CHAPTER 1
Aim of study

Verticillium-wilt is a major disease limiting potato (*Solanum tuberosum* L.) yield under arid growing conditions. It is caused by the soil-borne fungal pathogen, *Verticillium dahliae*. Current control approaches include sanitation of fields and equipment, crop rotation, soil fumigation and chemical control. Since current control strategies are not effective in managing the disease, and chemical treatments may be deleterious to the environment, alternative strategies are required. Breeding for resistant potato lines faces many challenges, therefore the transgenic manipulation to confer resistance against this disease is proposed.

The literature review will reveal the importance of polygalacturonase inhibiting protein (PGIP), a plant defence gene, in resistance against fungal pathogens. It has been shown previously to be successful in conferring fungal resistance to a heterologous plant. Pear PGIP was able to confer resistance to tomato *in vivo* against the fungal pathogen *Botrytis cinerea*, whereas endogenous tomato PGIP was not effective (Powell *et al.*, 2000). Apple and pear PGIP was reported to be 97% identical in nucleotide sequence, and it was thus anticipated that apple PGIP would also be a successful inhibitor of *B. cinerea* infection. A crude total PGIP extract from apple, *Malus domestica* cv. Granny Smith, was indeed able to inhibit *B. cinerea* PGs (D.K. Berger, D. Oelofse and M. Arendse, personal communication). An apple *pgip1* gene was isolated using inverse polymerase chain reaction methods and cloned at ARC-Roodeplaat (Arendse *et al.*, 1999). A project was thus proposed to transform tomato with apple *pgip1* to confer resistance against *B. cinerea*. This fungal pathogen causes substantial losses of the crop during cold storage. Subsequent transformation of the apple *pgip1* gene into tobacco cv. LA Burley and transgenic expression of the protein showed it to be an active inhibitor of other fungal PGs, but not PGs from *B. cinerea*. It is thus possible that apple *pgip1* is not the gene coding for the PGIP in the crude apple extract that is responsible for the inhibitory activity against *B. cinerea* PGs. It has been shown that only one amino acid substitution can change the PG:PGIP interaction specificity (Leckie *et al.*, 1999). Apple contains more than one *pgip* gene (Yao *et al.*, 1999), so one of the other genes could code for the PGIP responsible for *B. cinerea* PG inhibition.

Apple PGIP1 expressed in transgenic tobacco was, however, successful in inhibiting PGs from *V. dahliae*, the fungus causing *Verticillium*-wilt of potato (see Chapter 6 for inhibition results). This protein may therefore be useful in the genetic engineering of potato for resistance against this fungal pathogen. The focus of this MSc study therefore changed from evaluating apple PGIP1 in transgenic tomato against *B. cinerea*, to evaluating apple PGIP1 in potato vs. *V. dahliae*. Transgenic potato cv.

BP1 plants, containing the apple *pgip1* gene under control of the constitutive enhanced CaMV 35S promoter, had been generated subsequently by A. Veale (ARC Roodeplaat) and maintained in tissue culture.

The aim of this study was the molecular characterisation of the transgenic potato lines containing the apple *pgip1* gene, and to assess whether it confers enhanced fungal resistance to *V. dahliae* during a glasshouse trial. In addition, the expression of resistance genes under control of a pathogen inducible plant promoter, the *gst1* promoter from *Arabidopsis thaliana* (L.) Heynh (Yang *et al.*, 1998), was evaluated by creating transgenic *A. thaliana* plants containing apple *pgip1* under control of this promoter.

The hypothesis is that the apple PGIP1 will be effective against pectin-degrading enzymes from the potato pathogen *V. dahliae*, and will confer enhanced resistance to the pathogen under glasshouse conditions. It is also hypothesised that the *gst1* promoter will drive pathogen-inducible expression of the apple *pgip1* gene in transgenic *A. thaliana*.

The literature review in Chapter 2 focuses on the role of polygalacturonase-inhibiting proteins in the defence against phytopathogenic fungi. It also reviews the specific interaction between a plant and a fungus, namely *Verticillium*-wilt of potato. Chapter 3 describes the cloning of the apple *pgip1* gene under control of a pathogen-inducible promoter, whereas Chapter 4 describes the transformation of *A. thaliana* with this construct and molecular analysis of the transgenics. Chapters 5 to 7 will deal with the main aim of this project, which is to evaluate whether PGIP confers enhanced resistance against *V. dahliae* to potato. The molecular characterisation of apple *pgip1* transgenic potato plants is covered in Chapter 5, and the inhibition of *V. dahliae* PGs by transgenically expressed apple PGIP1 in Chapter 6. Chapter 7 reports on the glasshouse trial for screening transgenic potato lines for increased resistance to *V. dahliae* infection. Chapter 8 is the concluding discussion of the relevance of the results obtained.

CHAPTER 2

Literature review

2.1 Introduction

This literature review will describe the disease caused by the specific interaction between potato and *V. dahliae*, namely *Verticillium*-wilt. It will continue to briefly introduce the principle of resistance of plants to fungal pathogens. The topics of fungal polygalacturonases as pathogenicity factors, and plants' mechanism to defend against fungal infection, polygalacturonase-inhibiting proteins, are reviewed. Some background information on the pathogen-inducible promoter that will be used in this study, the *gst1* promoter from *Arabidopsis thaliana* (L.) Heynh, is given. Lastly, it will highlight the importance of *A. thaliana* as a model plant for research.

2.2 Potato as an economically important crop

Potato (*Solanum tuberosum* L.) is an important component in intensive agricultural systems in areas of high population density, since it can produce a large mass of high-value food in a short time (Hijmans, 2001). It is an important staple food and animal feed especially in the northern hemisphere. A study of global potato production over the period from 1950 to 1999 has shown a stable production of 30 million tonnes per year. This is due to overall increasing yields, despite the area grown with potato decreasing stably at 92000 ha per year. On a global scale, 51% of potato area is in Europe, 35% in Asia and the rest equally distributed between North America, South America and Africa. South Africa is the second largest potato producer in Africa, with 61000 ha potato area, whereas Egypt has 82561 ha. Ten thousand hectares of the total potato production area in South Africa are for seed potato production. Seed potatoes are certified disease free by a seed certification scheme administered by Potatoes South Africa. These are then supplied to table potato producers that produce potatoes for the consumer market (Berger, 2000). The total value of the potato harvest in South Africa is in excess of two billion rand (Potatoes South Africa, 2001 potato harvest figures).

2.3 *Verticillium*-wilt of potato

Verticillium-wilt is a major disease limiting potato yield under arid growing conditions (Corsini and Pavek, 1996). The main causal organism is *V. dahliae* and to a lesser extent *V. nigrescens* (Millard and Denner, 2001). *V. dahliae* belongs to the Kingdom Mycetae, Division Eumycota, subdivision

Deuteromycotina, class Hyphomycetes, and genus *Verticillium*. *V. dahliae* is a destructive soil-borne fungal pathogen that causes vascular wilt diseases on more than 160 plant species, which includes cotton, tomatoes and potatoes. The fungus penetrates the host through its roots, spreads systemically through the xylem, and subsequently leads to the appearance of wilt symptoms (James and Dubery, 2001).

2.3.1 Source of fungal inoculum

Infected seed potatoes serve as the inoculum source, although it can also occur on roots of natural plant growth. It has a wide host range, so the fungus can survive at low levels on roots of many crop and weed species (Rowe, 1985). The fungus can survive up to 10 years in soil as dormant microsclerotia. As soon as the amount of microsclerotia in the soil exceeds a certain threshold, the disease develops (Millard and Denner, 2001).

2.3.2 Symptoms of *Verticillium*-wilt

Visually there is no difference between *Fusarium*-wilt and *Verticillium*-wilt symptoms. The disease is characterised by chlorosis and wilting of lower leaves, followed by browning and drying-out, after which the symptoms spread to the rest of the stem or even the whole plant. Cross-sections of wilted stems and tubers show the discoloration of the vascular tissue. In potato early dying (PED) disease, where nematodes are also involved in the complex, dying-off of the plant occurs from the flowering stages. The yield of potatoes is then reduced because the growing period is shortened, resulting in a reduction in size of the daughter tubers (Millard and Denner, 2001).

2.3.3 Distribution of the disease in South Africa

Verticillium-wilt is predominantly distributed in the Sandveld-production area, with 53% of incidence in this region. Since 1998, however, it has also become a major problem in other areas of South Africa (Millard and Denner, 2001).

2.3.4 Life cycle of *V. dahliae*

V. dahliae overwinters as microsclerotia in the soil. After planting, root exudates from the developing potato plant stimulate the microsclerotia in the soil to germinate. Hyphae penetrate the roots and move to the inside of the vascular tissue, where they impede effective transport of water and nutrients by the xylem. The mycelium forms a specialised hypha (conidiophore) in which an asexual fungus spore (conidia) is produced. The conidia are produced continually by the mycelium as long as environmental conditions are favourable. The conidia that are formed within the xylem vessels are transported with the transpiration stream. They cause systemic colonisation of the entire plant by germinating at remote locations (Powelson and Rowe, 1994). Symptom development, including chlorosis and wilting, occurs as a result of toxin production and vascular dysfunction (Rowe, 1985).

Dying off of plants can occur from the flowering stages, leading to the name potato early dying disease. When the plant senesces and dies, the fungus colonises the dead tissues saprophytically. Microsclerotia are formed on the dead stems, which are often ploughed into the field during harvesting (Millard and Denner, 2001). Microsclerotia and conidia inoculum in the soil can be disseminated to uninfected fields by wind, water, etc. When the inoculum level in a field reaches a high enough value, it causes the infection of plants by the fungus.

2.3.5 Potato early dying (PED)

Potato early dying is a syndrome caused by a complex of interacting organisms. The same symptoms can be caused by infection by a variety of combinations between nematodes, *Colletotrichum*, *V. dahliae*, *V. albo-atrum*, *Erwinia* etc. (Rowe, 1985). PED is characterised by premature vine death and declining yields. Symptoms of PED are difficult to distinguish from normal senescence and initially the affected plants show only slower growth. Chlorosis and necrosis of the plant start at the bottom leaves, often occurring on only one side of the plant or on individual leaves.

The environment also has an influence on the development of PED. External abiotic factors such as temperature and moisture are implicated (Powelson and Rowe, 1993). Fungal infection of a plant is usually enhanced when the plant is stressed. The optimum growth temperature of potato plants is 18 - 20°C, while the optimal temperature for *V. dahliae* growth is 21 - 27°C. Therefore, the disease severity in potatoes infected with *V. dahliae* increases when the average air temperature rises from 20 to 28°C.

Wheeler *et al.* (1994) noted that when nematodes and *V. dahliae* both infect roots of potato, a disease complex can result that leads to yield losses higher than the sum of those expected for the two pathogens separately. The root lesion nematode *Pratylenchus penetrans* has been shown to interact with *V. dahliae* as a critical component of PED disease (Wheeler *et al.*, 1994). The exact mechanism of the *Pratylenchus-Verticillium* interaction in PED is unknown. Root wounding by nematodes was first believed to provide entry into the root for fungal pathogens. However, evidence from several pathosystems indicates that the interaction is biological or physiological, rather than physical, in nature. Perhaps the nematode feeding on the plant elicits physiological changes in the plant, which then favour infection by *V. dahliae* (Powelson and Rowe, 1993).

The role of nematodes was not determined in a recent survey of *Verticillium*-wilt of South Africa, so the prevalence of PED disease in this country is not known (C. Millard, personal communication).

2.3.6 Management of *Verticillium*-wilt

Management of this disease focuses on integrated control measures, where more than one strategy is followed simultaneously. Seed tubers are tested by enzyme-linked immunosorbent assay (ELISA) to certify them free from *Verticillium* and suitable for planting. Sanitation of equipment, such as farm implements, shoes, crates etc., that comes into contact with contaminated soil and tubers are important to prevent spreading of the fungus from one plot to another. Ploughing the topsoil layer, containing the microsclerotia, deep into the soil may bring cleaner soil to the surface and prevent contact between the pathogen and the roots.

2.3.6.1 Crop rotation

The crop used in crop rotation with potatoes should be chosen with care, since *V. dahliae* has a wide host range. Long rotation periods are necessary to decrease the inoculum levels of microsclerotia. At least three years without potato is required to adequately lower microsclerotia levels (Wheeler *et al.*, 1994). Brassicas, such as cabbage crops, are good choices since they act as biofumigants. They provide sulphur-containing compounds called glucosinolates that form fungal toxins upon degradation. In intact cells, the glucosinolates are separated from the enzyme myrosinase, a thioglucosidase, by subcellular compartmentalisation. In response to tissue damage, the myrosinase hydrolyses the glucosinolates to form unstable aglycone intermediates. These are converted into biologically active molecules, which include the volatile isothiocyanates, nitriles and thiocyanates (Buchanan *et al.*, 2000). Crop rotation with broccoli was shown to be effective against *Verticillium*-wilt of cauliflower (Millard and Denner, 2001).

2.3.6.2 Soil fumigation

Examples of chemical soil fumigants include methyl-bromide and metamsodium. It has been shown that it can control PED effectively in sandy soils, but the potential of this management tool is reduced by limited effectiveness in heavier soils (Wheeler *et al.*, 1994). Soil fumigation is also not a popular control strategy because of cost and safety concerns.

2.3.6.3 *Verticillium*-resistant cultivars

Resistant or tolerant cultivars are the most efficient approach to control of *Verticillium*-wilt of potato and PED disease. Most local cultivars, including BP1, are susceptible. Potato breeding programs of commercial potato resistant to *Verticillium* spp. began in 1949 in Idaho, USA (Wheeler *et al.*, 1994). A few *V. dahliae* resistant cultivars include Reddale, Russette, Gemchip and Ranger Russet (Wheeler *et al.*, 1994; Corsini and Pavek, 1996). Cultivars labelled *V. dahliae* tolerant include Alpha, Kondor, Desiree and Spunta (Tsror and Nachmias, 1995; Nachmias *et al.*, 1990). Désirée is a susceptible cultivar according to Kawchuk *et al.* (2001).

South African cultivars have been tested for *Verticillium*-resistance, but none were found (unpublished results). Currently there is no breeding program in South Africa for resistance against *Verticillium*-wilt (C. Millard (ARC-Roodeplaat), personal communication). ARC-Roodeplaat is involved in breeding strategies against common scab, bacterial wilt and *Fusarium* dry rot of potato.

2.3.7 Problems with breeding for resistance to *Verticillium*-wilt

Disease resistance breeding is made difficult by the complex inheritance of many traits of interest in potato and the problems associated with inbreeding. In particular, resistance to *Verticillium*-wilt appears to be a multigenic trait that makes selecting for it difficult. The breeders developing potato varieties have to select simultaneously for a combination of multigenic traits, including acceptable levels of resistance or susceptibility to diseases. When selecting for a single resistance gene, the population can be screened first for resistance to the disease, and a high proportion of the population will remain as the basis for variety development. However, early screening for resistance to *Verticillium* tends to eliminate clones with acceptable tuber maturity characteristics (Corsini and Pavek, 1996). This is because plants that stay in a non-tuberising juvenile condition do not become systemically infected with *Verticillium* and do not show wilt symptoms. Thus, researchers found it inefficient to select for *Verticillium* resistance at the early stages of variety development. They suggested to rather select for yield and other agronomic criteria in *Verticillium* infested fields.

2.3.8 Resistance to *V. dahliae*

Plant resistance to this disease is defined as the inability of the pathogen to penetrate the roots, to colonise the host tissue, to inactivate toxic elements or to inhibit fungal sporulation (Tsror and Nachmias, 1995). The genetic basis of *Verticillium* resistance has only been characterised in a few crops. Cotton has two dominant genes that control resistance to the disease (Tsror and Nachmias, 1995). Tomato has a single dominant resistance gene (*Ve*) that confers race-specific resistance to infection by *Verticillium* species. Some biotypes are not pathogenic on (*Ve*) tomato. *V. dahliae* is able to overcome the resistance by the formation of new pathotypes or races. A *V. dahliae* biotype called race 2 is pathogenic on tomatoes that contain the *Ve* gene (Nachmias *et al.*, 1987). The genetic basis of resistance has not been characterised in other crops.

Some potato varieties are tolerant to *Verticillium*. They show delayed or reduced colonisation by *V. dahliae* but do not show severe wilt symptoms or yield loss (Nachmias *et al.*, 1990). This phenomenon is referred to as 'Alpha-type tolerance' since it was first observed in the *Verticillium* disease-tolerant cultivar Alpha. This mechanism of tolerance, which does not exert genetic pressure on the pathogen, has not been explained yet (Tsror and Nachmias, 1995).

During reciprocal rootstock grafting experiments between resistant and susceptible cultivars, the root was established as the major site of resistance to *Verticillium*-wilt in tomato and tolerance in potato (Tsror and Nachmias, 1995). However, it appears that potato has additional defence mechanisms that are controlled by other genes. Two races of tomato *V. dahliae* show differential pathogenicity on susceptible and race 1-resistant (*Ve*) tomato cultivars (Nachmias *et al.*, 1987). *V. dahliae* toxin peptides were isolated from the culture fluid of the two races and used in several bioassays. (*Ve*) tomato cultivars are resistant to race 1, but race 2 forms symptoms on all (both resistant and susceptible) cultivars. The two toxins differ in amino acid composition, and it is believed that the toxin from race 2 has evolved to escape recognition by the (*Ve*) tomatoes. Race 1 *V. dahliae* does colonise (*Ve*) plants, but doesn't lead to the development of any disease symptoms. The data therefore suggests that the *Ve* gene doesn't have an effect on *V. dahliae* multiplication but instead limits damage. Experiments with these two races on four potato cultivars (three tolerant and one susceptible) did not show any differential activity between race 1 and race 2. Thus, it appears that a gene similar to *Ve* in tomato does not exist in potato (Tsror and Nachmias, 1995).

The *Ve* locus of tomato was characterised and two closely linked inverted genes were isolated (*Ve*1 and *Ve*2) (Kawchuk *et al.*, 2001). The *Ve* disease resistance genes were found to encode cell surface-like receptors. When the two *Ve* genes conferring resistance in tomato against *V. dahliae* race 1 was transformed into susceptible potato (cultivar Désirée), both genes independently conferred resistance to an aggressive race 1 isolate of *V. albo-atrum*. This resistance to a different pathogen species is a contradiction of the view of highly specific interaction between race-defining *R* genes. Thus, the *Ve* gene of tomato is pleiotropic because it can distinguish between race 1 and 2 of *V. dahliae* and possess the capacity to recognise another *Verticillium* species in a different host, the potato. Resistance of the transgenic potato lines against *V. dahliae* was not tested. *R* genes are thus able to retain their biological activity in other plant genera, and may be valuable in other agricultural crops that are infected by *Verticillium* species.

The deduced primary structure of *Ve*1 and *Ve*2 indicated that they are cell-surface glycoproteins with receptor-mediated endocytosis-like signals (Kawchuk *et al.*, 2001). Both have a hydrophobic N-terminal signal peptide, leucine-rich repeats (LRR) containing 28 or 35 potential glycosylation sites, respectively, a membrane-spanning domain and a C-terminal domain with endocytosis signals. The LRR is often associated with protein-protein interactions and ligand binding (Kobe and Deisenhofer, 1995). Since it is of the extracytoplasmic type, it could facilitate the recognition of an extracellular pathogen ligand. Eukaryotic cells use receptor-mediated endocytosis to communicate or to respond to external stimuli. In *Ve*, receptor-mediated endocytosis may be a mechanism to selectively bind ligands and then to remove the receptor-ligand complex from their surfaces, thereby responding to changing disease pressures.

The *Ve* cell surface-like receptor may transmit a ligand-induced conformational change from the extracytoplasmic to the cytoplasmic domain (Kawchuk *et al.*, 2001). Since it lacks a kinase domain, the cytoplasmic domain will need to interact with a kinase that phosphorylates the various downstream signalling proteins, thereby activating the subsequent signal transduction pathway. Alternatively, the receptor-mediated endocytosis may allow the extracellular domains and the ligands of the *Ve* receptor to directly stimulate signal transduction.

Elemental sulphur (S^0) was implicated as a phytoalexin of fungal vascular pathogens in the xylem of resistant lines of tomato (Williams *et al.*, 2002). It was accumulated much more rapidly in response to *V. dahliae* infection in disease-resistant than in disease-susceptible lines. It is fungitoxic to *V. dahliae*, inhibiting spore germination and mycelium growth. Elemental sulphur was predominantly localised in vascular structures that are in potential contact with the xylem-invading pathogen *V. dahliae* and thereby linked to defence against it.

2.3.9 Phylogeny of *V. dahliae*

The genus *Verticillium* includes various species, which falls into diverse econutritional groups (Bidochka *et al.*, 1999). Some are pathogens of insects, plants, mushrooms, nematodes, spiders and saprobes. The two commercially important plant pathogens, *V. dahliae* and *V. albo-atrum*, have very wide host ranges with little host specificity. Therefore, few physiological races have been found and characterised. Prior to the early 1970's, both species were considered forms of *V. albo-atrum*, but they are now considered to be separate species (Rowe, 1985). The two are separated on the basis of their survival structures, with *V. dahliae* forming true microsclerotia while *V. albo-atrum* only forms melanised hyphae. They also differ in their temperature sensitivity, with *V. albo-atrum* preferring cooler temperatures (up to 24°C) while *V. dahliae* grows well up to 27°C.

The internal transcribed spacer (ITS) and small nuclear (SN) rRNA regions of several *Verticillium* isolates were used to construct a phylogenetic tree (Bidochka *et al.*, 1999). Strains capable of infecting insects (*V. indicum*, *V. lecanii*) are present in divergent groups in the consensus tree, suggesting that this ability has evolved independently a few times. The plant pathogens (*V. dahliae*, *V. albo-atrum* and *V. nigrescens*) form a clade. They all produce pectinase enzymes able to degrade plant cell walls, while the insect and mushroom (*V. fungicola*) pathogens cannot degrade pectin. They, on the other hand, produce high levels of a subtilisin-like protease, capable of degrading insect cuticles. Together with the nematode pathogens, the insect and mushroom pathogens are distinguishable from the plant pathogens by their ability to produce chitinases. Strains of *Verticillium* therefore show enzymic adaptation to the polymers present in the integument of their particular host, being either plant or insect depending on their ecological niche (St Leger *et al.*, 1997).

