An assessment of the mutation patterns in South African isolates of *Potato leafroll virus* and the expression of recombinant viral coat protein genes in *Escherichia coli*

Adri Hilda Rothmann

Thesis presented in partial fulfillment of the requirements for the
degree of Master of Science (Biochemistry)
at the University of Stellenbosch

Supervisor: Prof. D. U. Bellstedt

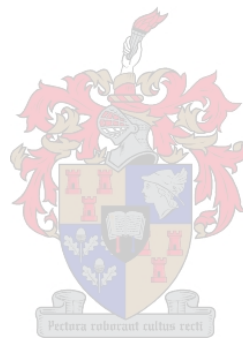Department of Biochemistry

University of Stellenbosch

March 2007

Declaration

I, the undersigned, hereby declare that the work contained in this thesis is my own original work and that I have not previously in its entirety or in part submitted it at any university for a degree.

Signed: Date:

Summary

Presently, the observed variation in symptoms of *Potato leafroll virus* (PLRV) infection in potato cultivars in South Africa cannot be reconciled with PLRV symptoms obtained 10-15 years ago, even if the different interactions between the pathogen and the cultivar are taken into account.

In an effort to analyze this variation, mutations in the coat protein (CP) gene of South African isolates of PLRV were assessed. The CP gene of PLRV isolates from different areas within South Africa was amplified by reverse transcription-polymerase chain reaction (RT-PCR), cloned and sequenced. Significant sequence variation in the CP gene was found within the analyzed South African isolates of PLRV. Phylogenetic analysis revealed two major clades with most South African isolates and an Australian and North American isolate grouped together and the remainder grouped with isolates from diverse countries worldwide. The deduced amino acid sequences from representatives of these two clades indicated differences in CP three-dimensional structure.

In an effort to produce recombinant PLRV CP for the production of antibodies specific for South African isolates of PLRV for use in enzyme-linked immunosorbent assay (ELISA), the CP gene of a South African isolate of PLRV was subcloned into a bacterial expression vector (pET14-b). Expression of full length recombinant PLRV CP was attempted in *Escherichia coli* strains BL21(DE3)pLysS, Rosetta-gami B(DE3)pLysS and Rosetta-2(DE3)pLysS. As this was not successful, the PLRV CP gene was subcloned in another expression vector (pGEX) for expression as an N-terminal fusion protein with glutathione-*S*-transferase (GST) in *E. coli* strains BL21(DE3)pLysS and Rosetta-2(DE3)pLysS. The recombinant GST-PLRV CP fusion protein was purified and used for antibody production in rabbits. Using western blots, the effectiveness of antibodies produced to recombinant GST-PLRV CP fusion protein was assessed for PLRV recognition. It was found that antibodies to the recombinant GST-PLRV CP fusion protein were more effective for the detection of GST than PLRV CP and that production of antibodies to the cleaved PLRV CP product would be necessary if antibodies are required for ELISA applications.

Opsomming

Huidiglik kan die waargeneemde simptome van infeksie met aartappelrolbladvirus (*Potato leafroll virus*, PLRV) in aartappelkultivars in Suid-Afrika nie vereenselwig word met PLRV simptome wat 10-15 jaar gelede verkry was nie, selfs al word die verskillende interaksies tussen die patogeen en kultivar in ag geneem.

In 'n poging om hierdie variasie te analiseer, was mutasies in die mantelproteïen (CP) geen van Suid-Afrikaanse isolate van PLRV bepaal. Die CP geen van PLRV isolate van verskillende areas in Suid-Afrika was ge-amplifiseer met behulp van die tru transkripsie-polimerase ketting reaksie (RT-PCR), gekloneer en die nukleotiedvolgorde bepaal. Noemenswaardige nukleotied variasie is in die CP gene van die ge-analiseerde Suid-Afrikaanse isolate van PLRV gevind. Filogenetiese analises het gedui op twee hoof klades met die meeste van die Suid-Afrikaanse isolate wat saam met 'n Australiese en Noord-Amerikaanse isolaat gegroepeer en die res wat met isolate van verskillende lande wêreldwyd gegroepeer. Die afgeleide aminosuurvolgordes van verteenwoordigers van bogenoemde twee klades het gedui op verskille in die CP drie-dimensionele struktuur.

