGAAA GCCTTCGGCG ACCCCTATGT [174]

pCR-ScaI-GW4           ---------- ----------
pGEM-Eucpgip110B      CTTGGCCTCA TGGAAATGCA
```

Figure 2.15 Alignment of the nucleotide sequence of pCR-ScaI-GW4 with the sequence of the pGEM-*Eucpgip*110B clone. The genome-walking oligonucleotide primers used to generate the upstream sequence are highlighted in yellow. The blue arrow indicates the beginning of the *Eucalyptus pgip* insert in the pGEM-*Eucpgip*110B clone and the start (ATG) codon is indicated in blue. The sequences overlap by 91bp in the region surrounding the ATG codon. The partial *Sca*I site is indicated by the ? symbol. The sequence indicated in pink is derived from the adaptor.

Figure 2.16 Alignment of the nucleotide sequence of pCR-*Dra*I-GW2 with the downstream sequence of the pGEM-*Eucpgip*110B clone. Oligonucleotide primers PGIP-Inv-R3 and ADP2 are highlighted in yellow. The stop (TAG) codon is indicated in blue and by the asterisk (*). The pCR-*Dra*I-GW2 clone shows only 39bp overlap with the sequence of the pGEM-*Eucpgip*110B clone. The partial DraI site is indicated by the ? symbol. SPPGIP2 primer, which shares an overlap of 10bp with the PGIP-Inv-R3 primer, is indicated in green. The sequence indicated in pink is derived from the adaptor.

⇐ PGIP-Inv-L2

```
1    Atg GAG CTC AAG TtC TCC ACC TTC CTC TCC CTA ACC CTA CTC TTC  45
1     M   E   L   K   F   S   T   F   L   S   L   T   L   L   F  15

46   TCC TCC GTC CTA AAC CCC GCT CTC TCC GAT CTC TGC AAC CCC GAC  90
16    S   S   V   L   N   P   A   L   S   D   L   C   N   P   D  30

91   GAC AAA AAA GTC CTC CTA CAA ATC AAG AAA GCC TTC GGC GAC CCC  135
31    D   K   K   V   L   L   Q   I   K   K   A   F   G   D   P  45

136  TAT GTC TTG GCC TCA TGG AAA TCA GAC ACC GAC TGT TGT GAT TGG  180
46    Y   V   L   A   S   W   K   S   D   T   D   C   C   D   W  60

181  TAC TGC GTC ACC TGT GAC TCA ACC ACA AAC CGC ATC AAC TCC CTC  225
61    Y   C   V   T   C   D   S   T   T   N   R   I   N   S   L  75

226  ACC ATC TTT GCC GGC CAG GTA TCC GGC CAA ATC CCC GCC CTA GTT  270
76    T   I   F   A   G   Q   V   S   G   Q   I   P   A   L   V  90

271  GGA GAC TTG CCG TAC CTT GAA ACC CTT GAA TTC CAT AAG CAA CCC  315
91    G   D   L   P   Y   L   E   T   L   E   F   H   K   Q   P  105

316  AAT CTC ACT GGC CCA ATC CAA CCC GCC ATT GCC AAG CTC AAA GGA  360
106   N   L   T   G   P   I   Q   P   A   I   A   K   L   K   G  120

361  CTC AAG TCT CTC AGG CTC AGC TGG ACC AAC CTC TCA GGC TCT GTC  405
121   L   K   S   L   R   L   S   W   T   N   L   S   G   S   V  135

406  CCT GAC TTC CTC AGC CAA CTC AAG AAC CTC ACA TTC CTC GAC CTC  450
136   P   D   F   L   S   Q   L   K   N   L   T   F   L   D   L  150

451  TCC TTC AAC AAC CTC ACC GGT GCC ATC CCC AGC TCG CTT TCT GAG  495
151   S   F   N   N   L   T   G   A   I   P   S   S   L   S   E  165

496  CTC CCA AAC CTC GGC GCT CTT CGT CTA GAC CGC AAT AAG CTC ACA  540
166   L   P   N   L   G   A   L   R   L   D   R   N   K   L   T  180
```

⇐ N-PGIP-L1

```
541   GGT CAT ATT CCG ATA TCG TTT GGG CAG TTC ATT GGC AAC GTT CCA   585
181    G   H   I   P   I   S   F   G   Q   F   I   G   N   V   P   195