V. dahliae is not host-specific, since it can infect a wide host range. Its host range includes trees, ground covers, shrubs, vines, fruits, vegetables, field crops, herbaceous ornamentals and many weeds (Powelson and Rowe, 1994). Plants regarded as resistant to *V. dahliae* include ferns, gymnosperms, many monocots and the cactus family. Isolates from a specific host are able to infect other hosts. *V. dahliae* isolates from cotton, potato, tomato and avocado were all pathogenic on tomato and were able to induce foliar symptoms (Visser, 1999). *V. dahliae* races 1 and 2 from tomato were used in potato inoculation experiments, and caused disease symptoms (Tsror and Nachmias, 1995). They did not, however, show differential pathogenicity on tolerant potato cultivars as is observed with (*Ve*) tomatoes.

2.4 Fungal virulence and plant resistance genes

There are many fungal - plant interactions in nature, some causing fungal plant diseases and some that do not (Laugé and De Wit, 1998). Interactions that do not lead to disease mostly occur on plants that are not hosts to the fungus, perhaps because the fungus lacks the pathogenicity factors to cause disease. Plants are hosts when the fungus is known to be a pathogen of the given plant. Not all combinations of fungal strains and host plant cultivars will, however, lead to disease. Therefore, interactions of the "host"-type can be divided into compatible and incompatible interactions. Compatible interactions occur when a susceptible plant is attacked by a virulent pathogen, and become diseased. Incompatible interactions are when a resistant plant is attacked by an avirulent pathogen, and does not become diseased.

The fact that specific interactions occur during compatible and incompatible interactions were demonstrated when Flor (1946) found that virulence appeared to be recessive and avirulence dominant. The gene-for-gene concept was formulated to state that for every dominant gene determining resistance (*R*) in the host, there is a matching dominant avirulence (*Avr*) gene in the pathogen. This accounts for the species-specificity of some fungi that can infect various plant species, but each strain can only infect one or a few host plants.

The elicitor-receptor model has been proposed for the *Avr* and *R* genes to give a biochemical explanation of the gene-for-gene concept. The specific elicitor, a product of an *Avr* gene of the pathogen, is recognised by a receptor, the product of a matching *R* gene in the resistant plant. This interaction activates a signal transduction pathway that leads to disease resistance, often through the hypersensitive response (HR). *Avr* genes and their matching *R* genes can be exploited in molecular resistance breeding against any pathogen that can be inhibited by HR (Laugé and De Wit, 1998).

From the perspective of the pathogen, why would it make sense to have an *Avr* gene if its product leads to a defense response in the plant? A recent idea is that the *Avr* gene products may act as virulence factors under certain conditions, for example on different host plants, or it may have another role that is not related to plant infection. Some *Avr* genes are always maintained within a pathogen population. There often appears to be a fitness penalty when avirulence mutates to virulence. This indicates that the gene products have important roles in pathogenicity (Hammond-Kosack and Jones, 1997).

Plant *R* genes have been used for a long time in breeding for disease control (Rommens and Kishore, 2000). Most *R* genes have only limited durability, which necessitates the continued discovery and introgression of new *R* genes. Because they can be linked to undesirable traits and the phenotype difficult to score, introgression using classical breeding approaches can be time-consuming and laborious. *R* gene-mediated resistance through genetic engineering is suggested to be the solution. *R* genes can be transferred within and across plant species, and have been shown to retain their activity. This approach also eliminates the retention of unwanted genetically linked germplasm. Non-host plants can be a source of extremely durable *R* genes, and they can be transferred to susceptible plants even though they are sexually incompatible.

2.5 Plant defence mechanisms

A plant`s first line of defence against colonisation by microbial pathogens is a preformed physical and chemical barrier. Superimposed on this is an array of inducible responses, and their activation depend upon recognition of the invading pathogen (Hammond-Kosack and Jones, 1997). This recognition event is very specific and is mediated by the interaction between the products of a microbial avirulence *Avr* gene and the corresponding plant disease resistance *R* gene. Following successful pathogen recognition, an oxidative burst and the hypersensitive response (HR) will occur. HR is the programmed death of challenged host cells, producing a visible area of cell death around the site of attempted pathogen invasion.

The oxidative burst is the production of reactive oxygen intermediates (ROIs), such as superoxide (O_2^-) and hydrogen peroxide (H_2O_2), at the site of attempted invasion. This is one of the most rapid responses following pathogen recognition. The oxidative burst and the transduction of the cognate redox signals play an important role in the diverse array of plant defence responses (Grant and Loake, 2000). ROIs integrate a diverse set of defence mechanisms, resulting in plant disease resistance. The oxidative burst is biphasic, the first phase being associated with wounding or infection by virulent microbial pathogens, while the second is correlated with the establishment of disease resistance. ROIs mediate redox signalling due to their ability to carry unpaired electrons. They oxidise nucleophilic

centres or coordinate interactions with transition metals at the allosteric or active sites of target proteins, thereby modulating their activity.

ROIs are generated by various potential mechanisms of which NADPH-dependent oxidase and peroxidases are the most studied (Grant and Loake, 2000). Nitric oxide (NO) has also been shown to accumulate during HR formation. A plant gene encoding NO synthase (NOS) has not been identified, but NO can be produced as a by-product from NO_2 accumulation by alternative mechanisms such as respiration, denitrification and nitrogen fixation. ROIs and NO may function in combination, possibly after reacting to form peroxynitrite ($ONOO^-$), to drive host cell death during HR (Delledonne *et al.*, 1998).

ROI production needs to be tightly regulated since it is highly cytotoxic (Grant and Loake, 2000). Ca^{2+} release following pathogen recognition is thought to be a trigger. It can either directly activate NADPH oxidase or indirectly by activating NAD kinase by binding to its activating protein, calmodulin. Ca^{2+}- mediated modulation of peroxidase activity is also involved in the regulation.

ROIs have various actions in facilitating defence against pathogens. They can cause strengthening of the cell wall, a physical barrier against pathogen penetration, by oxidatively cross-linking cell wall structural proteins, phenolics and polysaccharides (Bradley *et al.*, 1992). It also has direct antimicrobial activity. ROIs generated by the oxidative burst play a role in host cell death by initiating the development of the HR. ROIs stimulate Ca^{2+} influx into the cytoplasm, which is important for HR cell death, supposedly because of the requirement of cell death effectors for Ca^{2+} for their activity (Grant and Loake, 2000; Grant *et al.*, 2000).

Systemic acquired resistance (SAR) is the establishment of immunity to secondary infections in systemic tissues. It is long lasting, extends to tissues distant from the initial infection site and provides protection against a broad spectrum of microbial pathogens. It requires the accumulation of salicylic acid (SA) for its expression (Gaffney *et al.*, 1993; Delaney *et al.*, 1994). SA is proposed to act as an endogenous signal molecule required for inducing SAR. SA inhibits the activity of some antioxidant enzymes and positively enhances the production of superoxide (O_2^-). It can induce plant defence gene expression and systemic acquired resistance, but not cell death (McDowell and Dangl, 2000). ROIs may also act as systemic signals to establish plant immunity (SAR), by causing the deployment of cellular protectant functions in distal cells (Grant and Loake, 2000; Grant *et al.*, 2000).

Two other plant hormones frequently implicated in plant disease resistance are jasmonic acid and ethylene. Jasmonic acid and its cyclopentanone derivatives are synthesised by the octadecanoic pathway from linolenic acid in undamaged tissues and another pathway in wounded tissues (Xie *et al.*,

1998). It affects a variety of processes in plants, such as root growth, fruit ripening, senescence, pollen development and defence against insects and pathogens. Jasmonic acid alters gene transcription, RNA processing and translation. Ethylene is a simple two-carbon olefin, a plant hormone that is a potent modulator of plant growth and development (Wang *et al.*, 2002). It is involved in many aspects of the plant life cycle, such as seed germination, root hair development, flower senescence, abscission and fruit ripening. Its production is regulated by internal signals and from external stimuli such as pathogen attack and abiotic stresses. Ethylene plays an important role in plant disease resistance pathways. Depending on the type of plant and pathogen, its role may be dramatically different. Plants deficient in ethylene signalling may show increased susceptibility or increased resistance. In general, ethylene seems to inhibit symptom development during necrotrophic pathogen infection, but enhances cell death caused by other types of pathogen infection. The ethylene and jasmonic acid (ET-JA) and salicylic acid responses are mutually inhibitory. ET-JA dependent defence responses are activated by necrotrophic pathogen attack, while the SA-dependent response is triggered by biotrophic pathogens (McDowell and Dangl, 2000).

2.6 Polygalacturonases (PG)

Endopolygalacturonases (endoPGs) are found in a variety of organisms, such as bacteria, fungi and plants, where they are involved in the degradation and remodelling of the plant cell wall. They are poly[1,4-α-D-galacturonide] glycanohydrolases (EC 3.2.1.15), and they hydrolyse α-1-4 glycosidic linkages between galacturonic acids in homogalacturonans. Plants use their PGs in processes such as growth, fruit softening, root formation, organ abscission and pollen development. Phytopathogenic organisms use their PGs to penetrate and colonise their host plant tissues (De Lorenzo *et al.*, 2001). This review will focus on fungal PGs, and endoPGs in particular.

2.6.1 Fungal endoPGs

Endopolygalacturonases are the first detectable enzymes secreted by plant fungal pathogens when cultured *in vitro* on isolated plant cell walls (English *et al.*, 1971). The order, in which polysaccharide-degrading enzymes are secreted by plant fungal pathogens when they are cultured on isolated cell walls, may reflect the order in which they must work to degrade the cell walls. A "wall modifying enzyme" was found to be necessary before other enzymes, such as glycosidases, cellulases, hemicellulases and pectinases (pectin hydrolase, lyase and esterase), could degrade the cell wall polysaccharides (Karr and Albersheim, 1970). EndoPGs facilitate the ability of other fungus-secreted plant cell wall-degrading enzymes to attack their substrates, which are other plant cell wall components.

Fungal endopolygalacturonases play an important role during the early stages of plant pathogenesis (Karr and Albersheim, 1970). This enzyme spreads into the host tissue in advance of the invading fungal mycelium, and hydrolyses the pectic components in primary plant cell walls and middle lamellas (Yao *et al.*, 1996). This causes the cells to separate and the host tissue macerates, facilitating pathogen penetration and colonisation of the plant tissues. Extensive degradation of the plant cell walls lead ultimately to the death of the host cell.

Functional evidence for the role that endoPGs play in the virulence of a fungal pathogen was obtained when a mutation in one member of the PG family in *Botrytis cinerea* (*ΔBcpg1*) resulted in a reduction of virulence on tomato and apple (ten Have *et al.*, 1998). The mutants were still pathogenic and produced the same primary infections as control strains, but secondary infection, i.e. expansion of the lesion, was significantly decreased. The *Bcpg*1 gene is therefore required for full virulence.

While the fungal endoPG is disrupting the plant cell wall, the products of the degradation process are used by the fungus as a nutrient source for growth (Karr and Albersheim, 1970). It is also a potential avirulence factor by releasing cell wall fragments that signal the plant defence responses.

An example of an endoPG eliciting defence responses in plants is its induction of β-1,3-glucanases in *Phaseolus vulgaris* (Lafitte *et al.*, 1993). Purified endoPG from race β of *Colletotrichum lindemuthianum* was absorbed into near-isogenic lines of *Phaseolus vulgaris*, the one resistant and the other susceptible to this fungus. Induction of β-1,3-glucanase activity was earlier and the level higher in the resistant than in the susceptible isoline. This endoPG-mediated defence seems to be dependant on the release of pectic fragments of a critical size from the cell wall, because defence elicitation was abolished by the addition of an exopolygalacturonase to the bioassay. It degraded the oligogalacturonides released by endoPGs to elicitor-inactive monomers.

EndoPGs and mycelium from the fungus *Aspergillus niger* elicited necrosis on pods from cowpea (*Vigna unguiculata*) (Cervone *et al.*, 1987a). Oligogalacturonides with a degree of polymerisation greater than four and PG-released oligosaccharides from *Vigna* cell walls also elicited necrosis. PGs inactivated by heat or by antibodies were unable to elicit this response, indicating that the catalytic activity of the PG is required for its function as an elicitor. The plant may thus sense either the PG or the oligogalacturonides produced by it, or both, as elicitors of plant defence responses. This experiment supports the view that oligogalacturonides of a critical length are necessary for the induction of defence responses.

2.6.2 Fungal exoPGs

In addition to endoPGs, fungi also produce exo-polygalacturonases. They are not subject to inhibition by PGIPs (described in the next section) as are endoPGs (Hoffman and Turner, 1984; Cervone *et al.*, 1990). Two exoPGs from *B. cinerea* are described by Johnston *et al.* (1993). They release reducing sugars from pectin at a much lower rate than endoPGs, and because they are not inhibited by PGIP, they are responsible for part of the background activity in reducing sugar assays with fungal PG preparations. They also reduce the viscosity of a pectin solution at a slower rate than endoPGs. They are implicated in degrading oligogalacturonides released by endoPGs to elicitor-inactive monomers (Lafitte *et al.*, 1993).

2.6.3 The plant cell wall as defence mechanism

The plant cell wall is composed of complex polysaccharides, phenolics and structural proteins. It has important functions in maintaining the cell and tissue integrity. It also plays a complex role in resistance to invading pathogens. It is firstly a physical barrier to infecting pathogens (Karr and Albersheim, 1970). It acts as a source of nutrients for the pathogen and controls the production of degradative enzymes by the pathogen. It is composed of polysaccharides capable of regulating gene expression and defence responses. Enzymes and proteins involved in host defence mechanisms are localised in the plant cell wall.

Pectin is a complex saccharide that is present in the plant cell wall. It contains large amounts of galactosyluronic acid residues (York *et al.*, 1985), and can be broken down by a range of enzymes, such as endo- and exo-polygalacturonases, pectate lyases, pectin lyases and pectin methylesterases.

2.7 Polygalacturonase-inhibiting proteins (PGIPs)

Polygalacturonase inhibiting proteins (PGIPs) are basic proteins present in the cell wall of most dicotyledonous plants. PGIPs are specific, reversible, saturable, high-affinity 'receptors' for fungal, but not plant, endopolygalacturonases (Cervone *et al.*, 1987b, 1989, 1990). PGIPs reduce the activity of endoPGs to different extents and are highly specific. PGIP is structurally related to several resistance gene products, since it belongs to the super-family of leucine-rich repeat (LRR) proteins (Mattei *et al.*, 2001). These LRR proteins are specialised for the recognition of non-self molecules and rejection of pathogens.

2.7.1 Action of PGIP

When fungal endoPGs attack plant cell wall pectic polymers, they produce oligomeric α-1,4-linked oligogalacturonides. Oligomers with especially 10-13 residues are elicitors of plant defence

responses. The endoPGs then rapidly depolymerise them into shorter inactive molecules (Cervone *et al.*, 1989).

Cervone *et al.* (1987b) performed several experiments *in vitro*, which lead to the formulation of a hypothesis for the action of PGIP in plant defence responses. Fungal PGs were incubated with polygalacturonic acid in the absence and presence of bean PGIP. PGIP affected the amount of oligogalacturonides with a degree of polymerisation of four and higher, by retarding the PG-catalysed hydrolysis of polypectate to mono- and digalacturonate.

The elicitor-activity of oligogalacturonides produced from polygalacturonic acid by fungal PGs in the presence and absence of bean PGIP was also assayed (Cervone *et al.*, 1989). The degree of polymerisation of oligogalacturonides was reduced at a much slower rate in the presence of PGIP. The elicitor activity of the digestion products in the presence of PGIP was much higher. The most active group of oligomers was of an intermediate size, with polymerisation between 10 and 13.

The hypothesis based on the *in vitro* evidence states therefore that the complex formed between the endoPG and PGIP leads to more stable oligogalacturonides that have elicitor activity (Cervone *et al.*, 1989; De Lorenzo *et al.*, 1990). The elicitor active molecules accumulate at the site of infection. The PGIP can therefore act against fungal invasion of the plant by causing fungal PGs to increase their elicitation of plant defence responses (Cervone *et al.*, 1987b).

In an experiment with pear PGIP and a complex mix of *B. cinerea* PGs, Sharrock and Labavitch (1994) attempted to verify the hypothesis of Cervone, namely that PGIP modifies cell wall degradation to result in more elicitor active pectic fragments of a specific size class. They attempted to replicate the *in vivo* environment of PGIP by incubating the enzymes with pear cell walls. They found, however, that Cervone's hypothesis is only a model system that either works only *in vitro* or it may only work in some PG:PGIP interactions e.g. bean PGIP but not pear PGIP. Oligomeric breakdown products did not accumulate due to the presence of a component that rapidly degraded intermediates to monomers and dimers. This component was demonstrated to be isozymes of *B. cinerea* polygalacturonase that remained uninhibited by pear PGIP.

2.7.2 History of discovery of PGIP

Proteins were extracted from various tissues that were able to inhibit the activity of fungal polygalacturonases to various extents (Albersheim and Anderson, 1971). The protein from bean hypocotyls formed a complex with a very low dissociation constant with the polygalacturonase from *C. lindemuthianum*. The existence of PGIP was discovered by the apparent lack of extractable fungal PG from infected plant tissues. Lack of pectinase activity in a 1 M NaCl extract of *Cladosporium*

cucumerinum infected cucumber hypocotyls was thought to occur by the co-extraction of a component from the cell wall that specifically inhibited the pectinase (Skare *et al.*, 1975). The apparent absence of PG in various fruit infected with plant pathogenic fungi lead to the discovery of a protein capable of inhibiting the fungal PG (Fielding, 1981). The author found the inhibitory activity of PGIP to separate from the PG during isoelectric focusing. No inhibitors were found in extracts of healthy tissues. It appears as if the infection process stimulated the formation of the inhibitors. Little or no endoPG activity was found in *Capsicum* fruits infected with *Glomerella cingulata* (Brown and Adikaram, 1982) and pear fruit infected with *B. cinerea* and *Dothiorella gregaria* (Abu-Goukh and Labavitch, 1983). It was shown that these fruits contained proteins that inhibited pectinases.

2.7.3 PGIP discovered in various plants

PGIP has since been discovered in the cell wall of all dicotyledonous plants that have been examined so far, as well as a few monocots (onion and leek). The gene encoding PGIP has subsequently been cloned from several plants. Table 2.1 provides a summary of the PGIP proteins and genes studied as well as the literature references.

2.7.4 Protein structure of PGIP

PGIP has not yet been crystallised, so the protein structure presented here is only a prediction. The predictions are based mostly on amino acid sequence predicted from the DNA sequences, and modelling by Mattei *et al.* (2001).

2.7.4.1 N-terminal signal peptide

All mature PGIPs are preceded by a 24 amino acid hydrophobic signal peptide. It is almost identical in pear and apple PGIP (only two substitutions, Yao *et al.*, 1999), and the potential cleavage site (Ala-Leu-Ser) for the signal peptidase is conserved between the apple, pear, bean and raspberry PGIPs (Yao *et al.*, 1999; Ramanathan *et al.*, 1997). The alanine is substituted with a serine in the cleavage site of tomato PGIP (Stotz *et al.*, 1994). The signal peptide targets the PGIP through the endomembrane system (Von Heijne, 1985), for targeting to the apoplast or translocation into the endoplasmic reticulum (De Lorenzo *et al.*, 2001). This is consistent with the proposed cell wall localisation of PGIP (Abu-Goukh *et al.*, 1983b), the observation that PGIP is in the extracellular matrix in bean hypocotyls and that it is secreted into the medium by suspension-cultured bean cells (Salvi *et al.*, 1990).

Table 2.1 **Summary of characterised PGIP proteins and genes.** The plants in which they were identified, the literature references as well as the tissue source from which the protein was isolated are indicated.

Plant	Scientific name	Reference		
		Protein/ Extract	Plant tissue	Gene
Alfalfa	*Medicago sativa*	Degrá *et al.*, 1988	callus	
Apple	*Malus domestica*	Fielding, 1981;	fruit	Yao *et al.*, 1999;
		Brown, 1984;		Arendse *et al.*, 1999
		Yao *et al.*, 1995		
		Müller and Gessler, 1993	leaves	
Arabidopsis	*Arabidopsis thaliana*			De Lorenzo *et al.*, 2001
Bean	*Phaseolus vulgaris*	Albersheim and Anderson, 1971;	hypocotyl	Toubart *et al.*, 1992;
		Anderson and Albersheim, 1972;		Berger *et al.*, 2000
		Cervone *et al.*, 1987b;		
		Berger *et al.*, 2000		
Green peppers	*Capsicum annuum*	Brown and Adikaram, 1982	fruit	
Cotton	*Gossypium hirsutum*	James and Dubery, 2001	hypocotyl	
Cucumber	*Cucumis sativus*	Skare *et al.*, 1975	hypocotyl	
Eucalyptus spp.				Chimwamurombe *et al.*, 2001
Leek	*Allium porrum*	Favaron *et al.*, 1993, 1997;	stalk	
		Favaron, 2001		
Lupin	*Lupinus albus*	Costa *et al.*, 1997	root	
Onion	*Allium cepa*	Favaron *et al.*, 1993	bulb	
Pea	*Pisum sativum*	Hoffman and Turner, 1982, 1984	leaflets	
Pear	*Pyrus communis*	Abu-Goukh *et al.*, 1983a, 1983b;	fruit	Stotz *et al.*, 1993
		Abu-Goukh and Labavitch, 1983;		
		Stotz *et al.*, 1993		
Potato	*Solanum tuberosum*	Machinandiarena *et al.*, 2001	leaves	
Raspberry	*Rubus idaeus*	Johnston *et al.*, 1993;	fruit	Ramanathan *et al.*, 1997
		Williamson *et al.*, 1993		
Soybean	*Glycine max*	Favaron *et al.*, 1994	seedlings	Favaron *et al.*, 1994
Tomato	*Lycopersicon esculentum*	Brown and Adikaram, 1983;	fruit	Stotz *et al.*, 1994
		Stotz *et al.*, 1994		

2.7.4.2 Leucine-rich repeats (LRR)

The mature PGIP protein consists of three domains, namely the central LRR region, and two cysteine-rich flanking regions (Mattei *et al.*, 2001). Figure 2.1 is a schematic drawing of the secondary structure of PGIP indicating these domains. The internal domain consists of tandem repeated units, each a modification of a 24 amino acid peptide. Alignment of 24 different mature PGIP sequences revealed that they consist of 10 repeats of modifications of the 24 aa leucine-rich peptide (De Lorenzo *et al.*, 2001). Leucine residues are regularly spaced in the consensus sequence of GxIPxxLxxLKnLxxLdLSxNxLx, with residues conserved in at least four repeats indicated with capital letters, x's being non-conserved residues and small letters residues identical in five or less repeats. The LRR matches the extracytoplasmic consensus also found in other *R* genes that participate in gene-for-gene resistance.