In 'n poging om rekombinante PLRV CP te produseer vir die produksie van antiliggame spesifiek teen Suid-Afrikaanse isolate van PLRV om in "enzyme-linked immunosorbent assay" (ELISA) te gebruik, was die CP geen van 'n Suid-Afrikaanse isolaat van PLRV gesubkloneer in 'n bakteriële ekspressie vektor (pET14-b). Daar was gepoog om vollengte rekombinante PLRV CP in die *Escherichia coli* rasse BL21(DE3)pLysS, Rosetta-gami B(DE3)pLysS en Rosetta-2(DE3)pLysS te produseer. Aangesien dit nie suksesvol was nie, was die PLRV CP gesubkloneer in 'n ander ekspressie vektor (pGEX) sodat die proteïen as 'n N-terminale fusie proteïen met "glutathione-*S*-transferase" (GST) in *E. coli* rasse BL21(DE3)pLysS en Rosetta-2(DE3)pLysS geproduseer kon word. Die rekombinante GST-PLRV CP fusie proteïen was gesuiwer en gebruik vir antiliggaam produksie in konyne. Die effektiwiteit van die antiliggame wat teen rekombinante GST-PLRV CP fusie proteïen geproduseer was vir PLRV herkenning is deur middel van "western blots" geanaliseer. Dit was gevind dat antiliggame teen die rekombinante GST-PLRV CP fusie proteïen meer effektief was vir die herkenning van GST as PLRV CP. Gevolglik sal dit nodig wees om antiliggame teen die gesnyde PLRV CP produk te maak vir gebruik in ELISA.

Acknowledgements

Abbreviations

ASPV	*Apple stem pitting virus*
BWYV	*Beet western yellows virus*
BSA	bovine serum albumin
BYDV	*Barley yellow dwarf virus*
BYDV-PAV	the PAV serotype of *Barley yellow dwarf virus*
C$_{bis}$	concentration N,N'-methylene-bis-acrylamide
CP	coat protein
CTV	*Citrus tristeza virus*
CYSDV	*Cucurbit yellow stunting disorder virus*
DAS-ELISA	double-antibody sandwich (DAS) enzyme-linked immunosorbent assay (ELISA)
dNTP	deoxyribonucleotide triphosphate
DTT	dithiothreitol
EDTA	ethylene diamine tetra-acetic acid di-sodium salt
ELISA	enzyme-linked immunosorbent assay
FBNYV	*Faba bean nectrotic yellow virus*
GLRaV-3	*Grapevine leafroll associated closterovirus*-3
gRNA	genomic RNA
GRSPaV, RSPaV	*Grapevine rupestris stem pitting associated virus*
GST	glutathione-*S*-transferase
HRP	horseradish peroxidase
IPTG	isopropyl β-D-thiogalactopyranoside
kb	kilo-base
kDa	kilo-Dalton
LB	Luria-Bertani
MAbs	monoclonal antibodies
MBP	maltose-binding protein
M$_r$	relative molecular mass
nt	nucleotides
ORF	open reading frame
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PDV	*Prune dwarf virus*
PLRV	*Potato leafroll virus*
PVA	*Potato virus A*
R-domain	arginine rich domain
RdRp	RNA-dependent RNA polymerase
Rf-values	relative mobilities
RT-PCR	reverse transcription polymerase chain reaction
RYMV	*Rice yellow mottle virus*
SDS-PAGE	sodium dodecyl sulphate polyacrylamide gel electrophoresis
sgRNA	subgenomic RNA
S-domain	shell domain
SMYEAV	*Strawberry mild yellow edge associated potexvirus*
SPLSV	*Sweet potato leaf speckling virus*
ssRNA	single-stranded RNA
T	total concentration of acrylamide and N,N'-methylene-bis-acrylamide
TAE	Tris-base, glacial acetic acid, EDTA
TMV	*Tobacco mosaic virus*
TSWV	*Tomato spotted wilt tospovirus*
VPg	virus protein, genome-linked
X-Gal	5-bromo-4-chloro-3-indolyl-β-D-galactoside

Chapter 1: Introduction

Potatoes are one of the world's most important food crops and are subject to plant virus infection which results in serious crop losses (Van den Heuvel *et al.*, 1995; Bustamante and Hull, 1998; Haliloglu and Bostan, 2002). *Potato leafroll virus* (PLRV) is economically the most important and destructive virus affecting potato crops (Massalski and Harrison, 1987; Haliloglu and Bostan, 2002). Economic losses of potato as a result of PLRV infection are due to the reduction in both the quality and yield of potato crops (Kawchuk *et al.*, 1990; Johnson and Pappu, 2006). PLRV infection of potatoes causes starch to be retained in the leaves resulting in smaller tuber formation (McKay, 1955; Rich, 1983). Furthermore, the quality of potatoes is affected due to net necrosis (Johnson and Pappu, 2006).