586   GAC CTG TAT CTC TCC CAC AAC CAG CTT TCT GGT AAC ATT CCA ACC   630
196    D   L   Y   L   S   H   N   Q   L   S   G   N   I   P   T   210

631   TCA TTC GCT CAG ATG GAC TTC ACT AGC ATA GAC TTA TCA CGG AAC   675
211    S   F   A   Q   M   D   F   T   S   I   D   L   S   R   N   225

676   AAG CTC GAA GGT GAC GCA TCC GTG ATA TTT GGG CTG AAC AAG ACA   720
226    K   L   E   G   D   A   S   V   I   F   G   L   N   K   T   240

721   ACC CAG ATT GTG GAC CTG TCC AGG AAC TTG CTG GAA TTT AAT CTG   765
241    T   Q   I   V   D   L   S   R   N   L   L   E   F   N   L   255

766   TCA AAG GTG GAG TTT CCG ACA AGC TTG ACC TCG CTG GAT ATC AaC   810
256    S   K   V   E   F   P   T   S   L   T   S   L   D   I   N   270

811   CAC AAT AAG ATC TAC GGG AGT ATC CCA GTG GAG TTT ACG CAA CTG   855
271    H   N   K   I   Y   G   S   I   P   V   E   F   T   Q   L   285

                                N-PGIP-R1-1 ⇒
856   AAT TTC CAG TTC CTG AAC GTG AGC TAC AAC AGG CTG TGT GGT CAG   900
286    N   F   Q   F   L   N   V   S   Y   N   R   L   C   G   Q   300

901   ATT CCT GTG GGT GGA AAG TTG CAG AGC TTC GAC GAG TAT TCT TAT   945
301    I   P   V   G   G   K   L   Q   S   F   D   E   Y   S   Y   315

                      PGIP-Inv-R3 ⇒
946   TTC CAT AAC CGA TGC CTG TGT GGT CCT CCC CTC GGT CCC TGC AAG   990
316    F   H   N   R   C   L   C   G   A   P   L   G   P   C   K   330