Leucine-rich repeat (LRR) proteins are found in a variety of organisms and have diverse functions and cellular locations. In most of the cases, they play a role in protein-protein or protein-ligand interactions (Kobe and Deisenhofer, 1995). It is the LRR motif in these proteins that is responsible for the protein-protein interactions (De Lorenzo *et al.*, 2001). They are mostly membrane associated or involved in signal transduction, so proteins containing a domain of tandem Leu-rich repeats may be receptors for other macromolecules (De Lorenzo *et al.*, 1994). The consensus sequence for the repeat motif is very similar between proteins, considering their diverse functions and wide distribution of organisms. This indicates a strong selection pressure for the conservation of this structure.

In plants, LRR proteins play an important role in defence, by facilitating host-pathogen interactions. Most plant *R* genes encode leucine-rich repeat (LRR) proteins (Hammond-Kosack and Jones, 1997), indicating that protein-protein interactions play an important role in plant disease resistance. The functions of plant *R* genes are to recognise pathogen *Avr* gene products and to initiate an induced response. Some of the *R* gene products do this through the LRR that acts as a putative receptor. Sequence variation in the LRRs is thought to influence the recognition specificity of these *R* gene products. The role of the LRR domain in other *R* gene products can, however, also be dimerisation or interaction with upstream or downstream signalling components.

As already mentioned, PGIP of plants also has the LRR structure. PGIP is not directly involved in pathogen recognition, but does bind to fungal endoPGs (Stotz *et al.*, 2000). It is one of the few plant LRR protein for which the ligand is known, so it may prove useful in studying the structural bases of recognition specificity of other plant LRR *R* proteins (Leckie *et al.*, 1999). PGIP is proposed to be the secreted receptor component of the cell-surface signalling system involved in the recognition event between plant and fungi. The LRR of PGIPs may be required for the interaction with and inhibition of fungal PGs (Stotz *et al.*, 1994).

The porcine ribonuclease inhibitor (PRI) was studied, and proved valuable in understanding the interaction between PGIP and PG (Kobe and Deisenhofer, 1995). It consists of a repeated β-strand/β-turn structural unit, arranged into a parallel β-sheet, with the molecule having a horseshoe shape. The motif xxLxLxx is repeated so that the leucines form a hydrophobic core and the sidechains of the flanking amino acids are solvent exposed and able to interact with the ligand. It is proposed that PGIP is also composed of a parallel stacking of β-strand/β-turns, forming an arch-shaped protein with a solvent exposed surface (Leckie *et al.*, 1999). The amino acids that determine the specificity and affinity of PGIP to its target fungal PG may be displayed on the solvent-exposed area of β-strand/β-turn region (De Lorenzo *et al.*, 2001).

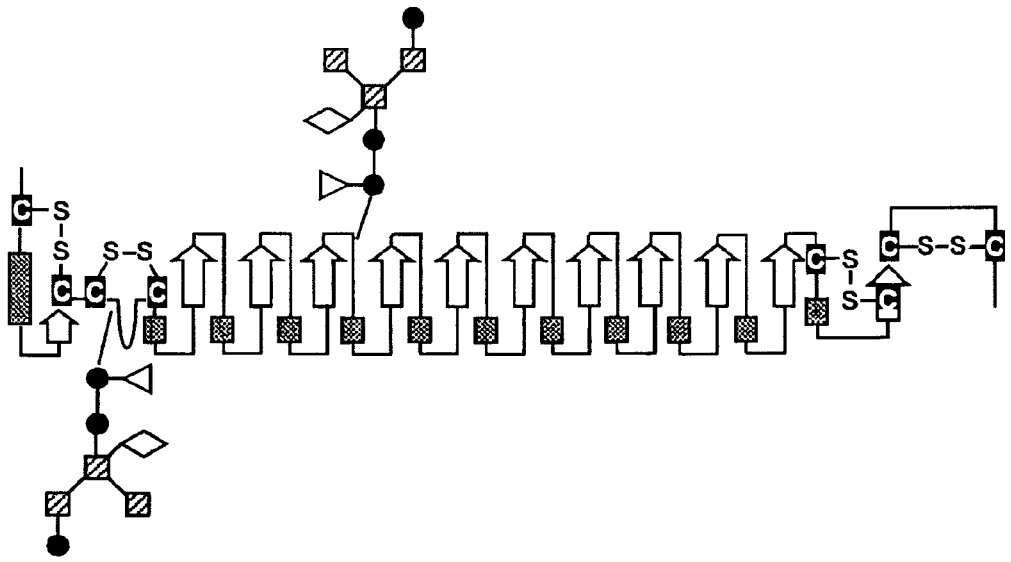


Figure 2.1 **Schematic drawing of PGIP secondary structure elements.** Arrows indicate β-strands and boxes indicate α-helices. Glycan structure has been sketched as follows: (●) GlcNAc (*N*-acetyl glucosamine), (□) Man (mannose), (▽) Fuc (α(1,3)-fucose), (◊) Xyl (β(1,2)-xylose). (Mattei *et al.*, 2001).

2.7.4.3 Cysteine residues

Conserved positions of eight cysteine residues, clustered at the N- and C-terminus of the mature peptide (Figure 2.1), suggest disulphide bonds are involved in stabilising the tertiary structure of PGIP (tomato, pear and bean PGIP: Stotz *et al.*, 1994). This may be responsible, at least partially, for the heat stability and resistance to proteases of PGIPs. They may also be required for the correct folding of the extracellular peptide (Ramanathan *et al.*, 1997).

2.7.4.4 Glycosylation

PGIP is a glycoprotein, containing seven conserved N-glycosylation sites. Differentially glycosylated forms of PGIP account for the heterogeneity in molecular mass (44 - 54 kDa) observed for PGIP that has not been deglycosylated (soybean: Favaron *et al.*, 1994; tomato: Stotz *et al.*, 1994; transgenic pear PGIP in tomato: Powell *et al.*, 2000). Chemically deglycosylated PGIP from pear, tomato, apple and lupin has the same molecular mass of 34 kDa (Stotz *et al.*, 1993, 1994; Yao *et al.*, 1995; Costa *et al.*, 1997). Glycosylation of PGIP may influence inhibitor activity, since differentially glycosylated pear PGIP expressed in transgenic tomato were not equally effective as fungal PG inhibitors (Powell *et al.*, 2000). Figure 2.1 shows the typical structure of plant N-linked oligosaccharides as was determined for bean PGIP-2. The innermost *N*-acetyl glucosamine (GlcNAc) contains an α(1,3)-fucose (Fuc). N-glycans were identified to be attached to the Asn residuces of the predicted Asn-Xxx-Thr/Ser consensus site (Mattei *et al.*, 2001).

2.7.5 Gene structure of PGIP

The gene for PGIP is a single open reading frame of approximately 1000 bp. It codes for a mature polypeptide between 300 and 315 aa in length and with a molecular mass around 44 kDa (De Lorenzo *et al.*, 2001). The nucleotide sequence of the PGIP transcript is colinear with the genome sequence in apple, pear, tomato and bean PGIP (Yao *et al.*, 1999; Stotz *et al.*, 1993, 1994; Toubart *et al.*, 1992) but *A. thaliana* and raspberry PGIP are exceptions (De Lorenzo *et al.*, 2001; Ramanathan *et al.*, 1997). They contain introns of which the positions are conserved in both the *A. thaliana* and raspberry genes (De Lorenzo *et al.*, 2001).

2.7.6 Evolution of PGIP genes

PGIP inhibition specificities and kinetics vary within and between species. This reflects counter-adaptations between fungal PGs and plant PGIPs that lead to specialisation (Stotz *et al.*, 2000). PGIP genes diverged by evolutionary adaptation to different forms of PG encountered, either because of variation in the distribution of different fungal pathogens producing the PGs or because of the evolutionary response of the PG to effective inhibition.

Evolution of PGIP genes occurred by advantageous substitutions, adapting in response to natural selection. Most amino acid substitutions occurred in the solvent-exposed β-strand/β-turn segment of the LRR of PGIPs (Stotz *et al.*, 2000). Amino acid replacements in the hydrophobic core were mostly conservative, indicating the structural role. Differences among PGIPs from different plants mainly consist of substitutions and insertions or deletions of a few amino acids. This indicates that duplication and point mutations are the major driving force in the evolution of the *pgip* families (De Lorenzo *et al.*, 2001).

Compared to other resistance genes containing LRRs, *pgip* appears to be constrained in the number of sites that could change adaptively (Stotz *et al.*, 2000). This can be explained by the fact that PGIP might need to recognise much less divergent forms of PG whereas *R* genes need to evolve rapidly to recognise highly divergent elicitors. Other possible reasons include that PGIP needs to maintain inhibition of multiple fungal PGs and that it must not inhibit endogenous plant PGs. For these reasons, there might only be a few residues that can change inhibition specificity advantageously and without deleterious effects.

2.7.7 Specificity of interaction with PGs

Conclusions on the specificities of PGIP-PG interactions were based on experiments with purified PGIP and PG extracts. Plant PGIPs can interact with endoPG from fungi, but not endoPGs of plant or bacterial origin. Plant PGIP can also not inhibit fungal pectin and pectate lyases or fungal exoPGs (Hoffman and Turner, 1982, 1984; Brown and Adikaram, 1983; Abu-Goukh and Labavitch, 1983; Cervone *et al.*, 1990; Johnston *et al.*, 1993).

Fungi produce many PGs, and each has a different expression pattern *in planta* and *in vitro* (Wubben *et al.*, 1999). PGIPs differ in their PG-target specificity within and between plant species. A particular plant PGIP can also selectively inhibit individual PG isoforms produced by a single fungus. For example, the major pear PGIP that inhibits only certain PG isozymes from *B. cinerea* (Sharrock and Labavitch, 1994).

Positively selected amino acids of fungal PGs are surface exposed and able to interact with other proteins, such as PGIPs. They are distributed into several distinct regions, making it impossible for a single plant PGIP to contact them all at once. Rather, different PGIPs target different domains of PG, none of which are the active site. This explains the competitive and non-competitive inhibition kinetics found for different PGIPs (Abu-Goukh *et al.*, 1983a; Johnston *et al.*, 1993; Stotz *et al.*, 2000), the fact that endogenous fruit PG is not inhibited (pear PG: Abu-Goukh and Labavitch, 1983) and that mutations of the active site abolishes PG activity but does not prevent PGIP binding (Caprari *et al.*, 1996).

Variations in the LLR structure of PGIP influence recognition specificities. The solvent-exposed residues in the β-strand/β-turn motifs of the LRRs determine the specificity and affinity of PGIP to its target fungal PG (De Lorenzo *et al.*, 2001). Even one amino acid substitution is sufficient to alter the ability of PGIP to interact with its ligand (Leckie *et al.*, 1999). This can be illustrated by the example of bean *pgip1* and *pgip2*, where single amino acid differences in the β-strand/β-turn region causes them to have distinct specificities. Each bean *pgip*'s PG-interacting specificities can be conferred onto the other when certain amino acids are substituted with their counterparts'. The ability of bean *pgip2*

to inhibit *Fusarium moniliforme* PG was conferred to bean *pgip1* when one amino acid was changed from a lysine to a glutamine, the corresponding amino acid in bean *pgip 2*. Synonymous nucleotide changes between bean *pgip1* and *pgip2* corresponded mostly to residues located outside the β-strand/β-turn structural motif.

2.7.8 Gene families of PGIP

Most plants studied contain a small family of PGIP genes. Small variations in the structure of different PGIPs from a specific plant might provide resistance to a variety of pathogens (Pressey, 1996). A pure extract of PGIP from bean contains a mixture of PGIPs that is difficult to separate biochemically due to their similar characteristics. Nevertheless, they have different specificities against fungal PGs, and may therefore protect the plant against a variety of fungi (Desiderio *et al.*, 1997; Pressey, 1996).

At least two closely related copies of PGIP are found in pear (Stotz *et al.*, 1993) and tomato (Stotz *et al.*, 1994). The different PGIPs from the same plant tissue may have different target PG specificities. PGIP sequences are particularly divergent in the LRR domains, while the N- and C-terminal portions of the proteins are more conserved (De Lorenzo *et al.*, 2001). It may be these sequences that are responsible for the differences in inhibition kinetics and interaction with different fungal PGs (Stotz *et al.*, 1993, 1994). These differences in specificity might explain the observed host-pathogen compatibility. Heterologous PGIPs may be especially useful to improve disease resistance of crops to certain fungal pathogens.

PGIPs from pear and apple (*Malus domestica)* are very closely related, and may be orthologous genes, that is, direct evolutionary descendants from a common ancestral gene. Because bean and soybean PGIP differ so much, they may be paralogues, descendants of duplication events that occurred prior to speciation. Duplications also occurred after divergence of the species, as evident by the pairs of PGIP genes in *A. thaliana*, bean and soybean (Stotz *et al.*, 2000).

Apple and raspberry have low copy number gene families (Yao *et al.*, 1999; Ramanathan *et al.*, 1997). There are multiple copies of PGIP genes in the bean genome (Frediani *et al.*, 1993). They are clustered and cytologically localised to one chromosomal region, the pericentromeric heterochromatin of chromosome pair X. PGIP genes from five commercially important *Eucalyptus* species were cloned and the sequences compared (Chimwamurombe *et al.*, 2001). They were found to be highly conserved, with 98 to 100% identity. The fact that PGIPs have high homologies within genera strengthens the belief that PGIP plays an important role in plants.

2.7.9 Localisation of PGIP in the plant

Various indications exist that PGIP is localised in the plant cell wall (Albersheim and Anderson, 1971; Skare *et al.*, 1975; Abu-Goukh *et al.*, 1983b; Salvi *et al.*, 1990; Johnston *et al.*, 1993). It is bound ionically to cell walls since it is extractable by salt. PGIPs are basic proteins, and their high pI values may facilitate binding of the positively charged molecules to negatively charged pectin in the plant cell wall. Soluble forms of the PGIP protein have also been discovered (Hoffman and Turner, 1982).

A basal level of PGIP is present in the cell wall of uninfected cells, but levels increase significantly upon fungal attack. PGIP accumulates preferentially in the epidermal cells immediately surrounding the infection site, as in the example of *C. lindemuthianum* infection of *P. vulgaris* leaves and hypocotyls (Bergmann *et al.*, 1994; Devoto *et al.*, 1997). This indicates the role of PGIP as one of the first defence gene products in plant disease resistance, since epidermal cells present the first structural barrier to invading pathogens. Induction of PGIP in the surrounding epidermal cells may occur due to the local signals generated during fungal penetration, namely the oligogalacturonides produced by endoPG:PGIP interaction, or fungal oligoglucosides generated from the fungus itself.

PGIP is localised in the apoplast of *P. vulgaris* stems. PGIP levels are the lowest at the roots and highest at the vegetative apex and the flower (Salvi *et al.*, 1990). It is speculated that PGIP activity is low in roots so as not to prevent invasion by beneficial mycorrhizal fungi, which are expected to produce the same kind of polygalacturonase as phytopathogenic fungi. Because of the occurrence of pectic enzymes in pollens and PGIP in flowers, such as *P. vulgaris* flowers, it is an interesting question whether PGIP might act as a recognition factor for alien pollens.

2.7.10 Timing of PGIP expression

Levels of PGIP activity are higher in immature than in mature fruit: pear (Abu-Goukh *et al.*, 1983b), apple (Yao *et al.*, 1999), raspberry (Johnston *et al.*, 1993) and tomato (Powell *et al.*, 2000). It is suggested that the pre-existing PGIP in young, developing tomato fruit could limit tissue colonisation by inhibiting fungal PGs (Powell *et al.*, 2000). As fruit mature, less PGIP is present and defence against pathogens decline. The decomposition of the cell wall by pathogens may be advantageous to the plant, because it would facilitate the release of mature seeds into the environment.

Expression of the PGIP gene may, however, continue at a consistent level in all the developmental stages, as detected in raspberry using RT-PCR and northern analysis (Ramanathan *et al.*, 1997). If it is assumed that constant levels of PGIP transcript causes constant rates of PGIP expression, the drop in PGIP activity in ripe raspberry fruit may be caused by post-translational modification or a factor present during fruit ripening that downregulates the activity of PGIP.

2.7.11 Inhibition kinetics of PGIP

Kinetic studies have shown that PGIPs inhibit fungal PGs by different types of mechanisms. PGIP showing competitive inhibition is those of pear (Abu-Goukh *et al.*, 1983a) and bean PGIP-2 in its interaction with *F. moniliforme* PG (Federici *et al.*, 2001).

Mutation studies with *F. moniliforme* PG indicated that Histidine-234, located in the active site, is critical for the enzymatic activity (Caprari *et al.*, 1996). Modifying this residue did not alter the capacity of the PG molecules to interact with bean PGIP-2. The authors thus proposed that the site responsible for PGIP recognition must reside in a domain different from the active site. Normally this would represent a situation of non-competitive inhibition.

The crystal structure of a PG from *F. moniliforme* was determined and the amino acids important for interaction with *Phaseolus vulgaris* PGIP-2 elucidated by site-directed mutagenesis (Federici *et al.*, 2001). Three amino acids, located inside and on the edge of the active site cleft, were found to be critical for the formation of the complex. This was consistent with the competitive inhibition effect that PGIP-2 has on the PG. By substituting residues of the fungal PG with residues found in plant PGs, the enzyme was unable to interact with the PGIP. This suggests how plant PGs may escape recognition by PGIPs and maintain their function in the presence of PGIP. The two residues located inside the active site cleft, Arginine-267 and Lysine-269, are likely to play a role in substrate binding. Mutations of these two residues also negatively affected the interaction of the PG with PGIP. Thus, the binding of the PGIP to these two residues prevents the binding of the substrate to the enzyme and thereby inhibits its activity. Other catalytic residues of the active site were shown not to form contacts with PGIP-2. This result corresponds to the result obtained by Caprari *et al.* (1996). The proposed mechanism of PG inhibition by PGIP is therefore the prevention of substrate access while at the same time covering the active site cleft (Federici *et al.*, 2001).

PGIP showing non-competitive inhibition include that of raspberry (Johnston *et al.*, 1993) and tomato (Stotz *et al.*, 2000). In this type of inhibition, the inhibitor binds to a site on the PG molecule that is different from the active site. Mutations of the active site abolish PG activity but do not prevent PGIP binding.

PGIP from apple shows mixed inhibition kinetics with PG from *A. niger* (Müller and Gessler, 1993; Yao *et al.*, 1995). Cotton PGIP also has a mixed or non-competitive effect on *A. niger* PG (James and Dubery, 2001). Cotton PGIP interacts in a positive cooperative manner with an extracellular endoPG from *V. dahliae* (James and Dubery, 2001). A sigmoidal rather than typical Michaelis-Menten curve was obtained. Cooperative allosteric interactions occur when the binding of one ligand at one site is influenced by the binding of another ligand at a different (allosteric) site on the enzyme (Palmer,

1995). The positive cooperative inhibition indicates allosteric interactions between the PG enzyme, inhibitor protein and polygalacturonic acid (PGA) substrate. The reaction rate is reduced at low substrate concentrations, but not at others. The PGIP reduced the reaction rate and decreased the substrate affinity of the PG, thereby contributing to the accumulation of elicitor-active oligogalacturonides.

Since these studies used distinct PGs, the differences in inhibition kinetics may be due to the specific target PG, and not the PGIP properties. The mixed results may be due to impure preparations of PG or PGIP, or both.

2.7.12 Inducers of PGIP expression

PGIP is induced by various stress stimuli, such as:

- Wounding (Soybean: Favaron *et al.*, 1994; Bean hypocotyls: Bergmann *et al.*, 1994; Apple: Yao *et al.*, 1999; Potato: Machinandiarena *et al.*, 2001).
- Elicitors (of plant or fungal origin): elicitor-active oligogalacturonides or fungal glucan. Bergmann *et al.* (1994) showed that the accumulation of bean PGIP in suspension cultures in the response to elicitor treatment was correlated with an increased accumulation of *pgip* mRNA.
- Pathogen infection, especially incompatible interactions with the fungus, manifested by the hypersensitive response (*C. lindemuthianum* and bean hypocotyls: Bergmann *et al.*, 1994; Nuss *et al.*, 1996 and Devoto *et al.*, 1997).
- Salicylic acid treatment causes an increase in PGIP levels (Cotton hypocotyls: James and Dubery, 2001; Bean hypocotyls: Bergmann *et al.*, 1994; Potato leaves: Machinandiarena *et al.*, 2001). Salicylic acid is a molecule implicated in the systemic induction of plant defence responses. During pathogen attack, salicylic acid increases both locally and systemically.
- Developmental factors. PGIP activity and transcripts were found at different levels in different organs of bean seedlings and plants (Salvi *et al.*, 1990; Devoto *et al.*, 1997).

2.7.13 Examples of PGIP

2.7.13.1 Apple PGIP

Much work has been done on the PGIP from apple (*M. domestica*) (Brown 1984; Müller and Gessler, 1993; Yao *et al.*, 1995, 1999; Arendse *et al.*, 1999). Inhibitor activity against endopolygalacturonases from several fungi was first discovered in the cell walls of four apple cultivars, namely Granny Smith, Golden Delicious, Cox's Orange Pippin and Bramley's Seedling (Brown, 1984). It was shown that the cultivar with the highest PGIP levels displayed the least tissue maceration and rot expansion during fungal inoculation experiments.

Proteinaceous inhibitors of fungal PGs were previously only detected in the cell walls of infected apple fruits (Fielding, 1981), but Brown (1984) showed the endoPG inhibitor to be present also in healthy fruits of at least some cultivars. It is suggested that it plays a role in cultivar resistance to some pathogens. It has since been found also in leaves of apples (Müller and Gessler, 1993).

Protection against pectic enzymes by oxidised polyphenols declines as the sucrose levels of fruit increase (Brown, 1984). Therefore, PG inhibitors become more important as protectors against rot as fruit ripens. Unfortunately, the endoPG inhibitor was only present in pre-climacteric, ripening and rotted apple fruit, and seemed to decline when fruits became very ripe (Brown, 1984). This then leads to a conflict, with fruits becoming more susceptible to fungal infection as they near ripening. If PGIP can be expressed during the ripe stages, fruit will be much better protected against fungal infection. According to the hypothesis of the action of PGIP, it will firstly inhibit colonisation and secondly activate host defence responses at the site of infection.