Worldwide losses to potato yields by PLRV infection are estimated at 2 x 10^7 tons of tubers annually (De Souza-Dias, 1999b). PLRV infection can result in a reduction of up to 60% in crop yield and if the incidence in a seed crop is too high the latter cannot be certified (Shepardson *et al.*, 1980). In the Brazilian potato industry, PLRV was identified as a major cause for seed potato declassification (De Souza-Dias, 1999b). One of the most challenging and costly aspects of production in Brazil is the increasing demand for high-grade seed potatoes (i.e. virus-free seed). PLRV infection of plants from certified seed can be as high as 20% during one growing season in Brazil (De Souza-Dias, 1999b).

In the past in Ireland, PLRV was the most serious disease affecting potatoes with the reduction in yield never less than 40% of the normal for a variety, with losses for some varieties of potato exceeding 90%, practically giving no crop (McKay 1955). The most important crop/virus combination in the United States is considered to be potato and PLRV (Kaniewski and Thomas, 2004). In the Columbia Basin of the northwestern United States, PLRV causes major losses as a result of net necrosis which renders potatoes unfit for the fresh market and all except the least profitable of processed products (Thomas *et al.*, 1997).

In South Africa, the potato industry faces a crisis as a result of the serious increase in the incidence of PLRV during the past few years (Anon, 2005; Coetsee, 2005). The presence of PLRV is nothing new, but it has been increasing steadily over the past few years until the virus infestation became very problematic from 2003 onwards (Anon, 2004a). During the 2004 planting season, PLRV infection was prevalent at levels that have not been seen in South Africa before (Anon, 2004b; Coetsee, 2004). During 2005, the occurrence of PLRV reached what was viewed by the potato industry to be crisis levels (Coetsee, 2004; Coetsee, 2005). During the 2005 growing season PLRV infection occurred in regions which were traditionally known for low virus infection (Anon, 2005; Coetsee, 2005).

The whole South African potato industry is affected by high levels of PLRV infection which include the seed potato industry as well as table potato plantings (Anon, 2004b). Successful production of certified seed potatoes is especially affected by the presence of PLRV (Anon, 2004a). There is a serious PLRV infestation in seed and potato plantings in the Sandveld, Western Cape, which add to the already destructive effect the virus has on the seed potato industry (Anon, 2004a).

The Sandveld is the greatest producer of seed potatoes in South Africa and also plays a significant role in the production of table potatoes (Bonthuys, 2005). Exports from this region have grown considerably over the past three years (Bonthuys, 2005). The Sandveld produces 20 million bags of potatoes of 10 kg each per year, 8 million of this is for the Cape Metropole (Bonthuys, 2005). The prevalence of PLRV infection in the Sandveld region placed seed production under pressure and according to an article in a local newspaper (Bonthuys, 2005) this problem can impair the reputation of local seed production as the producer of quality seed potatoes (Bonthuys, 2005). According to the regional director of Potatoes South Africa, the situation is serious with financial losses as a result of viruses being in the order of hundreds of millions of rands (Bonthuys, 2005).

The seed certification program of Potatoes South Africa is experiencing problems with the detection of PLRV in potatoes in this country. In order for Potatoes South Africa to ensure disease free planting material, the organization requires kits for the specific detection of various potato diseases. For the detection of PLRV in potato leaves and tubers, Potatoes South Africa makes use of European enzyme-linked immunosorbent assay (ELISA) kits. Concerns were expressed that these kits may, however, be unable to detect South African isolates of PLRV.

In a preliminary study performed in this laboratory, it was determined that the coat protein (CP) gene of a South African isolate of PLRV had undergone a number of mutations which may result in a lack of detection by the European ELISA kits (Matzopoulos, 2005). For this reason, the objectives of the present study are to firstly assess the extent to which variation in the CP gene of the South African isolates of PLRV occurs, and secondly, to produce recombinant PLRV CP from a South African isolate for use as an antigen in antibody production. These antibodies would subsequently be used for PLRV detection by ELISA.