991   AGT TAG   996
331    S   *
```

Figure 2.17 Nucleotide and predicted amino acid sequence of the complete composite *Eucalyptus pgip* gene. The red arrow indicates the initiation codon and the asterisk indicates the termination codon (TAG). The open arrow indicates the end of the putative signal peptide and the predicted start of the mature PGIP protein. The cysteine residues are

indicated in blue and the poteintial N-glycosylation sites are highlighted in green. The regions used for synthesizing genome walking PCR primers are presented in blue coloured blocks.

Discussion

Polygalacturonase-inhibiting Proteins (PGIPs), associated with the cell wall of many plant species form part of the plants active defense against phytopathogenic fungi. Genes encoding PGIPs have been isolated from various plant species that include, bean, soybean, tomato, pear, apple, raspberry and recently, guava (Toubart *et al.,* 1992; Stotz *et al.,* 1993, 1994; Favaron *et al.,* 1994; Yao *et al.,* 1999; Ramanathan *et al.,* 1997; Deo and Shastri, 2003). In this chapter, the isolation and molecular characterisation of a full-length *pgip* gene from *Eucalyptus grandis* using conventional PCR and genome-walking PCR techniques is reported.

The reported composite *Eucalyptus pgip* gene is comprised of a single open reading frame of 996bp that is uninterrupted by introns. It encodes a protein of 331 amino acids with a predicted molecular mass of 36.5 kDa. A theoretical isoelectric point of 6.2 was calculated (Table 2.3). A comparison with published sequences of pear, apple, bean and soybean PGIPs, revealed the conservation of the potential processing site for signal peptidase (Toubart *et al.,* 1992) (Figure 2.18). The N-terminal highly hydrophobic region of the composite *Eucalyptus pgip* gene corresponds to the putative signal peptide, which is identical to that of the pear PGIP signal peptide sequence. The signal peptide has been proposed to have a role in targeting the PGIP through the endomembrane system (Stotz *et al.,* 1993).

The complete polypeptide sequence differs from the published pear polypeptide sequence by only three amino acids at the C-terminal region. When compared to the partial *Eucalyptus* polypeptide sequence, however, there is an observed difference of four amino acids (GKHR) within the peptide sequence (Figure 2.18). The *Eucalyptus* sequence obtained in this study was TSID, which is identical to that found in both pear and apple

PGIP1 (Figure 2.18). Several reasons can be attributed to the observed difference in amino acid sequence between the partial *Eucalyptus* PGIP and the complete *Eucalyptus* PGIP obtained in this study. The most obvious reason could be the use of different *E. grandis* clones in each study, which could possibly result in a difference in the polypeptide sequence. Alternatively, the differences observed could be attributed to the sequencing of a different PGIP allele or PGIP gene family member as compared to the one obtained in this study. The type of polymerase used to amplify the *pgip* gene also plays a significant role as to whether any mutations are incorporated in the gene sequence. In the current study, the DNA sequence printouts were analysed carefully in this region. Furthermore, the sequence of this region was confirmed in this study by direct sequencing from genomic *E. grandis* (TAG5) DNA (Figure 2.7). A possibility is that in the previous study by Chimwamurombe *et al.* (2001), 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.* (2001) obtained the same (GKHR) polypeptide region for each of the five *Eucalyptus* PGIP clones sequenced, thus strengthening the hypothesis that compared to this study, an alternative PGIP allele or PGIP gene family member was previously obtained.

The complete *Eucalyptus* PGIP polypeptide sequence shows relatively high sequence similarity to pear, apple, and tomato PGIPs, and in comparison to bean and soybean PGIPs (Table 2.4). It is interesting to note that the PGIP from *Eucalyptus*, which belongs to the family Myrtaceae, shows a high degree of sequence similarity to pear and apple PGIPs, which both reside in the Rosaceae. These families are phylogenetically rather distant from each other.

According to Chimwamurombe *et al.* (2001), the five *Eucalyptus* species studied each have more than one *pgip* gene. For each species only one of these *pgip* genes were partially characterized. Phylogenetic analysis of PGIPs shows that those from the Myrtaceae group together with those from the Rosaceae (Chimwamurombe *et al.*, 2001). A possible explanation for such a grouping could be that the partially characterized *Eucalyptus pgip* gene included in the phylogenetic analysis was non-homologous to the *pgip* genes from the

other plant species. The high similarity of PGIPs between the distantly related Myrtaceae and Rosaceae could also imply a possible divergence of these PGIPs from a common ancestral PGIP. However, it is tenuous to infer phylogenies based on single gene sequences since multigene analyses that include chloroplast and nuclear genes for example, is usually required before true relatedness of plants can be inferred.

The *Eucalyptus* PGIP contains nine cysteine residues, eight of which are conserved between the apple, pear, tomato, bean and soybean PGIPs. The cysteine residues are clustered at the N- and C-terminal regions of the mature protein and their location, therefore, indicates a potential role in stabilising the secondary structure of the protein (Stotz *et al.*, 1993, Favaron *et al.*, 1994). As in all other characterised PGIPs, the proposed *Eucalyptus* PGIP structure contains seven potential N-glycosylation sites (Asn-X-Ser/Thr) that are shared with the apple and pear PGIPs. Five of the sites are shared with the tomato PGIP, two with bean PGIP and only one site is shared with soybean PGIP.

Table 2.3 Predicted molecular mass and isoelectric point values of some cloned *pgip* genes.

Species	kDa	pI Value	Reference
Phaseolus vulgaris	34	8.81	Toubart *et al.* (1992); Leckie *et al.* (1999)
Pyrus communis	36.5	6.2	Stotz *et al.* (1993)
Glycine max	34	8.21	Favaron *et al.* (1994)
Lycopersicon esculentum	34.3	8.4	Stotz *et al.* (1994)
Eucalyptus grandis	36.5*	6.2*	This study

* http://www.expasy.ch

Table 2.4 Nucleotide and amino acid sequence comparison of the complete *Eucalyptus* PGIP with other published PGIP polypeptide sequences.

Plant Species	Accession No.	%DNA identity	%AA identity	%AA similarity
Soybean	X78274	59	46	60
Bean	X64769	59	49	66
Tomato	L26529	77	66	78
Partial Eucalyptus	AF159167	95	95	96
Apple PGIP1	U77041	97	93	66
Pear	L09264	98	99	99