PGIP from Golden Delicious apple fruit was characterised by Yao *et al.* (1995). It was found to be a glycoprotein, with the chemically deglycosylated polypeptide having a molecular mass of 34 kDa. It had a mixed inhibition effect on PG II produced by *B. cinerea* in liquid culture. Other PGs secreted in liquid culture were inhibited differentially, while PG produced in apple fruit inoculated with *B. cinerea* was not inhibited at all. An apple *pgip* gene from *M. domestica* cv. Granny Smith leaves was cloned using degenerate oligo-primed PCR and Inverse PCR (Arendse *et al.*, 1999). Its sequence was found to be identical to that of a PGIP cloned from cv. Golden Delicious by Yao *et al.* (1999). The gene has 990 nucleotides and contains no introns. It encodes a predicted polypeptide of 330 amino acids, of which the first 24 amino acids are the signal peptide. The full-length mRNA transcript is 1.3 kb in size. The predicted mature protein has a calculated molecular mass of 34 kDa and a pI of 7.0. It is predicted to contain 10 imperfect leucine-rich repeat (LRR) motifs that span 80% of the mature peptide. The apple PGIP polypeptide sequence is more homologous to that from pear (Stotz *et al.*, 1993) and other fruit than those from bean and soybean (vegetables). A Southern blot indicated that apple *pgip* might be part of a small gene family of PGIP homologous genes (Yao *et al.*, 1999).

The gene is developmentally regulated since different PGIP transcript levels were present in fruit collected at different maturities (Yao *et al.*, 1999). It is also induced by fungal infections and tissue wounding, with PGIP transcripts increasing locally in the decayed and surrounding areas, but the gene is not activated systemically in tissue distant from the site (Yao *et al.*, 1999). Storage of fruit results in the reduction of the PGIP transcript level, which coincides with increased susceptibility to fungal colonisation and maceration.

2.7.13.2 Pear PGIP

Abu-Goukh and co-workers (Abu-Goukh and Labavitch, 1983; Abu-Goukh *et al.*, 1983a, 1983b), made a detailed study of polygalacturonase inhibitors from "Bartlett" pear fruits. Very young fruit resisted infection by several fungal pathogens, but fruit resistance declined steadily with maturation. PGIP levels also decreased as the fruit matured (Abu-Goukh *et al.*, 1983b). This direct correlation between decreased resistance and decreased PGIP activity suggests a specific role of PGIP in defence.

The PG inhibitor of pear was found to have no effect on pear PG, indicating that the inhibitor has no direct role in fruit ripening (Abu-Goukh and Labavitch, 1983). PGIP might, however, have an effect on a fruit's disease susceptibility, since fungi whose PGs were not successfully inhibited by pear PGIP produced lesions that expanded much faster than those that were inhibited. Pear PGIP was shown to be specific against pathogen PG enzymes, but could distinguish between PG secreted by different pathogenic fungi (*B. cinerea*, *Dothiorella gregaria* and *Penicillium expansum*).

2.7.13.3 Potato PGIP

A PGIP was purified from potato cv. Spunta leaves (Machinandiarena *et al.*, 2001). It had a molecular mass of 41 kDa, and was cell wall bound. It showed broad inhibitory activity against crude PG preparations from several fungi, including *A. niger*, *F. moniliforme* and *F. solani*. Exo-enzymes might have been present in the crude PG extracts from the other fungal cultures that were not inhibited by potato PGIP. Otherwise, they might have contained endoPGs that escaped potato PGIP inhibition. PGIP expression was induced in potato leaves by wounding, treatment with salicylic acid and incompatible interactions with *Phytophthora infestans*.

2.7.14 Role for PGIP in plant defence

Several features point to an important role of PGIP in defence against plant pathogens (De Lorenzo *et al.*, 2001). Just like other known defence genes, expression of PGIP is induced by stress- and pathogen-derived signals. They also share similarities in structure and specificity with *R* gene products. Synthesis and accumulation of PGIP is an active as well as constitutive defence mechanism in dicots (Bergmann *et al.*, 1994). In incompatible interactions between plant and fungi, PGIP mRNA accumulation was much more rapid and intense than in compatible interactions, and is correlated with the expression of the hypersensitive response (Nuss *et al.*, 1996). Circumstantial evidence for their role in defence is that the level of PGIP correlates with increased resistance to fungi, e.g. increased susceptibility of ripening fruit to fungal attack as PGIP levels declines (pear and raspberry: Abu-Goukh *et al.*, 1983b; Johnston *et al.*, 1993). In an experiment where tomato was transformed with pear *pgip*, grey mould symptoms were less on transgenic plants than on control untransformed plants (Powell *et al.*, 2000). PGIP attenuated the disease symptoms by modulating the fungal pathogen

maceration of plant tissues. PGIP therefore does not prevent establishment of the initial plant-pathogen interaction, but influences the expansion of the fungal mass.

The following points summarise the evidence that PGIP plays an important role in defence against fungal pathogens:

- PGs are produced early in a plant-fungus interaction and contribute to the expansion of the infection site (ten Have *et al.*, 1998);
- PGIPs inhibit some (but not all) fungal PGs *in vitro* (Yao *et al.*, 1995; Sharrock and Labavitch, 1994), especially those from pathogens that are least virulent on a plant, suggesting their *in planta* function (Abu-Goukh and Labavitch, 1983; Abu-Goukh *et al.*, 1983b);
- PGs that are inhibited by PGIP *in vitro* produce oligogalacturonides that induce plant defence responses (according to the hypothesis by Cervone *et al.*, 1989);
- PGIP contain the LRR structural motif that is also present in other disease resistance genes (De Lorenzo *et al.*, 1994; Stotz *et al.*, 1994).

2.7.15 Use of PGIP in creating disease resistant crops

Discoverers of PGIP in immature raspberry fruit proposed to use this gene in breeding programs to enhance the disease resistance of ripe fruits against *B. cinerea* (Williamson *et al.*, 1993). Pear PGIP was 20 times more active against PG from *B. cinerea* than tomato PGIP (Stotz *et al.*, 1994). Tomatoes were transformed with pear *pgip*, and it was still an active inhibitor of *B. cinerea* endoPGs (Powell *et al.*, 2000). It slowed the expansion of disease lesions and associated tissue maceration on infected transgenic tomato fruit and leaves, compared with infections of control plants. In contrast, bean *pgip-1* did not enhance the tomato's resistance against fungal infection (Desiderio *et al.*, 1997).

Different PGIPs expressed in the same plant may have different fungal PG-specificities. When individual bean PGIPs were tested *in vitro* against fungal PGs, they exhibited inhibiting abilities different from bulk bean PGIP (Desiderio *et al.*, 1997). The broad activity of bulk bean PGIP appear thus to be the result of the presence of different PGIP molecules with narrow specificities. It must therefore be taken into consideration that by expressing one PGIP in a transgenic plant, inhibiting activity against only a certain number of specific fungi will be obtained. If a general enhanced resistance to fungi needs to be obtained, more than one PGIP gene needs to be transformed.

2.8 *GST1* promoter of *Arabidopsis thaliana*

During this project, a construct was prepared containing the apple *pgip1* gene under control of the *gst1* promoter of *A. thaliana*. The aim is to drive pathogen-inducible expression of the apple *pgip1* gene in transgenic *A. thaliana*, and ultimately in other crops of importance.

The glutathione S-transferases (GSTs) are a family of enzymes that protect cellular macromolecules from various toxic xenobiotics (Yang *et al.*, 1998). GST catalyses the conjugation of thiol groups of the tripeptide glutathione (GSH) (γ-L-glutamyl-L-cysteinylglycine) to a variety of electrophilic substrates, e.g. in plants to detoxify herbicides (Itzhaki *et al.*, 1994). They function as homodimers or heterodimers of subunits. *GST* gene expression in various plants has been shown to respond to many inducers. These include ethylene, herbicide safeners, auxin, pathogen attack, salicylic acid, H_2O_2, dehydration, wounding, low temperature, high salt and DPE (diphenyl ether) herbicide treatment (Dudler *et al.*, 1991; Itzhaki *et al.*, 1994; Yang *et al.*, 1998). Examples of plants of which the *GST* gene was studied include wheat, carnation and *A. thaliana*.

The *GST* gene family of *A. thaliana* consists of six members, and is induced by the various means as listed before. The cDNA and corresponding genomic sequence of *GST1* have been cloned (Yang *et al.*, 1998, accession number Y11727). The gene has an open reading frame coding for 208 amino acids, and is interrupted by two introns. *GST1* cDNA was found to be induced by pathogen infection and dehydration, as well as wounding, high salt, low temperature and DPE herbicide treatment. Its promoter region contains motifs (ethylene responsive elements (ERE) and other motifs such as the TCA motif and the G-box) conserved amongst stress-inducible gene promoters.

Ethylene is a phytohormone that is involved in the regulation of plant growth and development and response to biotic and abiotic stresses. It influences plants by changing gene expression, through ethylene-responsive elements (EREs) conserved in the promoter regions of several genes. The ERE in *A. thaliana GST1* gene promoter (ATTTCAAA) is inversely repeated and found at positions -183 and -737 (Yang *et al.*, 1998). The ethylene responsive *cis*-sequence elements of the *GST* gene of carnation were studied and the same 8 bp sequence was found (Itzhaki *et al.*, 1994). Two putative TCA motifs are found at positions -845 and +32 of the *GST1* gene promoter of *A. thaliana*, indicating that gene expression may be regulated by salicylic acid. A G-box motif, conserved in plant genes associated with response to diverse environmental stresses, was found at position -369 (Yang *et al.*, 1998).

Another example that is evidence for the pathogen-inducibility of *GST* genes, is the putative *GST1* gene of wheat (*Triticum aestivum*). Its transcript was highly induced in leaves infected with the incompatible pathogen *Erysiphe graminis* f. sp. *hordei* (Dudler *et al.*, 1991). Pathogen-induced genes

usually code for products that are involved in host defences against pathogens. Why would *GST* be one of the host genes induced by pathogens, and how can it be involved in defence? In plants it is usually associated with detoxifying herbicides, but in animal systems it protects against oxidative tissue damage by de-toxifying products of membrane lipid peroxidation. Reactive oxygen species (ROS) and membrane lipid peroxidation are also known to occur in plants in response to tissue damage, elicitors and pathogen attack. Thus, GST may play a similar protective role in plants. It may be a member of a class of general stress response genes that is activated by many different stimuli.

The *gst1* gene of *A. thaliana* was used as a molecular marker for ROI accumulation since it is an antioxidant defence gene (Grant *et al.*, 2000). The WS *GST1* sequence (Grant *et al.*, 2000) exhibited 98% DNA sequence identity to the sequence from the Ler accession (Yang *et al.*, 1998), indicating that they may encode the same protein. The promoter of the *gst1* gene was characterised to be 909 bp, and the translated region 1092 bp with two introns of 92 and 110 bp each. The *gst1* promoter contains two EREs, a G-box and a TCA element at the 3' end, the same as was found by Yang *et al.* (1998). The presence of the TCA and ERE motifs is in accordance with previous observations that *GST1* is induced following treatment with ethylene and salicylic acid.

The nucleotide sequence of the *gst1* promoter obtained from G. Loake, University of Edinburgh, contained eight nucleotide differences and had 99% sequence identity to that from the published sequence (Yang *et al.*, 1998, accession number Y11727). The sequence of the *gst1* promoter obtained from G. Loake and the published sequence (Yang *et al.*, 1998) are aligned in Figure 2.2. Identical nucleotides are indicated by an asterisk (*). Both sequences were selected to start at a *Kpn*I restriction site, and included the ATG initiation codon (boxed) as well as a few nucleotides of the *gst1* gene.

The promoter consists of a 909 bp sequence upstream of the translation start site (boxed). The differences between the two sequences may be explained by the fact that the *gst1* promoter obtained from G. Loake was isolated from the ecotype Ws (Grant *et al.*, 2000) while the Y11727 accession was from the Landsberg ecotype (Yang *et al.*, 1998). The nucleotide differences did not affect any of the putative regulatory motifs. G. Loake's *gst1* promoter was shown to be active in directing the expression of a *luciferase* reporter enzyme, a molecular marker for ROI accumulation (Grant *et al.*, 2000).

Seq 1 G. Loake

Seq 2 Yang *et al.*, 1998

KpnI

```
Seq 1  GGTACCATAAGAAGAAGAATTAATCTTATAATCTTGTGATGTACTTTCGGTGATTTCTTTAAGTTTGATG   70
       ******************************************************************* ******
Seq 2  GGTACCATAAGAAGAAGAATTAATCTTATAATCTTGTGATGTACTTTCGGTGATTTCTTTAAGATTGATG   70
                                                                      ERE

Seq 1  TTATGAATAAAAGGCAAAGCTTTTGCAAAAATCACTCTTTTTTTTG-CCATAATGATTTCAAAATTCCAA  139
       ************************************ ******** *************************
Seq 2  TTATGAATAAAAGGCAAAGCTTTTGCAAAAATCACTCATTTTTTTGGCCATAATGATTTCAAAATTCCAA  140

Seq 1  AGAATAATATATACTTCAATATACATGTCACAAGATTAAATGTCAAAGTTGTTTATAATGAGCATTTTTG  209
       ********************************** ***********************************
Seq 2  AGAATAATATATACTTCAATATACATGTCACAAGCTTAAATGTCAAAGTTGTTTATAATGAGCATTTTTG  210

Seq 1  TGGATGTGGAAAATGTGAAAATTACTCTGTTCCTTTGAATGTTTCTATACGAAATTATAGTTAGGATTTA  279
       ********************************** ***********************************
Seq 2  TGGATGTGGAAAATGTGAAAATTACTCTGTTCCTTCGAATGTTTCTATACGAAATTATAGTTAGGATTTA  280

Seq 1  GTAATGTATTTCTCCATAATTATGCTAAATTTAGTTAGTTACTTCAATATGATTAAATACTTTATTGACC  349
       ************************************************************************
Seq 2  GTAATGTATTTCTCCATAATTATGCTAAATTTAGTTAGTTACTTCAATATGATTAAATACTTTATTGACC  350

Seq 1  CCAAATTTGTAATTGTACCAGATTGTCAAAAGTGTGAAACCAAATTTTCTTTTCTTTTGATATTGTTTTG  419
       ********************************************** ********************
Seq 2  CCAAATTTGTAATTGTACCAGATTGTCAAAAGTGTGAAACCAAATTTTCCTTTCTTTTGATATTGTTTTG  420

Seq 1  TTTCTTATTATCTCTCGTTCTATCGAATGATCTAACTCGAGCATCCAACGATCTAACTCGAGCATCTAAC  489
       ************************************************************************
Seq 2  TTTCTTATTATCTCTCGTTCTATCGAATGATCTAACTCGAGCATCCAACGATCTAACTCGAGCATCTAAC  490
          G-BOX
Seq 1  GATCCACGTGGACCCAACAACGTCGGTCAGAGTTTGACTAGTAGATGAAGGACTATTCTTGTGGTCGTTG  559
       ***********************************************************************
Seq 2  GATCCACGTGGACCCAACAACGTCGGTCAGAGTTTGACTAGTAGATGAAGGACTATTCTTGTGGTCGTTG  560

Seq 1  TCACGCGGTGGCTGATATTTTCTCTTATTTTTATTTTATTTTAACTATTTTTTACGTTATATTTAAGTCT  629
       ************************************************************************
Seq 2  TCACGCGGTGGCTGATATTTTCTCTTATTTTTATTTTATTTTAACTATTTTTTACGTTATATTTAAGTCT  630
                                                        ERE
Seq 1  TGAACCAATAGAAACGACGAATCATACATTACTCAGCTTGACTTTGAAATAATCCTATAACAAAAGAGCA  699
       ************************************************************************
Seq 2  TGAACCAATAGAAACGACGAATCATACATTACTCAGCTTGACTTTGAAATAATCCTATAACAAAAGAGCA  700
```

```
Seq 1   TTCCAAGAATTTTATCCAAAAAA-CAAAATAAAAAAGAGTATTCAAGCTTGGTGGCGCCGTTTGTTTTTG   769
        ********************** ******************************** **************
Seq 2   TTCCAAGAATTTTATCCAAAAAAACAAAATAAAAAAGAGTATTCAAGCTTGGTGGAGCCGTTTGTTTTTG   770
                                    TATA-box
Seq 1   GTTTATTCACTAAAGTTACTCTGTTTTAGTTGTATAAATACACACTCCCATTTGTGTATTTCTTTTCATC   838
        ******************************************************************************
Seq 2   GTTTATTCACTAAAGTTACTCTGTTTTAGTTGTATAAATACACACTCCCATTTGTGTATTTCTTTTCATC   840
                                                        TCA
Seq 1   AATCACAAAGATCTCTCTACTTCAATAAATCTCCACCTTACTTTAAGAACAAGAAAAACACAGTATTAAC   908
        ************************************************************************
Seq 2   AATCACAAAGATCTCTCTACTTCAATAAATCTCCACCTTACTTTAAGAACAAGAAAAACACAGTATTAAC   910

Seq 1   AAATGGCAGGAATCAAAGTTTTCGGTCACCCAGCTTCCACAGCCACTAGA                       958
        ***********************************************
Seq 2   AAATGGCAGGAATCAAAGTTTTCGGTCACCCAGCTTCCACAGCCACTAGA                       960
```

Figure 2.2 **Comparison of the nucleotide sequence of the *gst1* promoter from G. Loake with the published sequence (Yang *et al.*, 1998).** The ethylene responsive elements (ERE), G-box, TATA box and TCA element are indicated in red. The restriction site *Kpn*I is indicated in green. The start codon is boxed.

2.9 *Arabidopsis thaliana* as a model experimental plant

A. thaliana is a small diploid angiosperm of the mustard family (Cruciferae or Brassicaceae) that has become a model system for plant research. Its life cycle, from germination of seeds, formation of the rosette plant, bolting of the main stem, flowering and maturation of the first seeds, is completed in six weeks. This rapid growth cycle makes it very useful for research. Its flowers are 2 mm long, self-pollinated and develop into siliques that contain seeds of 0.5 mm. The rosette plant grows to 2 to 10 cm in diameter while mature plants are 15 to 20 cm in height. Several thousand seeds can be collected from a single plant (Meinke *et al.*, 1998).

The *Arabidopsis* Genome Initiative (AGI) was established in 1996 to facilitate coordinated sequencing of the *A. thaliana* genome, which was completed in 2000. The genome sequence of *A. thaliana* was published in *Nature* (chromosomes 2 and 4 in 1999 and chromosomes 1, 3 and 5 in 2000). It is extensively discussed in several articles (Goodman *et al.*, 1995; The Arabidopsis Genome Initiative, 2000). The original idea behind using *A. thaliana* as a model system was to help in the identification of related genes important in crop plants. The *A. thaliana* genome is small, only 120 mega bases, organised into five chromosomes and contains an estimated 26 000 genes. It contains a relatively

small amount of interspersed repetitive DNA, making sequencing of its genome a cost-effective method in identifying every gene in a representative flowering plant (Meinke *et al.*, 1998).

Mutants defective in almost every aspect of plant growth and development have been identified and studied by various research groups over the world. Random large-scale insertional mutagenesis by T-DNA and transposon insertion is used to create gene knockouts so that reverse genetic screens can be applied to deduce the functions of the sequenced genes (Parinov and Sundaresan, 2000). To name a few examples: mutants were used for dissecting the mode of ethylene action in plants (Guzmán and Ecker, 1990); defence-related mutants were used to dissect the plant defence response to pathogens (Ausubel *et al.*, 1995; Glazebrook and Ausubel, 1994; Xie *et al.*, 1998); and to study the effect of plant hormones and signalling molecules on pathogen-induced defence gene expression (Feys *et al.*, 1994; Penninckx *et al.*, 1996).

2.10 *Agrobacterium tumefaciens*-mediated plant transformation

2.10.1 *Agrobacterium tumefaciens* transformation system

Agrobacterium tumefaciens causes crown gall on dicotyledonous plants. It confers this tumorous phenotype by introducing a DNA segment into the plant cell and stably integrating it into the plant chromosome (Hooykaas and Schilperoort, 1992). This was viewed as a useful method of delivering genetic material into plants to create stable transgenics. The transferred DNA is called the T-DNA and is carried on the tumour-inducing plasmid (Ti-plasmid), together with the virulence (*vir*) region that provides the *trans*-acting factors for creating the T-DNA copy, structural elements for the transfer intermediate and components of the transfer apparatus. Only sequences flanked by the T-borders (25 bp direct repeats) are transferred to the plant cell. In wild-type *Agrobacterium*, the genes contained between the T-borders are enzymes for plant growth regulators and opines. In transformation vectors, these sequences are replaced with the gene(s) of interest.

A different method of plant transformation involves a particle gun. Small tungsten or gold particles are coated with DNA and shot directly into plant tissues. The DNA with the particle that reaches the nucleus may integrate into the genome and be expressed. It has the disadvantages of scrambling the DNA copies and integration of multiple DNA copies that may lead to rearrangement, recombination or silencing.

Using the *A. tumefaciens*-mediated transformation method, the copy number of the T-DNA in transformed plant lines is usually low, varying from one to a few copies (Hooykaas and Schilperoort, 1992). Rarely, lines with up to a dozen copies are found. Multiple copies may be located at different loci in the plant genome, or occur at a single locus as direct or inverted repeats. The complex details of the formation of the transfer DNA complex are reviewed by Zupan and Zambryski (1997). The

Agrobacterium system thus doesn't have the disadvantages of the particle gun system, probably due to the structure with which the T-complex is delivered into the plant cells.

2.10.2 Plant transformation vectors

Binary vectors used for plant transformation were created to simplify the cloning of exogenous DNA into the large Ti-plasmid of *A. tumefaciens* (An, 1986). The *vir* region was removed from the Ti-plasmid, to result in a binary vector that contains an artificial T-DNA and carrying only the *cis*-acting elements. The *cis*-elements include the T-DNA borders, selectable marker expressible in plants (e.g. kanamycin resistance), cloning sites and a wide host range replicon. The binary vector needs to be transformed into an *Agrobacterium* strain containing a helper Ti-plasmid with an intact *vir* region but lacking the T-region (Hooykaas and Schilperoort, 1992). The helper Ti-plasmid and the *A. tumefaciens* genome provides *in trans* the other functions required for plant transformation.