In the first part of this project, the CP gene of 39 PLRV isolates from different potato growing areas in South Africa were amplified by reverse transcription-polymerase chain reaction (RT-PCR). These genes were cloned and sequenced in order to assess the sequence variation of PLRV isolates in South Africa and to compare with various PLRV CP nucleotide sequences from Europe and other countries. The amino acid sequences of the South African CPs were deduced and compared to European and other overseas isolates.

In the second part of this project, the cloned PLRV CP gene was sub-cloned into suitable expression vectors and the expression of the PLRV CP gene was attempted in a variety of *E.*

coli strains. Antibody production against recombinant PLRV CP was attempted and the specificity of the antibodies produced was assessed with a view to their potential use in ELISA. This approach obviates the necessity to isolate viral particles from plant material for antibody production. Antibodies produced in this way can give false positive results in an ELISA due to their cross-reaction with plant proteins. The ELISA produced in this way will subsequently be supplied to the South African potato industry for routine detection of PLRV in plant material.

In this thesis a general overview of the pathology and molecular biology of PLRV is given in Chapter 2. Furthermore, diagnostic techniques for the detection of viruses are discussed as well as conventional and recombinant means of producing antibodies to PLRV. Also in Chapter 2, a review is given of bacterial expression systems that were investigated in this study for the production of recombinant CPs followed by the literature cited for this section of the thesis. The results of the assessment of variation in the CP gene of South African isolates in comparison with each other and those from Europe and other countries are presented in Chapter 3. The results of the recombinant PLRV CP production for immunization are given in Chapter 4. A final conclusion and future perspectives are given in Chapter 5. The Appendixes containing the PLRV CP and the vector nucleotide sequences and the PLRV amino acid sequences that were determined in this study are listed at the end of the thesis.

Chapter 2: Literature review

2.1 PLRV: pathology and molecular biology

2.1.1 Virus transmission and spread

PLRV has a limited host range and is mainly restricted to the Solanaceae which includes all potato species (Kawchuk *et al.*, 1990; Taliansky *et al.*, 2003). About 20 solanaceous species have been infected experimentally and the virus is a common pathogen of potato (Taliansky *et al.*, 2003). The virus occasionally also attacks tomato, but does not affect other crops (Rich, 1983; Taliansky *et al.*, 2003). Several potato colonizing aphid species are responsible for the spread of PLRV from plant to plant (McKay, 1955; Rouzé-Jouan *et al.*, 2001; Taliansky *et al.*, 2003). The green peach aphid, *Myzus persicae* Sulzer is the principal and most efficient vector (Kassanis, 1952; Fisken, 1959; Van den Heuvel *et al.*, 1995; Rouzé-Jouan *et al.*, 2001; Johnson, 2003). Other possible aphid vectors include the buckthorn aphid, the potato aphid and the foxglove aphid (Rich, 1983; Robert *et al.*, 2000). In South Africa, *M. persicae* is also the most important primary vector (Laubscher, 2006).

In the case of PLRV, this virus can only be spread by aphids or it can be transmitted experimentally by grafting (Rich, 1983; Johnson, 2003). Other viruses can be spread by fungi, insects, nematodes, leafhoppers and mechanically by contact between plants, plant sap or humans (Rich, 1983; Johnson, 2003). However, PLRV is not transmitted mechanically via inoculation of plant sap (Rich, 1983; Kawchuk *et al.*, 1990; Taliansky *et al.*, 2003). The reason for this is that PLRV can multiply in infected plant cells, but it cannot be transported between the infected cells (Mayo *et al.*, 2000). This results in the primary infected cells not being able to yield an infective centre from which infection can develop (Mayo *et al.*, 2000).

However, Mayo *et al.* (2000) showed that it is possible to transmit PLRV mechanically from extracts of plants that had been inoculated by viruliferous aphids and then post-inoculated by pea enation mosaic virus-2. Other PLRV like Luteoviruses are also mainly dependent on aphid vectors to transmit them directly from the phloem tissue of one plant to another (Reavy and Mayo, 2002, Taliansky *et al.*, 2003). Aphids in both their larval and adult stages, either the winged or non-winged forms, transmit PLRV (Taliansky *et al.*, 2003). Winged aphids are able to spread the virus between fields and over long distances while non-winged aphids can spread the virus from infected source plants to adjacent plants, primarily within rows (Johnson and Pappu, 2006).