```
                          10         20         30         40         50         60         70         80         90        100
Pear                .....MELKF STFLSLTLLF SSVLNPALSD LCNPDDKKVL LQIKKAFGDP YVLASWKSDT DCCD..WYCV TCDSTTN..R INSLTIFAGQ VSG..QIPAL [89]
Composite_Eucalyptus.....MELKF STFLSLTLLF SSVLNPALSD LCNPDDKKVL LQIKKAFGDP YVLASWKSDT DCCD..WYCV TCDSTTN..R INSLTIFAGQ VSG..QIPAL [89]
Partial_Eucalyptus                                    VL LQIKKAFGDP YVLASWKSDT DCCD..WYCV TCDSTTN..R INSLTIFAGQ VSG..QIPAL [64]
Apple_PGIP1         .....MELKF SIFLSLTLLF SSVLKPALSD LCNPDDKKVL LQIKKAFGDP YVLTSWKSDT DCCD..WYCV TCDSTTN..R INSLTIFAGQ VSG..QIPAL [89]
Tomato              .........M NLSLLLVVIF LCFASPSLSV RCNPKDKKVL LQIKKDLGNP YHLASWDPNT DCCY..WYVI KCDRKTN..R INALTVFQAN ISG..QIPAA [85]
Bean               MTQFNIPVTM SSSLSIILVI LVSLRTALSE LCNPQDKQAL LQIKKDLGNP TTLSSWLPTT DCCNRTWLGV LCDTDTQTYR VNNLDLSGHN LPKPYPIPSS 100]
Soybean             .......... .......... .........E LCNPQDKQTL LQIKKELGNP TTLSSWHPKT DCCNNSWVGV SCDTVTPTYR VDNLDLSELN LRKPYPIPPS [71]
                                                    ↑

                         110        120        130        140        150        160        170        180        190        200
Pear               VGDLPYLETL EFHKQPNLTG PIQPAIAKLK GLKSLRLSWT NLSGSVPDFL SQLKNLTFLD LSFNNLTGAI PSSLSELPNL GALRLDRNKL TGHIPISFGQ [189]
Composite_Eucalyptus VGDLPYLETL EFHKQPNLTG PIQPAIAKLK GLKSLRLSWT NLSGSVPDFL SQLKNLTFLD LSFNNLTGAI PSSLSELPNL GALRLDRNKL TGHIPISFGQ [189]
Partial_Eucalyptus VGDLPYLETL EFHKQPNLTG PIQPAIAKLK GLKFLRLSWT NLSGSVPDFL SQLKNLTFLD LSFNNLTGAI PSSLSQLPNL NALHLDRNKL TGHIPKSFGQ [164]
Apple_PGIP1        VGDLPYLETL EFHKQPNLTG PIQPAIAKLK GLKFLRLSWT NLSGSVPDFL SQLKNLTFLD LSFNNLTGAI PSSLSQLPNL NALHLDRNKL TGHIPKSLGQ [189]
Tomato             VGDLPYLETL EFHHVTNLTG TIPPAIAKLT NLKMLRLSFT NLTGPIPEFL SQLKNLTLLE LNYNQFTGTI PSSLSQLPNL LAMYLDRNKL TGTIPESFGR [185]
Bean               LANLPYLNFL YIGGINNLVG PIPPAIAKLT QLHYLYITHT NVSGAIPDFL SQIKTLVTLD FSYNALSGTL PPSISSLPNL GGITFDGNRI SGAIPDSYGS [200]
Soybean            VGSLPCLKFL YITNNPNIVG TIPTTITKLT KLRELNIRYT NISGQIPHFL SQIKALGFLD LSNNKLSGNL PSWLPSLPDL YGISFDNNYI SGPIPDLFAS [171]