2.10.3 *Agrobacterium*-mediated transformation of *Arabidopsis thaliana* by the floral-dip method

The floral-dip method for *Agrobacterium*-mediated transformation of *A. thaliana* allows transformation of the plant without the need for tissue culture. The flowering plant is simply dipped into a solution of *Agrobacterium* and allowed to produce seed (Clough and Bent, 1998). The target for *Agrobacterium* transformation is the ovules of young flowers. To achieve efficient transformation, the *Agrobacterium* has to be delivered to the interior of the developing gynoecium before the locules close (Desfeux *et al.*, 2000). Transformants derived from seed from the same seedpod will contain independent T-DNA integration events. The *Agrobacterium*-treated T_0 plant is not treated with selection agents such as herbicide or antibiotics. Rather, the progeny seed is harvested and selection is applied to the resultant T_1 seedlings as they germinate. The T_1 transformants are found to be mostly hemizygous and there is an absence of homozygous self-fertilised offspring. Thus, the transformation event is thought to occur in the germ-line cells after divergence of female and male gametophyte cell lineages, or the T_1 embryo is transformed soon after fertilisation.

2.10.4 Selection of transgenic plants

During transformation, not all cells are transformed. Selecting for a plant selectable marker or a reporter gene can identify cells that have been transformed. Selection markers are based on the sensitivity of plant cells to antibiotics and herbicides (Hooykaas and Schilperoort, 1992). Expressing bacterial genes of detoxifying enzymes can make plants resistant to these compounds. For example, kanamycin resistance is mediated by neomycin phosphotransferase (NPTII), hygromycin resistance via the hygromycin phosphotransferase (HPT) gene and bialaphos (a herbicide) resistance via the *bar* gene from *Streptomyces hygroscopicus*.

Plant reporter genes can also be used to identify transformed cells, by linking the reporter gene to the gene of interest and screening for the expression of the reporter gene. Examples of reporter enzymes include luciferase, β-galactosidase and β-glucuronidase (Hooykaas and Schilperoort, 1992). The firefly luciferase enzyme generates emitted light when it hydrolyses its substrate luciferin. Since many plant tissues have endogenous β-galactosidase activity, β-glucuronidase is preferred as a reporter enzyme. β-glucuronidase enzyme activity can be measured quantitatively by fluorometrically detecting the umbelliferone that is released from umbelliferyl derivatives. It can also be determined histologically by staining transformed tissues. β-glucuronidase enzymatically converts 5-bromo-4-chloro-3-indolyl derivatives to indigo-blue compounds (Hooykaas and Schilperoort, 1992).

This chapter gave a brief overview of polygalacturonase-inhibiting proteins and how they play a role in the defence against phytopathogenic fungi. It also reviewed the specific interaction between a plant and fungus, namely *Verticillium*-wilt of potato. In the following chapters, a brief overview will be given on the topics covered and the methodology used in each.

CHAPTER 3

Cloning of the apple *pgip1* gene under control of the *gst1* promoter of *Arabidopsis thaliana*

3.1 Introduction

In an attempt to test the functionality and pathogen inducibility of the *gst1* promoter from *Arabidopsis thaliana* (L.) Heynh (Yang *et al.*, 1998; Grant *et al.*, 2000), a construct was prepared where it was cloned upstream of the apple *pgip1* gene. This gene codes for a protein product (apple PGIP1) that is able to inhibit the PGs secreted *in vitro* by the fungal pathogen *Verticillium dahliae*. It is reported that the *gst1* promoter is induced by various stress conditions (Yang *et al.*, 1998). The fungal inducibility of the *gst1* promoter would be valuable in the transgenic expression of antifungal resistance genes in crops of importance. The construct will first be tested in *A. thaliana*, a plant that can be rapidly transformed (Chapter 4). It is advantageous to use this model plant since transformants can be obtained without the methods of tissue culture regeneration, and the plant has a relatively short life-cycle (under optimal conditions, six weeks from germination to seed set).

The aim of this chapter was thus to prepare a plant transformation construct in which the apple *pgip1* gene is under control of the pathogen-inducible *gst1* promoter. This will be used in subsequent transformations of *A. thaliana* (Chapter 4).

Construction of plasmids for plant transformation

A. Apple *pgip1* gene under control of a constitutive promoter (CaMV e35S promoter)

Previously, a PGIP gene from apple was cloned at ARC-Roodeplaat using inverse polymerase chain reaction methods (Arendse *et al.*, 1999). The sequence obtained was identical to that published by Yao *et al.* (1999, accession number U77041). A chimeric e35S-*pgip1* construct for the constitutive overexpression of the apple *pgip1* in transgenic plants has been generated (Arendse and Berger, unpublished). The apple *pgip1* gene was amplified from genomic DNA by designing the appropriate primers, including restriction enzyme sites at the ends of the primers. The PCR fragment was directionally cloned into the pRTL2 vector (Cassidy and Nelson, 1995), placing the apple *pgip1* gene under control of the enhanced CaMV 35S promoter (e35S) (a standard dicot constitutive promoter) and the TEV leader (Tobacco Etch Virus leader element, which is a translational enhancer), followed by the CaMV 35S terminator. The cassette, containing the promoter, TEV leader, apple *pgip1* gene and terminator, was subcloned into the binary vector pCAMBIA2300 for plant transformation.

Constructs containing the cassette in either orientation were obtained. The construct pCAMBIA2300-apple*pgip*1B (**Appendix C**) was chosen as the constitutively expressed apple *pgip1* construct for transformation of *A. thaliana* (covered in Chapter 4).

B. Apple *pgip1* gene under control of a pathogen-inducible promoter (*gst1* promoter)

The preparation of a plant transformation vector containing the apple *pgip1* gene under control of the *gst1* promoter from *A. thaliana* required three steps. Firstly, the *gst1* promoter was amplified from a previous construct (p*GST1*-BluescriptSK$^-$) using the polymerase chain reaction, and cloned into the pMOSBlue blunt-end cloning plasmid vector. Secondly, the *gst1* promoter PCR fragment was subcloned into the plant transformation vector pCAMBIA2300. Lastly, the apple *pgip1* gene, as part of a cassette containing the TEV leader and CaMV 35S terminator, was cloned downstream of the *gst1* promoter in pCAMBIA2300. The resultant construct was called *GST1*prom-ap*pgip1*-pCAMBIA2300 (Figure 3.18).

The promoter consists of a 909 bp sequence upstream of the translation start site (Figure 2.2). A primer was designed to amplify the *gst1* promoter fragment from a previous construct (p*GST1*-BluescriptSK$^-$) using the polymerase chain reaction (PCR). The primer (GSTreverse) is complementary to a sequence at the 3' end of the *gst1* promoter, immediately upstream of the ATG start codon of the *gst1* gene (refer to Figure 3.1). The start codon needs to be excluded from the amplified promoter, to ensure correct translation initiation of the apple *pgip1* gene in the transgenic plant. A vector-specific primer was used as the second primer in the primer-pair needed for PCR. PCR exponentially amplifies selected regions of DNA using primers that anneal to denatured dsDNA. A thermostable DNA polymerase such as Taq DNA polymerase, isolated from *Thermus aquaticus*, is used to synthesise the new strands complementary to the denatured template. This polymerase enzyme remains active during the repeated cycles of denaturing of DNA hybrids, annealing of primer to template and extension of the primer complementary to the template.

Restriction enzyme recognition sites were engineered at the 3' end of the *gst1* promoter, to facilitate subsequent subcloning. It was incorporated into the amplification products by adding it to the 5' end of the promoter-specific primer, called GSTreverse. During the first cycle of PCR, a hybrid between template and a primer that is not completely complementary to it will form. After extension the primer will now be incorporated into a daughter molecule, which will serve as a template molecule during the next cycle. In this approach the primer sequence gets incorporated into the PCR amplification products.

This chapter reports on the cloning of the apple *pgip1* gene under control of the pathogen-inducible *gst1* promoter of *A. thaliana* into a plant transformation vector. Chapter 4 will continue with the

transformation of the model plant *A. thaliana* with the *gst1*- and e35S-promoter apple *pgip1* constructs. Molecular analysis of the transformants and PGIP expression studies will also be reported.

3.2 Materials and Methods

All chemicals and reagents used were either analytical or molecular biology grade. Buffers, solutions and media were all prepared using distilled water and were either autoclaved or filter-sterilised through 0.2 µm sterile syringe filters. All buffers, solutions and media used in this study are outlined in **Appendix A**. Ampicillin, calf thymus DNA, the Expand Long template PCR system, Hoechst 33258 DNA binding dye, restriction endonucleases, RNase A, T4 DNA ligase and X-gal (5-bromo-4-chloro-3-indolyl-β-galactoside) were obtained from Roche Diagnostics (Mannheim, Germany). Large scale (Qiagen Midi plasmid purification kit) and mini plasmid DNA preparation kits (Qiaprep and Qiafilter miniprep kits), as well as gel extraction kits (Qiaquick gel extraction kit) were from Qiagen, Germany. Plasmid maps were drawn using Vector NT. Sequence data can be entered and a map generated to scale, indicating restriction enzyme sites and other features at the correct positions.

3.2.1 Plasmid DNA isolation

Isolation of large amounts of plasmid DNA from *Escherichia coli* was performed using the Qiagen Midi plasmid purification kit. Mini preparations of plasmid DNA were performed using the Qiaprep and Qiafilter miniprep kits. Purification using the Qiagen kits relies on the alkaline lysis principle, and the DNA is purified by selective absorption to a silica column in the presence of a chaotropic salt. Contaminants are washed away and the DNA is eluted by a low salt buffer (10 mM Tris, pH 8.5).

For restriction digestion screening of possible recombinants, mini preparations of plasmid DNA was done by the alkaline lysis method as in Sambrook *et al.* (1989) with modifications. A single bacterial colony was grown overnight at 37°C with shaking in 5 ml LB medium containing the appropriate antibiotic (100 µg/ml ampicillin for pMOSBlue constructs and 100 µg/ml kanamycin for pCAMBIA2300 constructs). Cells from 3 ml of overnight culture were collected by centrifugation at 6500×g for 10 min at 4°C. The supernatant was removed and the pellet resuspended in 200 µl Solution I (**Appendix A**) by pipetting or vortexing. Five microlitres of 10 mg/ml RNase A was added and the sample incubated for 5 min at room temperature. The cells were lysed by the addition of 400 µl Solution II (**Appendix A**) and incubated on ice for 5 min. The sample was neutralised by the addition of 300 µl Solution III (**Appendix A**) and incubated on ice for 5 min. The debris was pelleted by centrifugation at 6500×g for 20 min at 4°C. Seven hundred and fifty microlitres of the supernatant was recovered and added to 750 µl isopropanol in a clean tube. The precipitated DNA was pelleted by centrifugation at 6500×g for 15 min at 4°C. The DNA pellet was resuspended in 540 µl dH_2O and 60 µl 5 M $NaClO_4$. Six hundred microlitres isopropanol was added and the mixture was incubated at room temperature for 10 min. The DNA was pelleted by centrifugation at 6500×g for 15 min at 4°C.

washed with 70% ethanol and re-pelleted. After air-drying, the DNA was resuspended in 30 µl 1× TE (pH 8.0) or 10 mM Tris (pH 8.5) by shaking at 37°C for a few minutes. Typically the yield was such that 3 µl was sufficient to be used in restriction digests for screening.

The DNA concentration was determined on a Sequoia-Turner 450 fluorometer (Sequoia-Turner Corporation) according to the manufacturer's instructions. The fluorometer was calibrated with 1× TNE buffer, pH 7.4, containing 0.2 µg/ml Hoechst 33258 DNA binding dye. Calf thymus DNA was used as a DNA standard.

3.2.2 Polymerase Chain Reaction

3.2.2.1 PCR amplification using Taq DNA polymerase (Primer testing)

PCR was conducted in 0.2 ml thin-walled tubes in a MJ Research Minicycler (MJ Research Inc.) with an internal temperature probe. Taq DNA polymerase and 10× reaction buffer [100 mM Tris-HCl, pH 9.0, 500 mM KCl, 1% Triton X-100] were from Promega. The reaction mixture, in a total volume of 10 µl, contained 0.5U Taq DNA polymerase, 1× reaction buffer, 200 µM of each dNTP, 1.5 mM $MgCl_2$, 0.5 µM of each primer (GSTreverse and SK primer, **Appendix B**) and 0.5 to 10 ng p*GST1*-BluescriptSK$^-$ plasmid DNA as template. The reaction volume was made up to 10 µl with sterile dH_2O and overlaid with mineral oil to prevent evaporation.

The PCR cycling conditions included an initial denaturation step of 94°C for 2 minutes. This was followed by 35 cycles of denaturation at 94°C for 90 s, annealing at 45°C for 90 s and extension at 72°C for 90 s. A final extension step of 7 min at 72°C was included.

3.2.2.2 PCR amplification of the *gst1* promoter using the Expand Long template PCR system

The Expand Long template PCR system was used to amplify the *gst1* promoter for subcloning. The reaction mixture, in a total volume of 50 µl, contained 1× Expand buffer 1 (with a final concentration of 1.75 mM $MgCl_2$), 1.5 µl Expand Long template DNA polymerase mix, 200 µM of each dNTP, 0.5 µM of each primer (GSTreverse and SK primer, **Appendix B**) and 10 ng p*GST1*-BluescriptSK$^-$ plasmid DNA as template. The reaction was duplicated to yield enough PCR products for subsequent cloning. The buffer and polymerase was added to the rest of the components just before the start of the PCR to minimise non-specific amplification. The reaction mixtures were overlaid with mineral oil to prevent evaporation. The same PCR cycling conditions were used as with the Taq DNA polymerase PCR.

3.2.3 Recovery of DNA fragments from agarose gels

DNA fragments from PCR or restriction digestions were separated by electrophoresis through a 1% (w/v) agarose/ 0.5× TAE (pH 8.0) gel containing 0.06 µg/ml ethidium bromide. The DNA was visualised under ultraviolet light and the fragment of interest excised from the gel. The weight of the gel slice was determined and the DNA eluted from the gel using the QIAquick gel extraction kit. The concentration of the eluted fragment was determined by agarose gel electrophoresis alongside a lambda DNA standard series ranging from 5 to 80 ng, and comparing the intensities of the bands to the standards.

3.2.4 pMOSBlue blunt-end ligation

The *gst1* promoter PCR product, that was extracted from the agarose gel, was cloned into the blunt-ended vector pMOSBlue. Ligation reactions were set up as set out in the pMOSBlue blunt-ended cloning kit instruction manual (AEC-Amersham, Little Chalfont, UK). The PCR product was first blunted and phosphorylated in a phosphokinase (pk) reaction containing 1× pk buffer, 5 mM DTT, 1 µl pk enzyme mix, 40 ng PCR product and dH_2O to a final volume of 10 µl. This reaction was incubated at 22°C for 30 min after which the enzymes were heat inactivated at 75°C for 10 min. Complete inactivation of the kinase enzyme is essential to avoid vector phosphorylation, which results in high vector religated background. The reaction was chilled on ice for 2 min to avoid heat inactivation of the ligase enzyme used in the subsequent step. The product of the kinase reaction was then ligated into the blunt dephosphorylated pMOSBlue vector. Ligation was set at a 1:2.5 vector-to-insert ratio. Fifty nanogram pMOSBlue vector (kit) and 1 µl T4 DNA ligase (4 Weiss units) was added to the 10 µl pk product and incubated at 22°C overnight.

3.2.5 Ligation

Vector and insert fragments were recovered from agarose gels and their concentrations determined as described before. Ligation was set at various insert-to-vector ratios and vector DNA quantities. Ligation reactions were prepared containing the appropriate quantities of vector and insert, 1× ligase buffer [66 mM Tris-HCl, 5 mM $MgCl_2$, 1 mM dithioerythritol, 1 mM ATP, pH 7.5], an additional 5 mM dATP and 5U T4 DNA ligase. The reaction volume was made up to 10 µl with sterile dH_2O. A vector-religated control was included in which insert was omitted to test the ability of vector to re-ligate. The ligation reactions were incubated at 16°C overnight after which they were stored at -20°C until transformation.

3.2.6 Preparation of competent *E. coli* DH5α cells

Competent *E. coli* DH5α cells were prepared using a method involving $CaCl_2$ and $MnCl_2$. The genetic background of these cells is *supE44* Δ(*lac*U169 (Φ80d*lac*ZΔM15)*hsd*R17 *rec*A1 *end*A1 *gyr*A96 *thi*-1 *rel*A1 (Hanahan, 1983).

A single colony of DH5α *E. coli* cells was inoculated into 25 ml of 2× LB medium (containing 0.2% sterile glucose). It was grown overnight at 37°C with shaking. A hundred millilitres of preheated 2× LB medium was inoculated with 1 ml of overnight culture, and grown for 1 hour at 37°C with shaking. One millilitre of sterile 2 M $MgCl_2$ was added and the culture incubated for another 35 minutes at 37°C with shaking. The culture was quick cooled on ice-water, and left on ice for 1 h 30 min. The cells were pelleted in 40 ml sterile centrifuge tubes by centrifuging at 2000×g (Beckman rotor JA-20) for 5 min at 4°C. The supernatant was poured off and the pellets resuspended in 50 ml of ice-cold Ca^{2+} /Mn^{2+} solution (40 mM NaAc, 100 mM $CaCl_2$, 70 mM $MnCl_2.4H_2O$; pH 5.5 and filter-sterilised). The suspension was left on ice for 1 h 30 min, after which the tubes were centrifuged at 700×g (Beckman rotor JA-20) for 5 min at 4°C. The cell pellets were resuspended in 3 ml ice-cold Ca^{2+} /Mn^{2+} solution containing 15% (v/v) sterile glycerol. The cells were kept on ice at all times. The resuspended cells were pooled and aliquotted in 100 µl quantities into pre-chilled sterile 1.5 ml Eppendorf tubes. They were frozen in liquid nitrogen and stored at -70°C. The competency of the prepared cells was tested by transformation with 1 ng pBluescript SK^+ plasmid DNA and plating onto LB-Agar plates containing 100 µg/ml ampicillin.

3.2.7 Transformation of ligation reaction into *E. coli* DH5α

Ligated plasmids were transformed into competent *E. coli* DH5α cells, prepared by the Ca^{2+} /Mn^{2+} method. The 10 µl ligation mixture was added to 100 µl competent DH5α cells and incubated on ice for 30 min. The cells were heat-shocked at 37°C for 1 minute and incubated on ice for 5 minutes. One millilitre of LB medium was added to the cells, after which they were incubated at 37°C for 1 hour with shaking. The cells were pelleted by centrifugation at 700×g for 6 min, and 600 µl of the supernatant was discarded. The cells were resuspended in the remaining supernatant and plated out in two 250 µl aliquots onto LB agar plates containing 100 µg/ml of the appropriate antibiotic. Ampicillin was used for pMOSBlue constructs and kanamycin for pCAMBIA2300 constructs. For blue/white selection, the agar plate was spread with 35 µl of 50 mg/ml X-gal in DMF and left to dry. The agar plates were incubated inverted at 37°C overnight. A positive transformation control in which 100 µl competent cells were transformed with 1 ng of pBluescript SK^+ or 50 ng pCAMBIA2300 plasmid DNA was included to determine the transformation efficiency of the competent cells.

3.2.8 Restriction enzyme digestion

3.2.8.1 Screening of putative transformants by restriction enzyme digestion

Mini preparations of plasmid DNA were performed and analysed by restriction enzyme digestions, as described by Sambrook *et al.* (1989). Plasmid DNA was digested overnight at 37°C in 10 µl reactions containing the appropriate restriction enzyme buffer. Restriction digestion products were analysed by agarose electrophoresis on a 1% (w/v) agarose/ 0.5× TAE (pH 8.0) gel containing 0.06 µg/ml ethidium bromide.

3.2.8.2 Preparation of insert and vector fragments by restriction enzyme digestion

Large-scale restriction enzyme digestions were performed to prepare vector and insert fragments for ligation. Reactions contained the applicable plasmid DNA, the appropriate 1× restriction enzyme buffer and typically 40U of restriction enzyme in a final volume of 50 µl. For the *Kpn*I restriction enzyme 0.01% BSA was included in the reaction. The reaction mixture was incubated at 37°C overnight, after which fragments were separated by electrophoresis on a 1% agarose gel. The applicable fragment was recovered from the gel as described before.

3.2.8.3 Preparation of the apple *pgip1* cassette fragment by partial restriction enzyme digestion

The apple *pgip1* cassette (TEV leader, apple *pgip1* gene and CaMV 35S terminator) was excised from a previous construct, pAppRTL2 (**Appendix C**: Arendse and Berger, unpublished), using *Xho*I and *Pst*I restriction enzyme digestion. The cassette contains an internal *Pst*I site, so to release the full-length cassette (1379 bp), a partial digestion had to be performed after complete digestion by *Xho*I has taken place.

3.2.8.3.1 Complete *Xho*I restriction digestion

Eight microgram pAppRTL2 plasmid DNA was completely digested with *Xho*I. Four replicate reactions were set up to generate enough linearised plasmid for subsequent partial *Pst*I digestions and cloning. After overnight digestion, the reactions were pooled and the linearised vector precipitated. This was done to separate the DNA from the restriction digestion buffer that will interfere with the subsequent partial digests. Precipitation was performed with $1/10^{th}$ volume 3 M NaAc, pH 5.5 and 2.5 volumes 100% ethanol. The sample was incubated at -20°C for 1 hour, after which the DNA was pelleted by centrifugation at 6500×g for 30 minutes at 4°C. The pellet was washed with 70% ethanol, air-dried and dissolved in an appropriate volume of 10 mM Tris (pH 8.5) to give an expected concentration of 450 ng/µl (if 100% recovery after precipitation is assumed).

3.2.8.3.2 Partial *Pst*I restriction digestion

Partial digests were carried out on the basis that limiting the concentration of Mg^{2+} in the reaction buffer can control restriction digestions. The $MgCl_2$ concentration was optimised to determine which concentration gave the highest yield of the 1379 bp cassette. Restriction digestion buffers were prepared which varied only in their $MgCl_2$ concentration [50 mM Tris-HCl, 100 mM NaCl, pH 7.5 with 0, 1 or 10 mM $MgCl_2$]. Four hundred and fifty nanogram samples of *Xho*I-linearised pAppRTL2 plasmid were digested with 0.5U *Pst*I in restriction buffers containing the varying concentrations of $MgCl_2$. The reactions were incubated at 37°C for varying amounts of time (20, 40 and 60 min).