A virus-free aphid acquires virus particles along with phloem sap while feeding on infected plants (Van den Heuvel *et al.*, 1995). Phloem sap contains abundant sugars which serve as

food for these insects (Raven and Johnson, 2002). The sugars are obtained by the aphids thrusting their stylets (piercing mouthparts) into phloem cells of leaves and stems to feed on the phloem tissue (Raven and Johnson, 2002). A virus-free aphid must feed on the phloem tissue of an infected plant for at least an hour to acquire the virus and the ability of the aphid to cause infections increases by increasing the length of infection feeding (Kassanis 1952; Johnson and Pappu, 2006; Taliansky *et al.*, 2003). Likewise, for transmission of the virus, a minimum feeding time of one hour is required (Kassanis, 1952; Taliansky *et al.*, 2003).

Before being able to transmit the virus, an incubation period of between 24 and 48 hours in *M. persicae* is required (Rich, 1983; Kassanis, 1952). This is referred to as circulative or persistent transmission, because the vector needs to feed for more than a day on the infected plant to become infective and retain its infectivity for long periods as the virus circulates in the aphid (Kassanis, 1952; Van den Heuvel *et al.*, 1994; Reavy and Mayo, 2002). This contrasts with non-persistent transmission where a virus does not require an incubation period in the body of the insect (Rich, 1983; Reavy and Mayo, 2002). In this type of transmission, virus particles have only a transient association with aphid mouthparts and do not circulate within other parts of the aphid body (Reavy and Mayo, 2002). Vectors in the non-persistent transmission group lose their infectivity in a matter of hours (Kassanis, 1952).

When aphids have acquired virus particles while feeding on phloem sap from infected hosts plants, the virions move from the gut lumen into the haemolymph via the posterior midgut (Fig. 2.1) (Van den Heuvel *et al.*, 1994; Van Regenmortel *et al.*, 2000; Rouzé-Jouan *et al.*, 2001). PLRV traverses the gut membrane by an exocytosis-endocytosis mechanism, presumably involving specific recognition between virus particles and aphid components (Rouzé-Jouan *et al.*, 2001). These virus particles are retained in an infective form in the haemolymph for the lifespan of the aphid (Van den Heuvel *et al.*, 1994). Virus particles may be protected in the haemolymph from proteolytic breakdown by associating non-specifically with symbionin, a chaperon protein produced by *Buchnera* endosymbionts (Rouzé-Jouan *et al.*, 2001). When the virus particles circulating in the haemolymph reach the accessory salivary glands, they enter this gland to arrive in the salivary duct (Van den Heuvel *et al.* 1994; Johnson, 2003). From the salivary duct the virus particles are excreted with the saliva when the aphid feeds (Garret *et al.*, 1993; Van den Heuvel *et al.*, 1994; Van Regenmortel *et al.*, 2000; Reavy and Mayo, 2002).
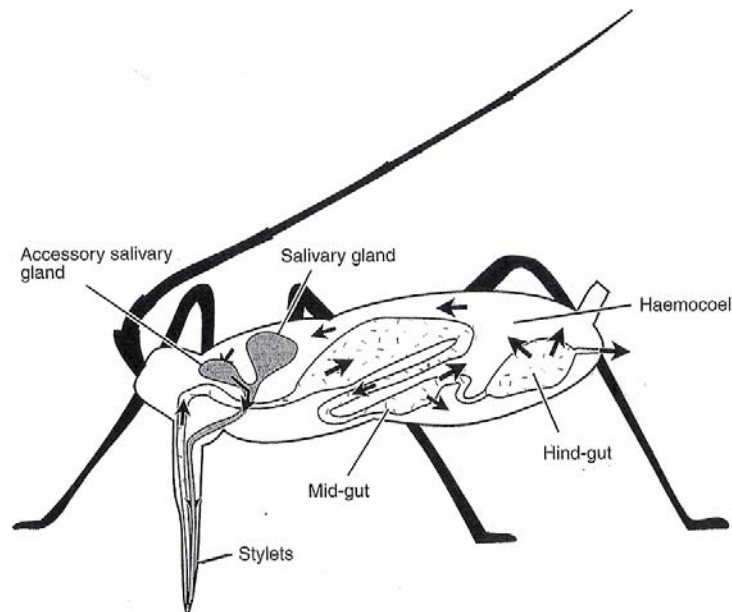
Fig. 2.1: The movement of transmissible luteovirus particles through an aphid vector. Arrows indicate the movement of luteovirus particles after acquisition by an aphid feeding on an infected plant. Phloem contents, containing virus particles, pass from the stylet into the alimentary canal. Virus particles that are absorbed and not excreted in honeydew, pass into the haemocoel from the midgut in the case of PLRV. In the final stage of the process, PLRV particles bind to the cells of the accessory salivary gland, pass through those cells and enter the salivary duct. From the salivary duct, the virus particles are expelled into the phloem tissue of a new plant during subsequent feeding (Reavy and Mayo, 2002).