                         210        220        230        240        250        260        270        280        290        300
Pear               FIG.NVPDLY LSHNQLSGNI PTSFAQMDFT SIDLSRNKLE GDASVIFGLN KTTQIVDLSR NLLEFNLSKV EFPTSLTSLD INHNKIYGSI PVEFTQLN.F [287]
Composite_Eucalyptus FIG.NVPDLY LSHNQLSGNI PTSFAQMDFT SIDLSRNKLE GDASVIFGLN KTTQIVDLSR NLLEFNLSKV EFPTSLTSLD INHNKIYGSI PVEFTQLN.F [287]
Partial_Eucalyptus FIG.NVPDLY LSHNQLSGNI PTSFAQMDFG KHRLSRNKLE .DASVIFGLN KTTQIVDLSR NLLEFNLSKV EFPTSLTSLD VNHNKIYGSI PVEFTQLN.F [260]
Apple_PGIP1        FIG.NVPDLY LSHNQLSGNI PTSFAQMDFT SIDLSRNKLE GDASVIFGLN KTTQIVDLSR NLLEFNLSKV EFPTSLTSLD INHNKIYGSI PVEFTQLN.F [287]
Tomato             FKGPNIPDLY LSHNSLTGHV PASLGDLNFS TLDFSRNKLE GDVSFLFGKN KTSQVIDLSR NLLEFDISKS EFAESLISLD LNHNRIFGSL PPGLKDVP.L [284]
Bean               FSK.LFTAMT ISRNRLTGKI PPTFANLNLA FVDLSRNMLE GDASVLFGSD KNTKKIHLAK NSLAFDLGKV GLSKNLNGLD LRNNRIYGTL PQGLTQLKFL [299]
Soybean          VSE.RFGFIS LSGNRLIGKI PASLGKPDMK IVDLSRNMLE GDASVLFGSE KHTERIYLAN NLFAFDLGKV RLSKTLGLLD VGHNLIYGTL PKGLTSLKDL [270]