The 450 ng digestion reactions were scaled up to 5.5 µg of linearised pAppRTL2 plasmid for cloning. It was digested with 5.5U *Pst*I in the appropriate buffer (of which the $MgCl_2$ concentration has been experimentally determined to be 1 mM $MgCl_2$) for 20 min at 37°C. The partial digestion products were separated on a 1% agarose gel until the fragments were sufficiently separated. The 1379 bp fragment was recovered from the gel.

3.2.9 Nucleotide sequencing

Plasmid DNA was isolated using the Qiagen Midi plasmid purification kit or Qiafilter miniprep kit. The inserts in the recombinant plasmids were subjected to nucleotide sequencing at the University of Stellenbosch, Department of Genetics. The fragment is amplified linearly in the presence of fluorescently labelled dideoxynucleotides (ddNTPs) (Old and Primrose, 1994). It is based on Sanger's method of sequencing, in which a single-strand primer anneals to a single-strand DNA template, a polymerase synthesises the complementary DNA, and at random positions a ddNTP is incorporated to result in the termination of chain lengthening. A thermostable DNA polymerase isolated from *Thermus aquaticus* is used since it lacks 5`-3` nuclease activity, has a reduced discrimination against fluorescently labelled ddNTPs, and permits a longer length of read. Fragments of every possible length are generated, each ending at the nucleotide incorporated as the ddNTP and labelled by one of the four fluorescent dyes. Each type of ddNTP is labelled with a different coloured dye, and the whole enzymatic reaction is done in the same tube. The labelled extension products are detected in a single lane during electrophoresis on a denaturing polyacrylamide gel, based on the different fluorescence emission properties of the four dyes. The sequence data is captured during the gel electrophoresis run by a focused laser beam. The individual migration fragments that pass a certain point fluoresce at different wavelengths for each of the four dyes. The information is stored electronically, and the DNA sequence output is presented in the form of a dye-specific intensity profile. The nucleotide sequences of the inserts of the recombinant constructs can be inferred from the intensity profiles. Sequence analysis and alignments were done using the computer software *Genepro Version 6.1* (Riverside Scientific Enterprises).

3.3 Results

3.3.1 Subcloning of the *gst1* promoter into pMOSBlue

This section describes the isolation of the appropriate part of the *gst1* promoter and subcloning into pMOSBlue. It is followed by the section in which the *gst1* promoter is subcloned into the plant transformation vector pCAMBIA2300.

3.3.1.1 Primer design and PCR of the *gst1* promoter

The nucleotide sequence information of the *gst1* promoter cloned into pBluescript SK$^-$ (acquired from G. Loake, University of Edinburgh) was available and confirmed by sequencing. The insert cloned into pBluescript SK$^-$ had a *Kpn*I site at the 5' end and an *Eco*RI site at the 3' end. The insert was sequenced from both sides using the T7 and T3 primers (see **Appendix B**). The *gst1* promoter is 909 bp in length. The nucleotide sequence of a section of p*GST1*-BluescriptSK$^-$ is presented in Figure 3.1.

This known sequence enabled the design of a primer complementary to a sequence at the 3' end of the *gst1* promoter, immediately upstream of the ATG start codon of the *gst1* gene. This primer was called GSTreverse, it was designed using the computer program *Primer Designer Version 3.0* (Scientific and Educational Software) and synthesised by Genosys Products. Characteristics of the primer are shown in **Appendix B.** The restriction enzyme sites *Sal*I and *Pst*I was incorporated to the 5' end of the primer to facilitate subcloning of the *gst1* promoter and subsequent cloning of the apple *pgip1* cassette downstream of the *gst1* promoter. The annealing sites of the GSTreverse primer and other primers are indicated in Figure 3.1.

To perform the polymerase chain reaction, a second primer at the 5' end of the *gst1* promoter was needed. Because the *gst1* promoter fragment in pBluescript SK$^-$ is flanked on the 5' end with a *Kpn*I restriction site, it can be used later in the subcloning of the *gst1* promoter PCR product. Therefore, a sequence specific primer was not necessary and a pBluescript SK$^-$ vector specific primer was used as the 5' primer of the primer-pair. Either the T3 or SK primer could have been used in combination with GSTreverse, but the latter was chosen. An amplification product of 1008 bp was expected.

```
                T3  primer                                  SacI                    NotI
   1  GCTCGAAATT AACCCTCACT AAAGGGAACA AAAGCTGGAG CTCCACCGCG GTGGCGGCCG

       SK primer     BamHI  SmaI          EcoRI EcoRV          ClaI    SalI
  61  CTCTAGAACT AGTGGATCCC CCGGGCTGCA GGAATTCGAT ATCAAGCTTA TCGATACCGT

      SalI                  KpnI
 121  CGACCTCGAG GGGGGGCCCG GTACCATAAG AAGAAGAATT AATCTTATAA TCTTGTGATG
 181  TACTTTCGGT GATTTCTTTA AGTTTGATGT TATGAATAAA AGGCAAAGCT TTTGCAAAAA
 241  TCACTCTTTT TTTTGCCATA ATGATTTCAA AATTCCAAAG AATAATATAT ACTTCAATAT
 301  ACATGTCACA AGATTAAATG TCAAAGTTGT TTATAATGAG CATTTTTGTG GATGTGGAAA
 361  ATGTGAAAAT TACTCTGTTC CTTTGAATGT TTCTATACGA AATTATAGTT AGGATTTAGT
 421  AATGTATTTC TCCATAATTA TGCTAAATTT AGTTAGTTAC TTCAATATGA TTAAATACTT
 481  TATTGACCCC AAATTTGTAA TTGTACCAGA TTGTCAAAAG TGTGAAACCA AATTTTCTTT
 541  TCTTTTGATA TTGTTTTGTT TCTTATTATC TCTCGTTCTA TCGAATGATC TAACTCGAGC
 601  ATCCAACGAT CTAACTCGAG CATCTAACGA TCCACGTGGA CCCAACAACG TCGGTCAGAG
 661  TTTGACTAGT AGATGAAGGA CTATTCTTGT GGTCGTTGTC ACGCGGTGGC TGATATTTTC
 721  TCTTATTTTT ATTTTATTTT AACTATTTTT TACGTTATAT TTAAGTCTTG AACCAATAGA
 781  AACGACGAAT CATACATTAC TCAGCTTGAC TTTGAAATAA TCCTATAACA AAAGAGCATT
 841  CCAAGAATTT TATCCAAAAA ACAAAATAAA AAAGAGTATT CAAGCTTGGT GGCGCCGTTT
 901  GTTTTTGGTT TATTCACTAA AGTTACTCTG TTTTAGTTGT ATAAATACAC ACTCCCATTT
 961  GTGTATTTCT TTTCATCAAT CACAAAGATC TCTCTACTTC AATAAATCTC CACCTTACTT
 021  TAAGAACAAG AAAAACACAG TATTAACAAA TGGCAGGAAT CAAAGTTTTC GGTCACCCAG
                GSTreverse  primer
                                                          SalI    EcoRI
1081  CTTCCACAGC CACTAGAAGA GTTCTCATCG CTCTTCACGA GAAGAATGTC GACTTTGAAT

        EcoRV        ClaI      SalI                  KpnI
1141  TCGATATCAA GCTTATCGAT ACCGTCGACC TCGAGGGGGG GCCCGGTACC CAATTCGCCC
1201  TATAGTGAGT CGTATTACAA TTCACTGGCC GTCGTTTTAC AACGTCGTGA CTGGGAAAAC
1261  CCTGGCGTTA CCCAACTTAA TCGCCTTGCA GCACATCCCC CTTTCGCCAG CTGGCGTAAT
        T7 primer
```

Figure 3.1 **Nucleotide sequence of a section of p*GST1*-BluescriptSK⁻ showing restriction enzyme sites and primer annealing sites.** The TATA-box of the *gst1* promoter is indicated in blue, restriction enzyme sites in green and red, the initiator codon of the *gst1* gene is boxed and primer-annealing sequences are underlined.

3.3.1.2 PCR amplification of the *gst1* promoter using Taq DNA polymerase (Primer testing)

The PCR using GSTreverse and SK primers was first optimised with Taq DNA polymerase (Promega) before using a proofreading DNA polymerase enzyme to generate high fidelity products for cloning. The annealing temperature for efficient amplification was determined experimentally to be 45°C (data not shown). PCR was performed in which all conditions were kept constant, except the amount of p*GST1*-BluescriptSK⁻ template was varied. The *gst1* promoter is 909 bp in length (Grant *et al.*, 2000), but with the 19 bp 5'-extension incorporated by the GSTreverse primer, and the 80 bp upstream of the promoter insert amplified by the SK primer, a product of 1008 bp was expected. An amplification

product of approximately 900 bp was obtained for all reactions (Figure 3.2, lanes 1 to 5). The fragment was smaller than expected but nucleotide sequencing later showed it to be correct. A possible reason for this is overloading of the gel (see Discussion). Figure 3.2 shows that a minimum of 0.5 ng template plasmid DNA is sufficient for amplification of the *gst1* promoter using the SK and GSTreverse primers (Figure 3.2, lane 1). As expected, the negative control, containing no template plasmid DNA, did not give any amplification products (Figure 3.2, lane 6).

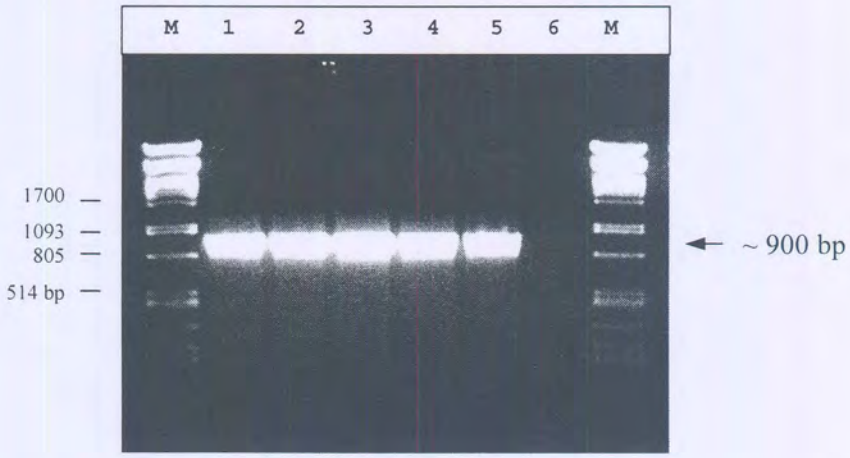


Figure 3.2 **PCR of the *gst1* promoter using Taq DNA polymerase.** PCR was performed with the SK and GSTreverse primers containing varying amounts p*GST1*-BluescriptSK⁻ plasmid as template. M: λDNA/ *Pst*I marker; lanes 1 to 5: 0.5, 1, 2, 5 and 10 ng plasmid as template, respectively; lane 6: negative control containing dH_2O.

3.3.1.3 PCR amplification of the *gst1* promoter using the Expand Long template PCR system

The Expand Long template PCR system was used to accurately amplify the *gst1* promoter for subcloning. It is a mixture of thermostable Taq DNA polymerase and a proofreading polymerase for high fidelity products. The proofreading polymerase is Tgo DNA polymerase, isolated from *Thermococcus gorgonarius*. It is a highly processive 5'-3' DNA polymerase, thermostable and has a better proofreading activity and a higher specificity than Pwo DNA Polymerase. It generates PCR fragments with a mixture of blunt and 3' single A overhangs. PCR using this system also yielded an amplification product of approximately 900 bp (Figure 3.3, lanes 1 and 2).

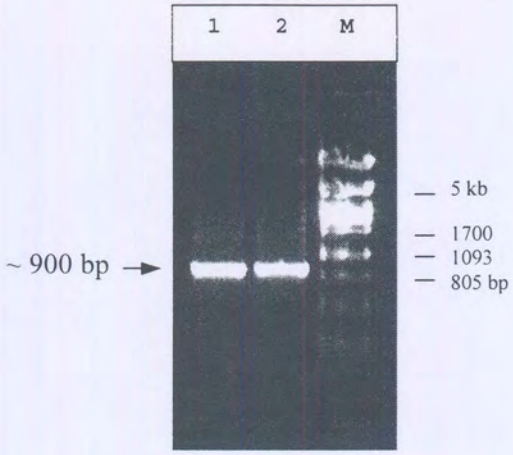


Figure 3.3 **PCR of the *gst1* promoter with the Expand Long template PCR system.** PCR was performed with the SK and GSTreverse primers containing 10 ng p*GST1*-BluescriptSK⁻ plasmid as template. Lanes 1 and 2: 2 µl samples of duplicate 50 µl Expand Long template PCR reactions: M: λDNA/ *Pst*I marker.

3.3.1.4 Recovery of the *gst1* promoter PCR product from the agarose gel

PCR products of the Expand Long template PCR system were pooled and purified from the contaminating PCR components by separating it with agarose gel electrophoresis and eluting the fragment from the gel. The concentration of the eluted PCR fragment was determined to be 40 ng/µl. Figure 3.4 shows the eluted PCR fragment. A faint band of a smaller size fragment (~700 bp) was observed together with the purified PCR fragment of ~900 bp.

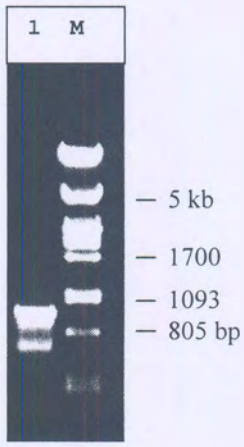


Figure 3.4 ***gst1* promoter PCR fragment eluted from an agarose gel.** Lane 1: 2 µl sample of the 30 µl eluted fragment: M: λDNA/ *Pst*I marker.

3.3.1.5 Ligation of the *gst1* promoter PCR fragment into pMOSBlue

Since ligation of the *gst1* promoter PCR product into the T-A cloning vector pGEM T-Easy (Promega) was not successful (results not shown), it was ligated into the blunt-ended cloning vector pMOSBlue (AEC-Amersham, Little Chalfont, UK).

3.3.1.6 Screening of transformants by miniprep and restriction enzyme digestions

Putative positive clones were identified using blue/white colony selection based on the α-complementation system. Ten white colonies were obtained when the ligation products of the *gst1* promoter PCR fragment and the pMOSBlue vector were transformed into competent DH5α *E. coli*. They were screened for the presence of recombinant constructs containing the *gst1* promoter PCR product insert by extracting plasmid DNA and digesting 3 µl with the restriction enzymes *Kpn*I and *Pst*I.

Of the ten colonies, only three contained putative recombinant plasmids with inserts that could be excised with *Kpn*I and *Pst*I (Figure 3.5, lanes 1, 6 and 7). Since *Kpn*I and *Pst*I sites flank the blunt *Eco*RV cloning site of the pMOSBlue vector, even an insert that is not the *gst1* promoter PCR product will be excised. The expected size of the correct insert excised with *Kpn*I and *Pst*I is 919 bp. While the insert from putative recombinant plasmid clone 1 was too large to be correct (larger than 1200 bp; Figure 3.5, lane 1), the other two clones contained inserts of approximately the correct size. These were clones 6 and 9, with inserts approximately 900 bp in size (Figure 3.5, lanes 6 and 7; and Figure 3.6, lanes 3 and 4).

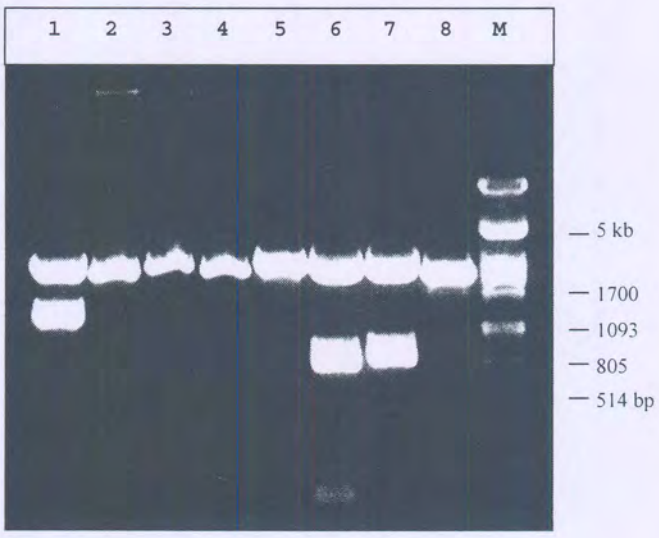


Figure 3.5 **Restriction enzyme screening of eight putative recombinant *GST1*prom-pMOSBlue clones.** Lanes 1 to 8: miniprep plasmid DNA isolated from *E. coli* colonies transformed with putative *GST1*prom-pMOSBlue clones 1, 2, 3, 4, 5, 6, 9 and 11, respectively, and digested with *Kpn*I and *Pst*I: M: λDNA/ *Pst*I marker.

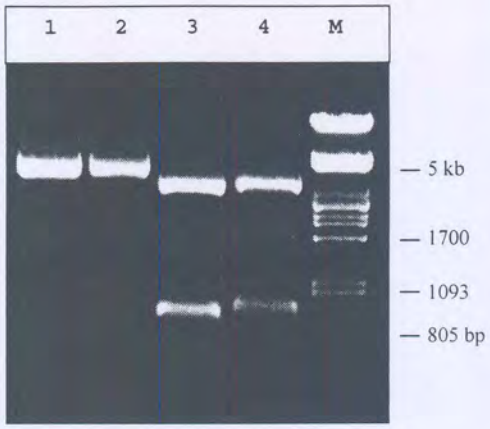


Figure 3.6 **Restriction digestion of putative recombinant *GST1*prom-pMOSBlue clones 6 and 9 plasmid DNA.** Lanes 1 and 2: undigested plasmid DNA from clone 6 and 9, respectively; lane 3 and 4: plasmid DNA of clone 6 and 9, respectively, digested with *Kpn*I and *Pst*I: M: λDNA/ *Pst*I marker.

3.3.1.7 Nucleotide sequencing of *GST1*prom-pMOSBlue clones 6 and 9

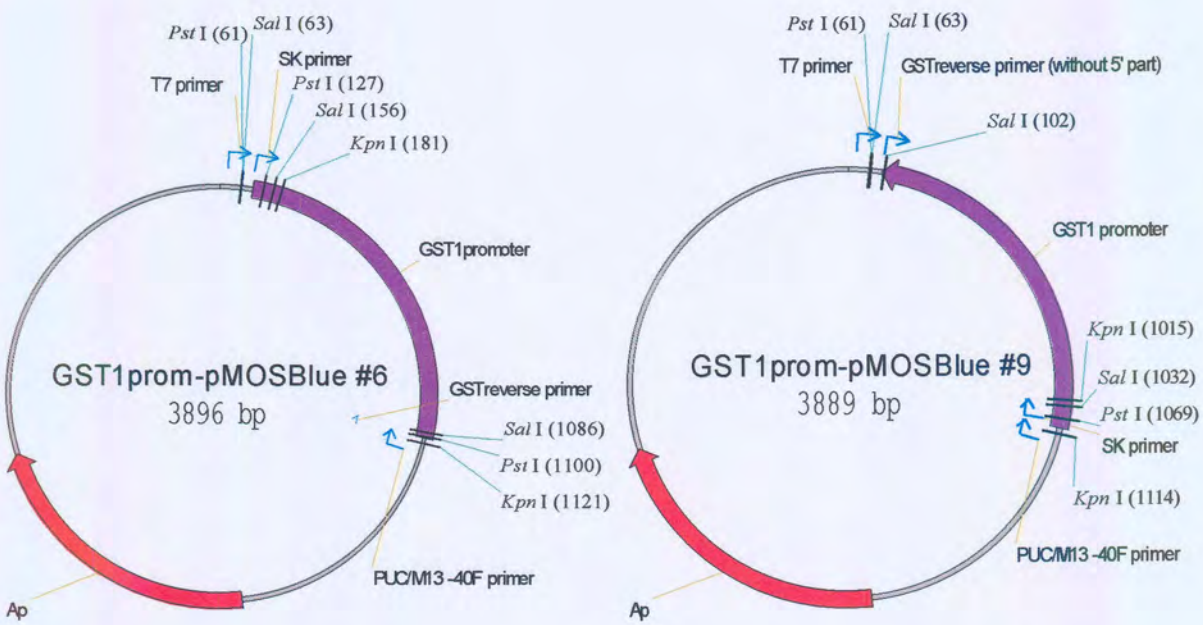
Colonies 6 and 9 were deduced to be likely candidates to contain recombinant *GST1*prom-pMOSBlue constructs. Recombinant plasmid DNA was sent for nucleotide sequencing of the insert using the T7 and PUC / M13-40F primers (**Appendix B**). From the agarose gel results (Figure 3.5 and Figure 3.6), the insert from clone 9 seemed to be slightly larger than the insert from clone 6. The nucleotide sequencing results explained why.

Figure 3.7 is a graphical representation of *GST1*prom-pMOSBlue clones 6 and 9 recombinant plasmids. Nucleotide sequencing of the inserts revealed that in clone 6 the complete GSTreverse primer sequence had been incorporated at the 3` end of the PCR fragment, resulting in a *Pst*I recognition site immediately downstream of the *gst1* promoter. The binding site of the GSTreverse primer is located at the 3` end of the *gst1* promoter, immediately upstream of the ATG start codon of the *A. thaliana gst1* gene. The rest of the *gst1* promoter sequence, up to the binding site of the GSTreverse primer, was identical to the original sequence in p*GST1*-BluescriptSK-. The *gst1* promoter was cloned into pMOSBlue with the same orientation as the ampicillin resistance gene (Ap). The recombinant *GST1*prom-pMOSBlue clone 6 plasmid was chosen for further experiments.

From the nucleotide sequencing results, it was learned that the 5` part (5`-AAACTGCA-3`) of the GSTreverse primer was absent in the sequence of the insert in clone 9. This resulted in the loss of the *Pst*I site incorporated with the GSTreverse primer to the 3` end of the *gst1* promoter. The insert is still excised from clone 9 since *Kpn*I and *Pst*I sites flank the blunt *Eco*RV cloning site in the multiple cloning site of the pMOSBlue vector. *Kpn*I and *Pst*I digestion of clone 9 therefore resulted in a fragment 35 bp larger than expected (954 bp instead of 919 bp). Thirty-five basepairs is the distance

between the *Pst*I site found in the MCS of pMOSBlue and where the *Pst*I site would have been at the 3' end of the *gst1* promoter if it had been incorporated with the GSTreverse primer. The loss of the *Pst*I site of the GSTreverse primer explains the restriction fragment of clone 9 being larger than that of clone 6, as was observed in Figure 3.5 and Figure 3.6.