No replication of PLRV particles takes place whilst they are circulating the haemolymph of *M. persicae* (Tamada and Harrison, 1981; Van den Heuvel *et al.*, 1994; Taliansky *et al.*, 2003). If no replication of virus particles in the vector takes place it is termed non-propagative transmission (Reavy and Mayo, 2002). If the virus replicates in the body of the insect it is characterized as propagative (Rich, 1983). Furthermore, PLRV cannot pass through the egg, so each progeny aphid has to acquire virus by feeding on an infected plant (Johnson and Pappu, 2006). For this reason, the PLRV content of aphids is related to the time they spend on the virus source plants (Tamada and Harrison, 1981). They found that the virus content of aphids fed on virus source plants increased with an increase in acquisition access period and decreased after they were removed from virus source plants. However, virus particles did not accumulate according to the estimated amount of virus in the plants used. This suggests that the vast proportion of virus acquired either passes out of the body in honeydew or is degraded. Only a small portion of the PLRV particles ingested seem to reach the haemolymph.

The efficiency of transmission of PLRV from *M. persicae* or other aphids to healthy potato plants depends on a number of parameters. These include aphid species, clone, morph and instar with as much variation between these factors as between aphid species (Rouzé-Jouan *et al.*, 2001; Taliansky *et al.*, 2003). For instance, non-winged morphs of *M. persicae* are found to be more efficient vectors of PLRV than the winged morphs (Robert *et al.*, 2000) and younger aphids also transmit PLRV most efficiently (Thomas *et al.*, 1997).

Furthermore, the virus isolate also plays a role in transmission efficiency (Rouzé-Jouan *et al.*, 2001). A poorly transmissible isolate of PLRV shows different efficiencies of transmission with different clones of *M. persicae*, indicating that the efficiency of transmission depends upon the interaction of properties of the virus particle and of the aphid vector (Reavy and Mayo, 2002). These interactions can produce a wide range of effects ranging from highly efficient to no transmission (Reavy and Mayo, 2002). Overall, the properties of the virus and the vector and the interaction between these two are involved in regulating the transmission process (Rouzé-Jouan *et al.*, 2001).

There are also factors that influence the efficiency with which *M. persicae* acquires PLRV. The age of plants plays a role as aphids acquire virus more readily from young plants than from older plants enabling aphids to more readily transmit the virus if they have been feeding on younger plants (Syller, 1980; Tamada and Harrison, 1981; Robert *et al.*, 2000). This is the case even if the PLRV concentration in young and older plants does not differ greatly (Tamada and Harrison, 1981). It has been observed early in the growing season that aphids feeding on young potato plants with a high virus titre acquire and transmit PLRV more readily than if they were feeding on potato plants with low titres (Van den Heuvel *et al.*, 1995) yet, later in the growing season, leaves with distinct symptoms were poorer sources for PLRV than those showing faint symptoms (Van den Heuvel *et al.*, 1995).

Older plants could be poorer sources for PLRV because the amount of virus available for acquisition had decreased (Syller, 1980; Van den Heavel *et al.*, 1995). The reduced availability of virus in older plants might have been the result of PLRV spreading to plant cells that are not normally fed on by aphids (Van den Heuvel *et al.*, 1995). On the other hand, Van den Heuvel *et al.* (1995) also found that the number of infected mesophyll cells were too low to describe the reduced availability of virus in older plants. In heavily infected plants, PLRV has been found to be unevenly distributed in the phloem tissue which suggests that systemic transport of the virus in heavily virus-infected vessels is considerably impaired (Van den Heuvel *et al.*, 1995). Aphids apparently avoid feeding on these heavily infected plant tissues and as a result, even though the disease is well developed in the plants, fewer aphids acquire PLRV (Van den Heuvel *et al.*, 1995).