                         310        320        330        340
Pear               QFLNVSYNRL CGQIPVGGKL QSFDEYSYFH NRCLCGAPLP SCK. [330]
Composite_Eucalyptus QFLNVSYNRL CGQIPVGGKL QSFDEYSYFH NRCLCGAPLG PCKS [331]
Partial_Eucalyptus QFLNVSYNRL CGQIPVGGKL QSFNEYSYFH NR.......... [301]
Apple_PGIP1        QFLNVSYNRL CGQIPVGGKL QSFDEYSYFH NRCLCGAPLP SCK. [330]
Tomato             QFFNVSYNRL CGQIPQGGTL QSFDIYSYLH NKCLCGSPLP KCK. [327]
Bean               QSLNVSFNNL CGEIPQGGNL KRFDVSSYAN NKCLCGSPLP SCT. [342]
Soybean            YYLDVSYNNL CGEIPRGGKL QEFDASLYAN NKCLCGSPLP SCT. [313]
```

Figure 2.18 Comparison of the deduced amino acid sequences of PGIPs from apple, pear, eucalyptus, tomato, bean and soybean (Yao *et al.*, 1999; Stotz *et al.*, 1993; Chimwamurombe *et al.*, 2001; Stotz *et al.*, 1994; Toubart *et al.*, 1992; Favaron *et al.*, 1994). The arrow indicates the N-terminal position for each of the mature peptides. The cysteine residues are highlighted in blue and the potential N-glycosylation sites are indicated in green. Amino acids that differ between the pear PGIP and the partial *Eucalyptus* PGIP are highlighted in green. Amino acids that differ between the composite *Eucalyptus* PGIP and the pear PGIP are highlighted in yellow.

Like all other PGIPs isolated to date, *Eucalyptus* PGIP is almost entirely composed of 10 imperfect Leucine-rich repeats (LRRs) characterised by the consensus sequence LxxLxxLxxLxLxxNxLxGxIPxx (Jones and Jones, 1997) (Figure 2.19). LRR motifs that are similar to those found in PGIPs, occur in both bacterial and eukaryotic proteins (De Lorenzo *et al.,* 1997; Marino *et al.,* 2000). In eukaryotes, LRR motifs are found in proteins encoding the disease resistance (*R*) genes of the plant immune system and by the *toll* and *toll*-like genes of *Drosophila* and mammals (Marino *et al.,* 2000). The consensus sequence of the repeating units is strikingly similar in all described LRR proteins, indicating a selection pressure for the conservation of this structure (De Lorenzo and Cervone, 1997). Many of these proteins are part of the cell wall receptors that respond to elicitors and are thus involved in protein-protein interactions and signal transduction pathways (Kobe and Deisenhofer, 1994; Ramanathan *et al.,* 1997).

Most plant resistance (*R*) genes encode leucine-rich repeat (LRR) proteins, which function as putative receptors that are able to trigger defence responses upon recognition of a pathogen (Jones and Jones, 1997). Similarly, PGIPs, which recognise extracellular polygalacturonases secreted by pathogenic fungi, contain LRR domains that are homologous to proteins encoded by plant disease (*R*) resistance genes that include tomato Cf-9, Cf-4 and Cf-4, and the rice Xa21 gene (Jones and Jones, 1997; Komjanc *et al.,* 1999). Within each LRR domain is a specific sequence xxLxLxx occurs, which is predicted to form a solvent exposed β-sheet/β-turn structure and is directly involved in ligand interaction (Leckie *et al.,* 1999; De Lorenzo and Ferrari, 2002). The amino acids of PGIP that determine specificity and affinity for fungal PGs reside within the β-sheet/β-turn structure. Thus, information on the structural and functional basis of the interactions between PGIPs and PGs can be obtained from manipulations of LRR receptor proteins. This information can ultimately lead to a better understanding of how plants recognise non-self molecules from pathogenic microorganisms.

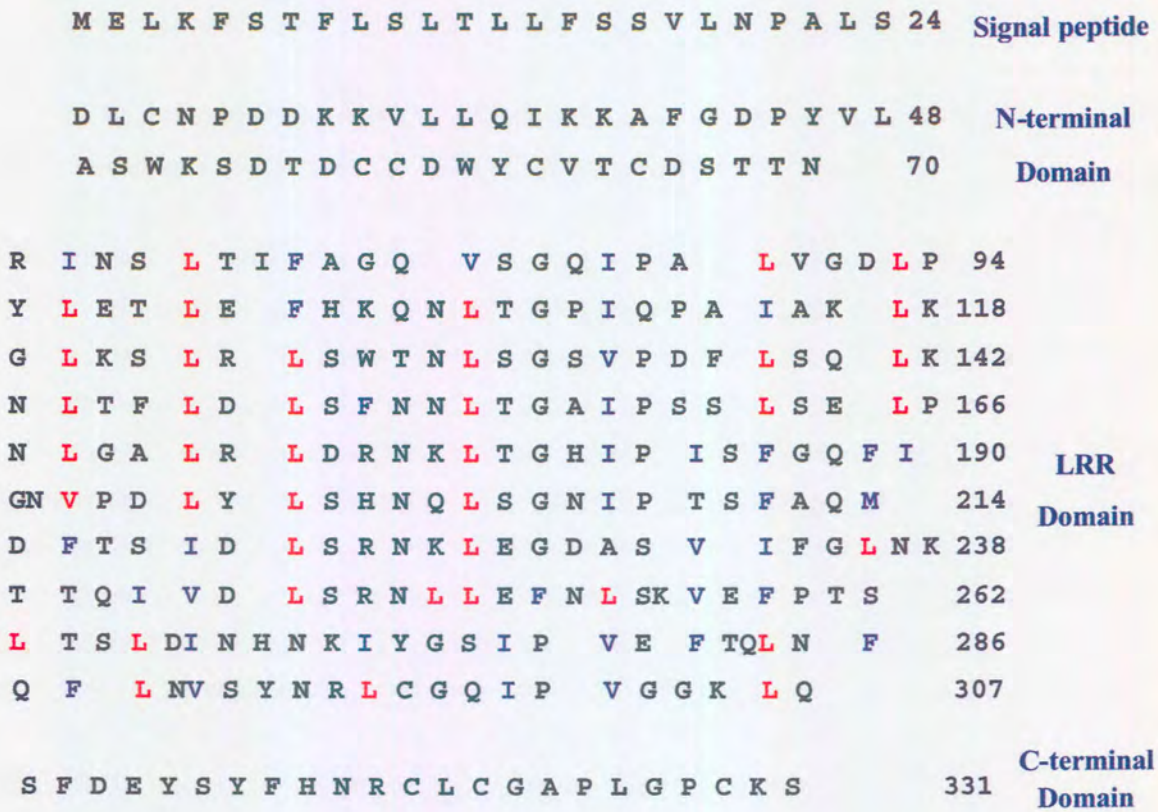

Figure 2.19 Predicted domain structure of the complete *Euclayptus* PGIP. The amino acid sequence is shown in single-letter code. Leucine residues are indicated in red. Changes in the consensus sequence, where L and I tend to be substituted by each other and by M, V and F are indicated in blue.

This chapter reports on the cloning and characterisation of a *Eucalyptus pgip* gene. The full-length *Eucalyptus pgip* gene is uninterrupted by intron sequences and shows 99% amino acid identity to pear PGIP (Stotz *et al.*, 1993). The high sequence homology to pear PGIP could imply that they have evolved from the same ancestral PGIP, however such conclusions can only be derived from comparative analysis of all PGIP copies from both plant species.