In clone 9 the *gst1* promoter was cloned into pMOSBlue with the opposite orientation as the ampicillin resistance gene (Ap). Clone 9 also contained an insertion mutation of an Adenine between the 5' end of the *gst1* promoter PCR product and the *Eco*RV blunt cloning site. This may be the result of the Taq DNA polymerase in the Expand enzyme mix, adding single 3' A overhangs to the PCR product during amplification.



***GST1*prom-pMOSBlue #6** ***GST1*prom-pMOSBlue #9**

Figure 3.7 **Graphical representation of *GST1*prom-pMOSBlue clones 6 and 9 recombinant plasmids.** The results obtained for the nucleotide sequencing of the inserts, as well as primer binding sites and selected restriction enzyme recognition sites, are indicated on the maps.

3.3.2 Subcloning of the *gst1* promoter into pCAMBIA2300

*GST1*prom-pMOSBlue #6 was chosen and plasmid DNA isolated using the Qiagen Midi plasmid purification kit for large-scale preparations of plasmid DNA. Plasmid DNA of pCAMBIA2300 was prepared in the same way. Both the insert (*gst1* promoter) and vector (pCAMBIA2300) were prepared with *Kpn*I and *Pst*I restriction enzymes for cloning.

3.3.2.1 Preparation of the *gst1* promoter pMOSBlue insert for subcloning into pCAMBIA2300

Isolation of the *gst1* promoter fragment after *Kpn*I/ *Pst*I digestion of *GST1*prom-pMOSBlue #6 yielded a fragment of ~900 bp (Figure 3.8, lane 1). This correlates well with the expected size of 919 bp. The concentration of the eluted fragment was determined to be ~30 ng/µl.

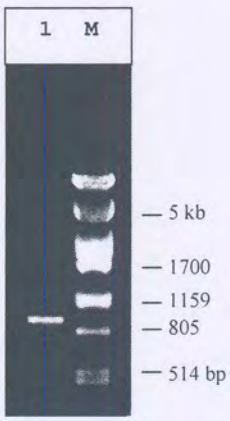


Figure 3.8 ***Kpn*I/ *Pst*I digested *gst1* promoter restriction fragment eluted from an agarose gel.** Lane 1: 2 µl sample of the 30 µl eluted fragment; M: λDNA/ *Pst*I marker.

3.3.2.2 Preparation of pCAMBIA2300

pCAMBIA2300 is a binary plant transformation vector that contains minimal heterologous sequences for plant transformation and selection of transformants (www.cambia.org). A map of pCAMBIA2300 is presented in **Appendix C**. Between the T-borders, it has the CaMV 35S-driven and terminated plant selection neomycin phosphotransferase II (*nptII*) gene that encodes resistance to kanamycin. Indicated also is the lacZα gene, that is interrupted when the insert is cloned into the pUC18 polylinker. This allows for blue/white screening of clones in *E. coli* cells. Outside the T-borders, it has the bacterial kanamycin resistance marker for selection in *E. coli* and *Agrobacterium* strains. It has the wide-host-range origin of replication from the *Pseudomonas* plasmid pVS1, which is very stable in the absence of selection. It also has the pBR322 origin of replication to allow high-yielding DNA preparations in *E. coli*.

Five microgram of pCAMBIA2300 plasmid DNA was digested with *Kpn*I and *Pst*I. The fragment corresponding to the linearised vector (8715 bp) was recovered from the gel. The concentration of the eluted fragment was determined to be ~120 ng/µl.

3.3.2.3 Ligation of the *gst1* promoter into pCAMBIA2300

The 919 bp *gst1* promoter fragment was ligated to the digested pCAMBIA2300 vector in ligation reactions set at various insert-to-vector ratios and vector DNA quantities. The ligation reactions were

transformed into competent *E. coli* DH5α. The cells were plated out onto LB agar plates containing 100 µg/ml kanamycin and that were spread with X-gal.

After ligation and transformation, eight colonies transformed with putative recombinant *GST1*prom-pCAMBIA2300 constructs were identified. Blue/white colony selection, based on the α-complementation system, was used. Two white vector-religated colonies were obtained. No blue colonies were obtained except in the positive transformation control containing pCAMBIA2300 plasmid.

3.3.2.4 Screening of transformants by miniprep and restriction enzyme digestions

Qiafilter miniprep plasmid DNA from eight putative *GST1*prom-pCAMBIA2300 transformed colonies and one vector-religated transformed colony was digested with *Kpn*I and *Pst*I to verify the presence of the *gst1* promoter insert. A fragment of approximately 900 bp was excised with *Kpn*I and *Pst*I from clone 3, indicating it possibly being a recombinant (Figure 3.9, lane 7).

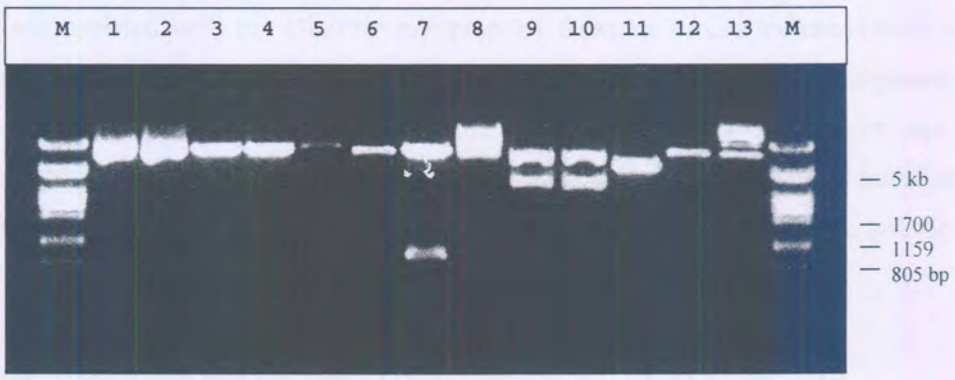


Figure 3.9 **Restriction enzyme screening of eight putative recombinant *GST1*prom-pCAMBIA2300 transformed colonies.** M: λDNA/ *Pst*I marker; lanes 1 and 2: undigested pCAMBIA2300 plasmid DNA; lanes 3 and 4: *Kpn*I/ *Pst*I digested pCAMBIA2300 plasmid DNA; lanes 5 to 12: *Kpn*I and *Pst*I restriction digestion of putative recombinant *GST1*prom-pCAMBIA2300 clones 1 to 8, respectively; lane 13: Religated vector plasmid digested with *Kpn*I and *Pst*I.

*GST1*prom-pCAMBIA2300 #3 plasmid DNA was chosen for sequencing of the insert. When the ligation and transformation was repeated, another eight colonies were obtained. Only one, clone 11, contained an insert that could be excised with *Kpn*I and *Pst*I restriction digestion (Figure 3.10, lane 3). The fragment seemed to be of the correct size, and clone 11 would have served as a backup if the sequence of clone 3 contained errors.

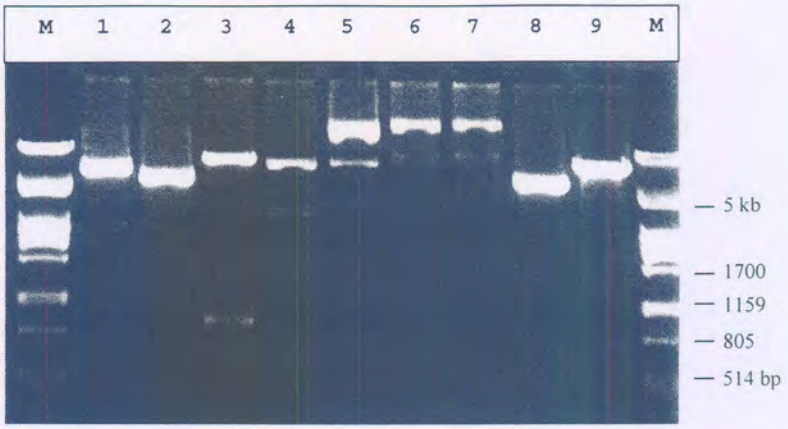


Figure 3.10 **Restriction enzyme screening of putative recombinant *GST1*prom-pCAMBIA2300 clones 9 to 16.** M: λDNA/ *Pst*I marker; lanes 1 to 8: miniprep plasmid DNA isolated from *E. coli* colonies transformed with putative *GST1*prom-pCAMBIA2300 clones 9 to 16, respectively, and digested with *Kpn*I and *Pst*I; lane 9: *Kpn*I/ *Pst*I digested pCAMBIA2300 plasmid DNA.

3.3.2.5 Nucleotide sequencing of *GST1*prom-pCAMBIA2300 #3

Plasmid DNA was isolated with the Qiafilter miniprep kit from an *E. coli* culture containing the *GST1*prom-pCAMBIA2300 #3 construct. The junction points and *gst1* promoter sequences were determined by nucleotide sequencing of the insert. The primers used were PUC / M13 R and PUC / M13-40F. These primer-binding sequences flank the multiple cloning site of the plant transformation vector pCAMBIA2300. Nucleotide sequencing showed a 100% correct sequence of the *gst1* promoter cloned into pCAMBIA2300, with the junction points as expected.

3.3.3 Subcloning of the apple *pgip1* cassette into *GST1*prom-pCAMBIA2300

The previous section described how the *gst1* promoter was cloned into a plant transformation vector to form the construct that is called *GST1*prom-pCAMBIA2300. The aim of this section was to clone the apple *pgip1* gene, in the form of a cassette including also the TEV leader and CaMV 35S terminator, downstream of the *gst1* promoter in pCAMBIA2300. The source of the apple *pgip1* cassette is pAppRTL2 (**Appendix C**), which was previously generated by cloning the apple *pgip1* gene into the pRTL2 vector (Arendse and Berger, unpublished). In this construct, the apple *pgip1* gene is under control of the enhanced CaMV 35S promoter. This vector also provides the TEV leader and CaMV 35S terminator.

3.3.3.1 Preparation of apple *pgip1* cassette with complete *Xho*I and partial *Pst*I restriction digestion

To clone the apple *pgip1* cassette into the engineered *Sal*I and *Pst*I sites of *GST1*prom-pCAMBIA2300 #3, the cassette fragment had to be excised from pAppRTL2 using *Xho*I (5` end) and *Pst*I (3` end)

digestion. *Sal*I and *Xho*I digestion o icky overhangs, thereby facilitating the annealing of the 5' *Xho*I- digested cassette fragment to the *Sal*I-digested vector fragment.

*Pst*I digestion of pAppRTL2, however, posed a problem, since this construct contains three *Pst*I sites (**Appendix C**). Two are located at each end of the cassette and a third site is present within the apple *pgip1* cassette to be subcloned. It lies between the 3' end of the apple *pgip1* gene and the CaMV 35S terminator sequence. Digestion of pAppRTL2 with *Xho*I and *Pst*I would therefore lead to the digestion of the apple *pgip1* cassette into two fragments, 1148 and 231 bp in size. In order to release the full-length cassette (1379 bp), a partial *Pst*I digestion of the plasmid was required, after complete digestion by *Xho*I has taken place. Complete digestion by *Xho*I is necessary to limit the number of fragments that would be generated if *Xho*I digestion, in addition to *Pst*I digestion, was also only partial. The sizes of the expected products are 1379 bp for the full-length cassette, 1148 bp for the truncated cassette and 231 bp for the short 3' piece containing the CaMV 35S terminator.

3.3.3.2 Partial *Pst*I restriction digestion

Partial restriction digests were performed by decreasing the $MgCl_2$ concentration of the restriction buffer. The optimum $MgCl_2$ concentration was determined by preparing buffers that differ only in their $MgCl_2$ concentration, and looking for the greatest yield of the full-length (1379 bp) cassette fragment following restriction enzyme digestion. The digestion times were also varied from 20 to 60 min. Digestion in the presence of buffer H (containing 10 mM $MgCl_2$) and buffer with $1/10^{th}$ the $MgCl_2$ concentration (1 mM) yielded almost similar products for the reactions incubated for 60 and 20 minutes (Figure 3.11, lanes 3, 4, 11 and 12), with 1 mM $MgCl_2$ yielding slightly more of the 1379 bp cassette (Figure 3.11, lanes 4 and 12). Digestion for 40 minutes with 1 mM $MgCl_2$ (Figure 3.11, lane 8) was more complete than at t = 60 min (Figure 3.11, lane 4) and in the presence of 10 mM $MgCl_2$ (Figure 3.11, lanes 3, 7 and 11). This was unexpected, because lanes 3 and 7 were expected to have the most complete digestion. This figure illustrates that all restriction enzymes require Mg^{2+} to function, since without it virtually no digestion took place (Figure 3.11, lanes 5, 6, 9, 10, 13 and 14).

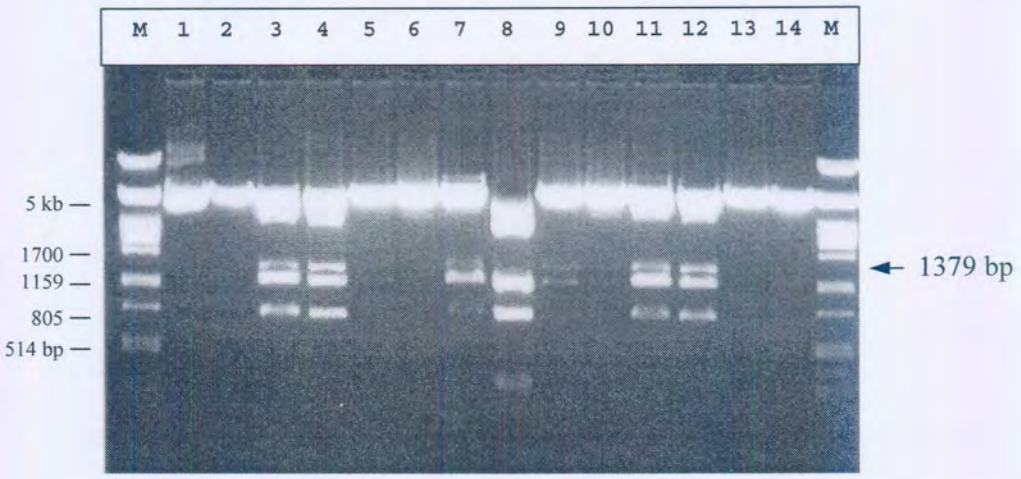


Figure 3.11 **Partial *Pst*I digestion of *Xho*I-linearised pAppRTL2.** *Pst*I digestion of *Xho*I-linearised pAppRTL2 in the presence of different $MgCl_2$ concentrations was performed to determine the optimum $MgCl_2$ concentration where the yield of the full-length cassette fragment is optimal. M: λDNA/ *Pst*I marker; lane 1: undigested pAppRTL2; lane 2: pAppRTL2 linearised with *Xho*I; lanes 3 to 6: 60 minutes digestion with *Pst*I of *Xho*I-linearised pAppRTL2 in the presence of buffer containing 10 mM $MgCl_2$, 1 mM $MgCl_2$, 0 mM $MgCl_2$ and dH_2O, respectively. Lanes 7 to 10 and lanes 11 to 14: 40 min digestion and 20 min digestion in the presence of the same buffers as at 60 min, respectively.

The buffer containing 1 mM $MgCl_2$ was chosen for the subsequent partial digestions. The 20 minute partial digest was scaled up and loaded into a single long well to separate the cassette fragment from the contaminating fragments by agarose gel electrophoresis (Figure 3.12, lane 1).

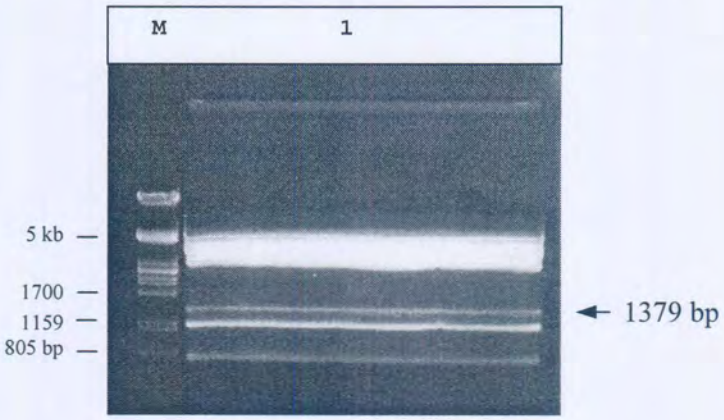


Figure 3.12 **Preparation of the 1379 bp apple *pgip1* cassette insert fragment for cloning.** M: λDNA/ *Pst*I marker, lane 1: Complete *Xho*I-digested, partial *Pst*I-digested pAppRTL2.

The excised 1379 bp apple *pgip1* cassette fragment was contaminated with the shorter 1148 bp cassette fragment (Figure 3.13, lane 5, white arrows). It was co-excised from the gel by accident, but it is unclear why since the two fragments were clearly separated (Figure 3.12, lane 1). Ligation of this

mix of insert fragments would therefore lead to colonies containing the full-length cassette and a shortened cassette, the latter lacking the 3' terminal *Pst*I fragment. Isolation of the 1379 bp fragment was repeated, and again found to be contaminated by the second fragment. The concentrations of the insert fragments recovered from the agarose gel were estimated to be between 15 ng/µl and 30 ng/µl.

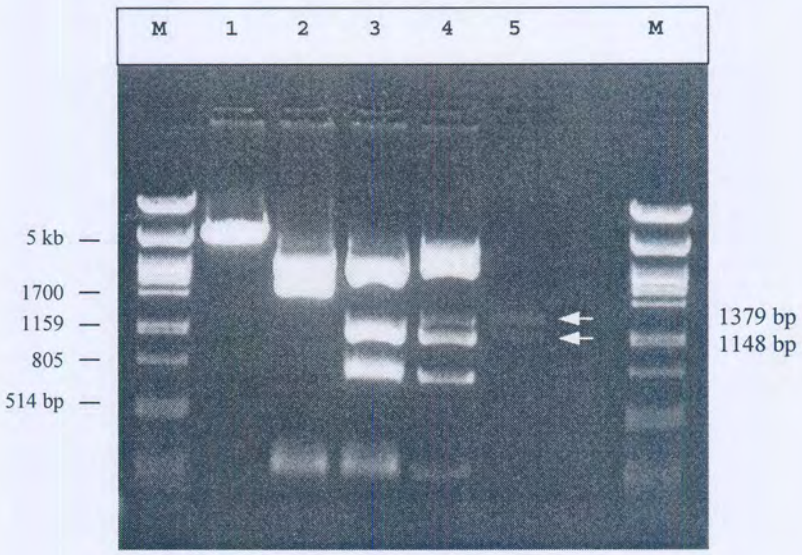


Figure 3.13 **Full and partial digested pAppRTL2 and gel extracted insert fragment containing the contaminating fragment.** M: λDNA/ *Pst*I marker; lanes 1 to 4: pAppRTL2 digested with *Xho*I, *Pst*I, *Xho*I and *Pst*I and *Xho*I & partial *Pst*I, respectively; lane 5: cassette fragment eluted from the gel.

3.3.3.3 Preparation of the *GST1*prom-pCAMBIA2300 #3 vector

The plant transformation vector containing the *gst1* promoter (*GST1*prom-pCAMBIA2300 #3) was prepared with *Sal*I and *Pst*I for cloning of the apple *pgip1* cassette downstream of the *gst1* promoter. Eight microgram *GST1*prom-pCAMBIA2300 #3 plasmid DNA was digested with *Sal*I and *Pst*I restriction enzymes. The fragment corresponding to the linearised vector (9635 bp) was excised and eluted from the gel. The concentration of the eluted fragment was determined to be ~400 ng/µl.

3.3.3.4 Ligation of the apple *pgip1* cassette into *GST1*prom-pCAMBIA2300 #3

The 1379 bp apple *pgip1* cassette fragment (*Xho*I/ *Pst*I digested) was ligated to the *Sal*I/ *Pst*I digested *GST1*prom-pCAMBIA2300 #3 vector in ligation reactions set at various insert-to-vector ratios and vector DNA quantities. Competent *E. coli* DH5α cells were transformed with the ligation reactions as described before. The cells were plated out onto LB agar plates containing 100 µg/ml kanamycin. Several colonies were obtained from two repetitions of ligation and transformation. 24 and 90 colonies, respectively.

3.3.3.5 Screening of putative *GST1*prom-ap*pgip1*-pCAMBIA2300 transformants

Forty-four of the colonies were screened for the presence of recombinant constructs by plasmid DNA isolation and restriction digestion. *Kpn*I and *Pst*I were chosen as restriction enzymes to screen for putative *GST1*prom-ap*pgip1*-pCAMBIA2300 clones, since they can discriminate between positive and negative clones and between full-length and truncated cassette constructs. Three fragments with sizes of 8715, 2053 and 231 bp are expected from positive clones containing the full-length cassette. Fragments corresponding to these sizes were obtained from seven clones (Figure 3.14, lanes 5, 6, 20, 25, 26, 30 and 38). The lanes are numbered the same as the clones. Only two fragments of 8715 and 2053 bp are expected for transformants containing the truncated cassette, lacking the 231 bp fragment. Ten clones that contained the truncated cassette were among the 44 clones screened (Figure 3.14, lanes 4, 24, 28, 29, 31, 32, 39, 40, 41 and 44).