Spread of PLRV is also influenced by factors not directly related to the aphids or virus, but those involving the crop and environment (Herrbach, 1992; Robert *et al.*, 2000). Some of these factors include the amount of initial virus inoculum in seed crops, agricultural practices in relation to seed potato production and effects of environmental factors on aphid population dynamics (Robert *et al.*, 2000). The initial level of crop infection plays a central role in the spread of PLRV, being aphid-borne (Robert *et al.*, 2000). Secondary spread of PLRV from primary sources accounts for a major portion of PLRV infection (Thomas *et al.*, 1997). Primary sources of PLRV

inoculum originate in the potato crop from infected seed tubers or by introduction by migratory alatae (winged morphs of aphids) (Thomas *et al.*, 1997; Johnson and Pappu, 2006).

Migratory alatae colonize crops in the spring, known as 'contamination' or 'emigration' flight, usually around the time potato shoots emerge above ground (Robert *et al.*, 2000). Emigration flight takes place after hibernation of *M. persicae*, when alatae fly from their winter hosts to potatoes (Hille Ris Lambers, 1955). The species can hibernate as eggs on peach trees or as viviparae on various herbaceous plants in mild climates (Hille Ris Lambers, 1955). When alatae arrive on potato crops, non-viruliferous aphids can acquire PLRV from infected plants and transmit them to further plants, demonstrating the importance of the initial level of crop infection (Robert *et al.*, 2000). The degree to which PLRV spreads in potatoes depends on the time and extent at which aphids arrive on the crops and on their subsequent activity within crops (Fisken, 1959).

After the influx of alatae aphids in the spring, apterous (non-winged morphs) populations build up on the plants (Robert *et al.*, 2000). Apterous offspring account for a spread of the virus to the remainder of the crop (Thomas *et al.*, 1997). These non-winged morphs have difficulty crossing an open space without a foliage bridge, so the virus is spread in expanding concentric areas around a primary inoculum, and may be rectangular following the rows of potato plants (Thomas *et al.*, 1997). Many aphids reach plants nearest to the central source but few reach maximum distances, resulting in short-distance transport of PLRV by aphids borne on already diseased plants (Hille Ris Lambers, 1955; Thomas *et al.*, 1997). From apterous populations of aphids, winged forms develop and disperse in the summer, called 'dissemination' flight (Robert *et al.*, 2000).

The secondary spread of PLRV by aphids can be limited by cultivation practices during seed-potato growing (Hille Ris Lambers, 1955). The practice of lifting seed potatoes early to prevent infection has become an important practice (Hille Ris Lambers, 1955). This means that the potatoes must be lifted before aphids become too dangerous and even a delay of a few days could mean heavy infection with PLRV (Hille Ris Lambers, 1955).

The environment affects aphid population dynamics and this has an influence on their effectivity as vectors for virus spread (Robert *et al.*, 2000). Temperature is particularly important in influencing aphid behavior during acquisition and/or inoculation (Robert *et al.*, 2000). Temperature may also affect plants as sources and new virus hosts and have an effect on virus survival (Robert *et al.*, 2000).

Climatic conditions affect aphid numbers as winter temperatures limit the survival of vector species and determine the date and size of the emigration flight (Gabriel, 1965; Robert *et al.*, 2000). Outbreaks of PLRV are associated with periods when the winters are milder than usual enabling aphids to survive in larger numbers on over wintering hosts (Robert *et al.*, 2000). Cold weather decreases the rate of aphid reproduction, while warmer conditions favor aphid

multiplication (Fisken, 1959). On the other hand, prolonged high temperature can cause reproduction to decline and cease completely (Gabriel, 1965). Optimum development of apterous aphid populations occurs within the range 20-25°C and is negatively affected at 30°C (Gabriel, 1965). Even when the initial level of virus infection in crops is low, any change in climatic conditions and aphid numbers can result in dramatic increases in virus incidence (Robert *et al.*, 2000).

Higher temperature does not only influence the aphid numbers, but also influences the acquisition feeding period considerably (Syller, 1987). PLRV acquisition by *M. persicae*, i.e. the time taken to acquire virus (Kassanis, 1952), is increased with higher temperature which in turn would increase the proportion of infected potato plants (Syller, 1987). Higher temperature reduces the latent period of PLRV in *M. persicae*, probably by increasing the speed with which the virus moves from the gut to the salivary system via the haemolymph (Syller, 1987). Therefore, the proportion of viruliferous aphids increases with higher temperature rather than the ability of individual aphids to transmit PLRV (Syller, 1987). However, at higher temperatures aphids are more active and more often retract their stylets and re-penetrate the phloem than at lower temperatures which increase the probability that a plant will become infected (Syller, 1987).