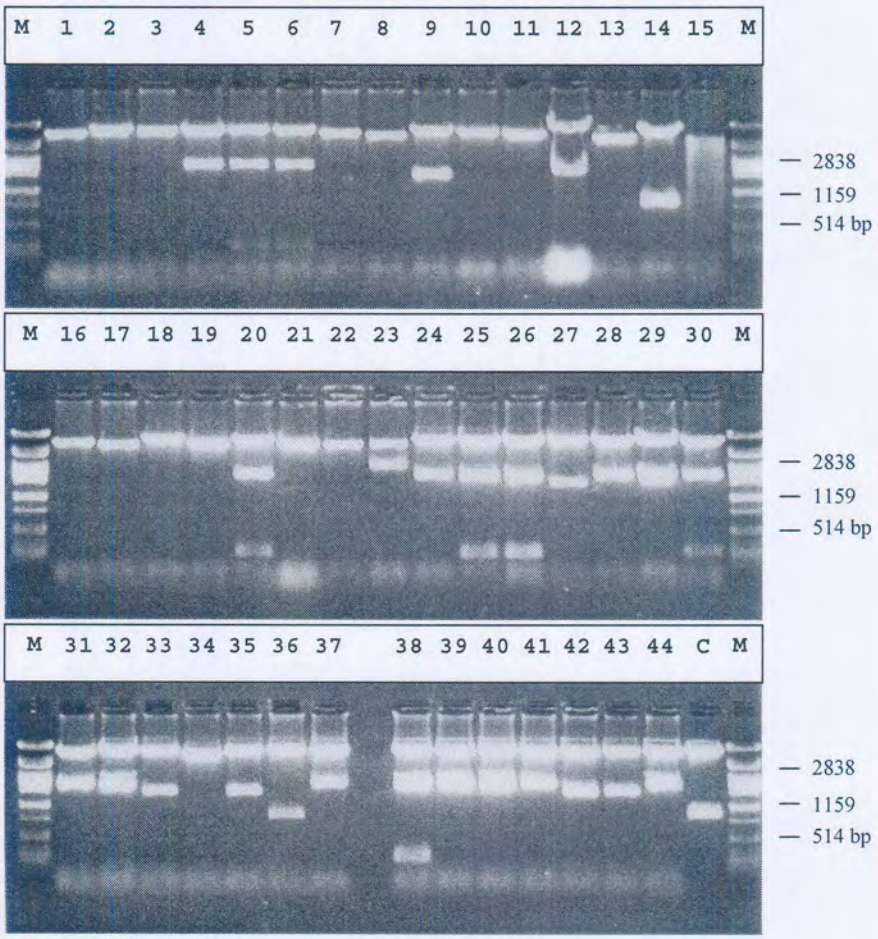


Figure 3.14 ***Kpn*I and *Pst*I digestion of plasmid DNA from 44 putative *GST1*prom-ap*pgip1*-pCAMBIA2300 clones.** M: λDNA/ *Pst*I marker; C: *Kpn*I/ *Pst*I digested *GST1*prom-pCAMBIA2300 #3 (negative control); lanes 1 to 44: lane numbers correspond to the putative recombinant *GST1*prom-ap*pgip1*-pCAMBIA2300 clones digested with *Kpn*I/ *Pst*I.

3.3.3.6 Restriction analysis of seven positive *GST1*prom-ap*pgip1*-pCAMBIA2300 clones

Restriction analysis was performed on plasmid DNA from the seven putative positive clones using the restriction enzymes *Hind*III, and a double digest with *Kpn*I and *Xba*I. The sizes of the expected fragments following *Hind*III digestion are 8789, 1094, 656 and 460 bp for a clone containing the full-length cassette and 8789, 1094, 656 and 229 bp for a truncated cassette. *Kpn*I & *Xba*I digestion liberates fragments with sizes 8936, 1565 and 498 bp from full-length cassette clones and 9203 and 1565 bp from truncated cassette clones. The seven clones showed the expected restriction pattern of a full-length apple *pgip1* cassette cloned downstream of the *gst1* promoter (Figure 3.15 A (*Hind*III digestions) and B (*Kpn*I and *Xba*I digestions), lanes 1 to 7). *Kpn*I and *Xba*I digestion showed conclusively that the 3' terminal *Pst*I fragment of the apple *pgip1* cassette is present in all seven clones, due to the presence of the 498 bp fragment which is not expected from truncated clones.

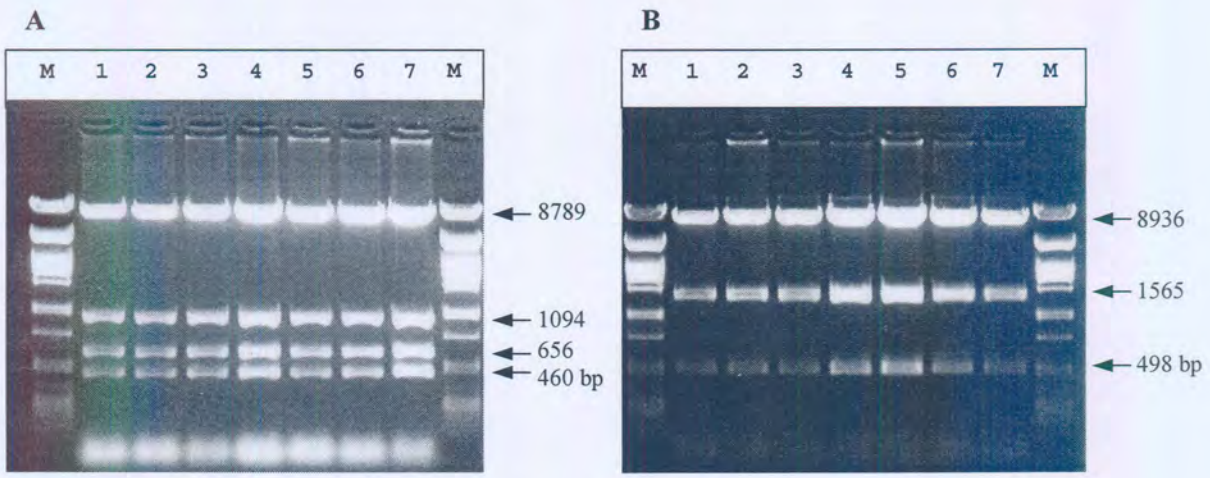


Figure 3.15 **Restriction analysis of seven positive *GST1*prom-ap*pgip1*-pCAMBIA2300 clones.** M: λDNA/ *Pst*I marker; A: lanes 1 to 7: *Hind*III restriction digestion of plasmid DNA from putative *GST1*prom-ap*pgip1*-pCAMBIA2300 clones 5, 6, 20, 25, 26, 30 and 38, respectively. B: lanes 1 to 7: *Kpn*I & *Xba*I digested plasmid DNA from the same clones used in the *Hind*III digests.

3.3.3.7 Nucleotide sequencing of *GST1*prom-ap*pgip1*-pCAMBIA2300

*GST1*prom-ap*pgip1*-pCAMBIA2300 #25 and #30 were selected and sent for nucleotide sequencing of the insert. The primers used for sequencing the junction sites were AP-PGIP-INVR and PUC / M13-40F. AP-PGIP-INVR anneals to a sequence in the middle of the apple *pgip1* gene (Figure 3.16). Using it in the sequencing reaction yielded the sequence upstream of it to the end of the *gst1* promoter. PUC / M13-40F anneals downstream of the CaMV 35S terminator, and gave the sequence up to the middle of the apple *pgip1* gene. Figure 3.16 shows the relative positions and lengths of the sequences obtained for the respective primers used. Sequence analysis showed 100% correct junction points for both clones. The details of the nucleotide sequence between the TATA-box of the *gst1* promoter and

the initiation codon of the apple *pgip1* gene is shown in Figure 3.17. The predicted amino acid sequences are also shown (see Discussion).

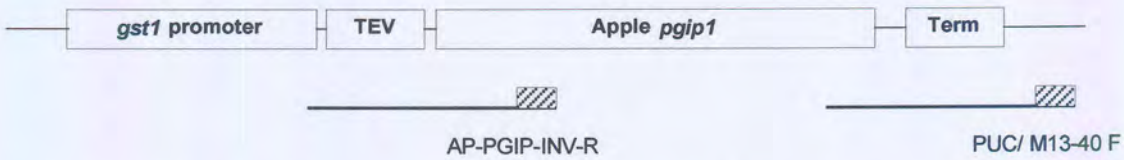


Figure 3.16 **Nucleotide sequencing of the *gst1* promoter - apple *pgip1* cassette inserted into pCAMBIA2300.** Schematic representation of the primer annealing positions and the regions sequenced during nucleotide sequencing of the *gst1* promoter and apple *pgip1* cassette cloned into the plant transformation vector pCAMBIA2300. The hatched boxes represent the primers used for sequencing (AP-PGIP-INVR and PUC / M13-40F) and the lines extending upstream from them represent the relative length of sequencing data obtained. TEV: Tobacco etch virus leader sequence; Term: CaMV 35S terminator.

```
TATAAATACACACTCCCATTTGTGTATTTCTTTTCATCAATCACAAAGATCTCTCTACTT      60
Y  K  Y  T  L  P  F  V  Y  F  F  S  S  I  T  K  I  S  L  L
 I  N  T  H  S  H  L  C  I  S  F  H  Q  S  Q  R  S  L  Y  F
  *  I  H  T  P  I  C  V  F  L  F  I  N  H  K  D  L  S  T  S
↓
CAATAAATCTCCACCTTACTTTAAGAACAAGAAAAACACAGTATTAACAGTCGAGAATTC     120
Q  *  I  S  T  L  L  *  E  Q  E  K  H  S  I  N  S  R  E  F
 N  K  S  P  P  Y  F  K  N  K  K  N  T  V  L  T  V  E  N  S
  I  N  L  H  L  T  L  R  T  R  K  T  Q  Y  *  Q  S  R  I  L

TCAACACAACATATACAAAACAAACGAATCTCAAGCAATCAAGCATTCTACTTCTATTGC     180
S  T  Q  H  I  Q  N  K  R  I  S  S  N  Q  A  F  Y  F  Y  C
 Q  H  N  I  Y  K  T  N  E  S  Q  A  I  K  H  S  T  S  I  A
  N  T  T  Y  T  K  Q  T  N  L  K  Q  S  S  I  L  L  L  L  Q

AGCAATTTAAATCATTTCTTTTAAAGCAAAAGCAATTTTCTGAAAATTTTCACCATTTAC     240
S  N  L  N  H  F  F  *  S  K  S  N  F  L  K  I  F  T  I  Y
 A  I  *  I  I  S  F  K  A  K  A  I  F  *  K  F  S  P  F  T
  Q  F  K  S  F  L  L  K  Q  K  Q  F  S  E  N  F  H  H  L  R

GAACGATAGCCATGGAACTCAAG                                          263
E  R  *  P  W  N  S
 N  D  S  H  G  T  Q
  T  I  A  M  E  L  K
```

Figure 3.17 **Nucleotide sequence and predicted amino acid sequence between the TATA box of the *gst1* promoter and the initiator codon of the apple *pgip1* gene in the *GST1*prom-a*ppgip1*-pCAMBIA2300 construct.** The TATA-box of the *gst1* promoter is indicated in blue, the TEV leader sequence is indicated in red, the initiator codon of the apple *pgip1* gene is boxed and the only methionine residue is indicated in green. Amino acid residues are indicated by their one-letter code and stop codons by asterisks (*).

Clone 30 was chosen for further experiments and was named *GST1*prom-ap*pgip1*-pCAMBIA2300#30. The plasmid map of *GST1*prom-ap*pgip1*-pCAMBIA2300 is included in Figure 3.18.

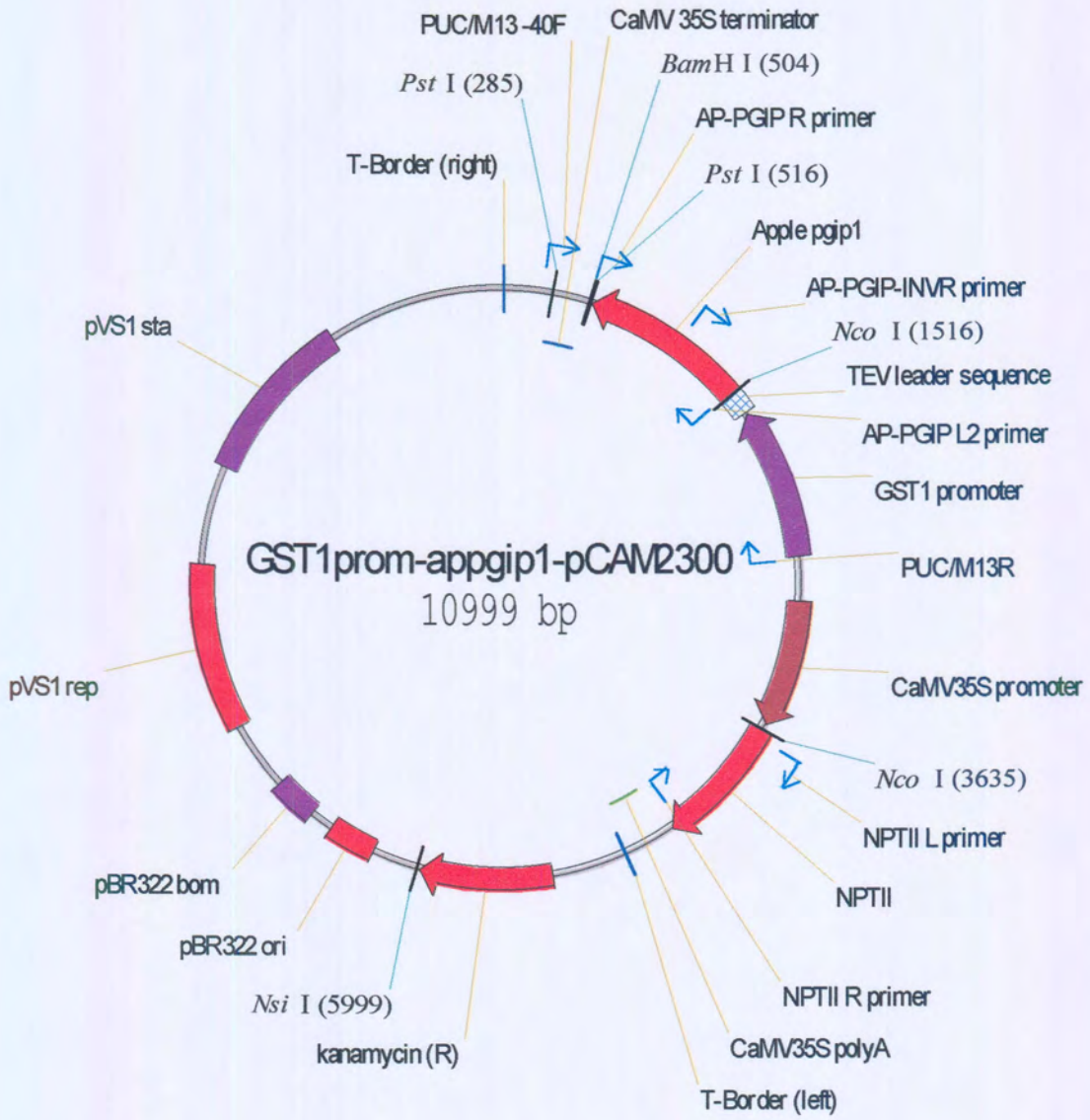


Figure 3.18 **Plasmid map of *GST1*prom-ap*pgip1*-pCAMBIA2300.**

3.4 Discussion

3.4.1 PCR primer design

Before subcloning the *gst1* promoter upstream of the apple *pgip1* gene, it was essential to isolate the appropriate part of the *gst1* promoter from a previous construct (p*GST1*-BluescriptSK$^-$). The polymerase chain reaction (PCR) was employed, since it can selectively amplify a specific region of DNA while at the same time incorporate desired sequences at the ends of the fragment. The GSTreverse primer was designed to amplify the promoter just upstream of the ATG start codon of the *gst1* gene. Translation by plant ribosomes is affected negatively if they encounter extra start codons before the correct one.

In prokaryotes, the initiator AUG codon is preceded by a purine rich "Shine/Dalgarno" sequence centred about ten nucleotides upstream. In contrast, eukaryotic ribosomes bind to the 5' end of the transcript and migrate along it, to start translation at the first AUG triplet it encounters, with a few exceptions. Studies with 79 higher plants confirmed the role of the first AUG codon on the processed mRNA as the translation initiation site. The consensus sequence for the plant initiator site was TAAACAATGGCT (on the plus strand of DNA) (Joshi, 1987), which conforms to the general eukaryotic model of a purine at position -3 and +4. The preferred consensus sequence for eukaryotic (mostly animal) translation initiation sequences is [(A/G)XXAUGG] (Kozak, 1981). Thus, purines are favoured in position -3, suggesting that purines in positions -3 and +4 might facilitate recognition of the AUG codon during formation of the initiation complex. The X's were later established to favour pyrimidines. The sequence around the initiator codon of the apple *pgip1* gene is **G**CCATG**G**. It conforms thus to both the plant and animal eukaryotic translation initiation consensus sequences, having purines at position -3 and +4 (indicated in bold) and pyrimidines at positions -1 and -2.

The nucleotide sequence between the TATA box of the *gst1* promoter and the initiator codon of the apple *pgip1* gene in the *GST1*prom-ap*pgip1*-pCAMBIA2300 construct was presented in Figure 3.17. The transcription initiation site of the *gst1* promoter was mapped by primer extension (Yang *et al.*, 1998) to be 50 bp upstream of the translation initiation codon (indicated by an arrow ↓). The translation into all three reading frames is indicated by the one letter amino acid code. The only methionine residue (M) in all three reading frames is the one corresponding to the initiator codon of the apple *pgip1* gene (indicated in green). This illustrates that there are no other initiator codons in the mRNA transcript than the correct one. Since it will be the first AUG that the eukaryotic ribosome will encounter while it migrates along the mRNA transcript, translation of the apple *pgip1* gene will proceed correctly.

3.4.2 PCR amplification and gel purification of the *gst1* promoter

During PCR amplification of the appropriate part of the *gst1* promoter using the GSTreverse and SK primers, a fragment of 1008 bp was expected. An amplification product of approximately 900 bp was obtained for all reactions using either Taq DNA Polymerase or the Expand Long template PCR system (Figure 3.2, lanes 1 to 5, and Figure 3.3, lanes 1 and 2). This was unexpected, but may have been an artefact of the agarose gel, perhaps being overloaded with the high yield of PCR amplification product. After gel extraction of the PCR product, a faint band of a smaller size (~700 bp) was observed together with the purified PCR fragment of ~900 bp (Figure 3.4, lane 1). The identity of this fragment was unknown, and no colonies transformed with recombinant pMOSBlue containing this smaller insert were obtained. Nucleotide sequencing of the recombinant constructs, *GST1*prom-pMOSBlue #6 and #9, showed that the correct sequence was amplified and cloned. It is speculated that the smaller size fragment is single-stranded DNA that is eluted from the silica column during purification of the fragment using the Qiaquick gel extraction kit. It migrates faster than the double stranded fragment, so that it seems smaller in size. Because it is single-stranded, it cannot ligate to the pMOSBlue vector to produce recombinant constructs.

3.4.3 Subcloning of the *gst1* promoter PCR product into pMOSBlue

The PCR product was cloned into the pMOSBlue vector for two reasons. Firstly, it was to enable nucleotide sequencing of the appropriate part of the *gst1* promoter to confirm the correct sequence before proceeding with the construct preparation. Secondly, it was to efficiently excise the promoter insert from the pMOSBlue vector for subcloning into pCAMBIA2300, since restriction digestion of PCR products is not very efficient. Two colonies transformed with recombinant *GST1*prom-pMOSBlue constructs were obtained, of which only clone 6 had a 100% correct sequence. The 5' part of the GSTreverse primer, containing the *Pst*I site, was missing from the insert cloned in clone 9. The rest of the *gst1* promoter sequence was, however, as expected. Clone 6 was therefore chosen for further experiments, so that the apple *pgip1* cassette could be later inserted into the *Sal*I and *Pst*I sites that was incorporated by the GSTreverse primer.

3.4.4 Subcloning of the *gst1* promoter into pCAMBIA2300

The plasmid pCAMBIA2300 is an *A. tumefaciens* binary vector (www.cambia.com). It is a building block for plant transformation because it has a multiple cloning site between the left and right border sequences in which a transgene can be inserted. It also contains the neomycin phosphotransferase II (*nptII*) gene, encoding kanamycin resistance under control of plant expression signals, to simplify selection of transformed plants. Two colonies transformed with recombinant *GST1*prom-

pCAMBIA2300 constructs were obtained, and nucleotide sequencing of clone 3 showed the correct junction sequences of the *gst1* promoter fragment ligated into pCAMBIA2300.

3.4.5 Subcloning of the apple *pgip1* cassette downstream of the *gst1* promoter in pCAMBIA2300

pAppRTL2 contains the apple *pgip1* gene in an expression cassette with the TEV leader and CaMV 35S terminator (**Appendix C**). Preparation of the cassette for subcloning it downstream of the *gst1* promoter in pCAMBIA2300 required digestion with *Xho*I and *Pst*I. *Xho*I cuts 5` to the TEV leader, and *Pst*I at the 3` end of the cassette. However, an additional *Pst*I recognition site lies between the apple *pgip1* gene and the CaMV 35S terminator. Partial restriction digestion thus had to be performed to isolate the full-length (1379 bp) cassette.

The $MgCl_2$ concentration, restriction enzyme (*Pst*I) concentration and incubation time are important factors to consider during partial digestion. Too much enzyme and too long digestion times will result in complete digestion. 1U Enzyme per 1 µg linearised plasmid, digested for 20 min in 1 mM ($1/10^{th}$ of the optimal) Mg^{2+} concentration, yielded the highest yield of excised full-length cassette (Figure 3.11, lane 12). The only reaction which gave complete *Pst*I digestion, was digestion for 40 minutes with 1 mM $MgCl_2$ (Figure 3.11, lane 8). It is unclear why this reaction was more complete than at t = 60 min or with 100% Mg^{2+} concentration (compare Figure 3.11 lane 8 with lanes 3, 7 and 11). The most logical explanation is that a pipetting error of too much enzyme mastermix was made to this tube, resulting in more enzyme digestion activity being present.

Because the 1379 bp apple *pgip1* cassette fragment, excised from the agarose gel, was contaminated with the 1148 bp shortened cassette fragment (Figure 3.13, lane 5), several colonies containing the truncated cassette (lacking the 3` terminal part) were obtained (Figure 3.14, lanes 4, 24, 28, 29, 31, 32, 39, 40, 41 and 44). Both the full-length and the truncated cassette are flanked by sticky ends of *Xho*I on the 5` end and *Pst*I on the 3` end. This enables them to anneal and ligate to the vector prepared with *Sal*I and *Pst*I. Fortunately, seven of the 44 colonies screened by restriction digestion contained the 3` terminal *Pst*I fragment. Further restriction analysis (Figure 3.15) confirmed the full-length identity of the subcloned apple *pgip1* cassette in all seven clones. Two clones, numbers 25 and 30, were sent for nucleotide sequencing of the junction points, and both contained the correct sequences.

This chapter reported on the cloning of the apple *pgip1* gene under control of the pathogen-inducible *gst1* promoter of *A. thaliana* into a plant transformation vector. The transformation of *A. thaliana* plants with the *gst1* promoter- and e35S- apple *pgip1* constructs, as well as molecular analysis of the transformants and PGIP expression studies, will be reported in Chapter 4.