Moreover, temperature does not only influence aphid population dynamics and behaviour during acquisition and/or feeding, but may also affect plants as virus sources or hosts (Robert *et al.*, 2000). High temperature at PLRV inoculation seems to increase the proportion of infected plants, but not significantly so (Syller, 1987). There is a possibility that higher temperature can make potato plants more susceptable to PLRV by lowering the resistance of the plants to infection or to virus multiplication or to both (Syller, 1987). Therefore, when taken together it becomes that clear that temperature considerably influences the dynamics of the PLRV transmission process (Syller, 1987).

Knowledge of the transmission characteristics of PLRV and its epidemiology is very important for the successful use of insecticides in preventing virus spread in the field (Robert *et al.*, 2000). Current seasonal spread of PLRV from one plant to another within a crop is entirely by aphids so that insect control is one of the most important means of reducing the spread (Johnson and Pappu, 2006). An aphid can acquire PLRV after a few minutes of feeding, but the incubation period of 24 to 48 hours before the aphid is able to transmit the virus, helps control virus spread if there are timely applications of insecticides (Johnson, 2003).

Managing aphids is one of a combination of means to manage PLRV, others include: planting PLRV free seed tubers, isolation from other potato fields, roguing out diseased plants to reduce sources of inoculum, eliminating refuse tubers, volunteer potatoes and weeds as well as early harvesting or vine killing of seed plots and certified fields (Rich, 1983; Johnson and Pappu, 2006) and planting virus resistant cultivars (Thomas *et al.*, 1997; Robert *et al.*, 2000).

The most important control measure for PLRV is to plant certified seed free from the virus (Rich, 1983; Johnson, 2003). Planting seed tubers that are free of PLRV is the only practical means of reducing the number of chronically infected plants, only seed tubers certified to be free or nearly free of PLRV should be used (Johnson and Pappu, 2006). Infected potato tubers used for seed are a main source of infection in commercial potato fields, contributing to the spread of PLRV (Kawchuck *et al.*, 1990; Johnson and Pappu, 2006).

When there are infected potato plants in the field, they should be rogued and removed and volunteers, which can act as an inoculum reservoir, such as weeds and mustards should be controlled (Johnson, 2003). Volunteer potatoes represent an important source of PLRV and volunteer plants can be destroyed by either rotating fields out of potatoes for one or more years, or by treating the potato crop with maleic hydrazide to reduce sprouting of tubers the following season (Johnson and Pappu, 2006). Scouting for PLRV infected plants as well as aphid vectors should begin early in the season and continue until mid to late season as late season infection can still result in tuber symptoms (Johnson and Pappu, 2006).

2.1.2 Pathogenesis

The most important symptom of PLRV infection of potatoes is that it reduces yield. PLRV infected plants have fewer tubers and these have a smaller size (McKay, 1955). Even though the yield of PLRV plants fluctuates with the seasons, it is never that of healthy plants (McKay, 1955).

Visual symptoms from potato plants infected with PLRV include leafrolling and stunting, the extent depending on the potato cultivar (Taliansky *et al.*, 2003). When these plants are shaken, their leaves make a sound similar to the rattling of parchment paper (Johnson, 2003). Tuber symptoms of some varieties include phloem net necrosis (net necrosis), which is small brown strands of discolored tissue extending throughout the stem end of the tuber after a month in storage (Johnson, 2003; Rich, 1983). Certain tubers, especially those that also exhibit net necrosis, develop long, spindly sprouts but this symptom is not unique to PLRV infection (Rich, 1983).

Potato plants can be infected with PLRV either by virus transmission to a healthy plant or by growing from an infected tuber. These differences in PLRV acquisition are distinguished as two phases of the disease, namely: a) Primary Leaf Roll and b) Secondary Leaf Roll (McKay, 1955).

Primary Leaf Roll (also known as current season infection) occurs when virus transmission to a healthy plant takes place during the growing season, usually through the agency of insects (McKay, 1955; Taliansky *et al.*, 2003). The symptoms are usually not severe unless plants become infected early in the season (Taliansky *et al.*, 2003). Early season infection usually results in a rolling of the top leaves, which have a purple-reddish appearance (Johnson, 2003;

Rich, 1983). The top leaves become erect and their margins turn upward and curl inwards toward the midrib, being tough and leathery (McKay, 1955). These symptoms only occur later in the growing season in plants that appeared to be healthy (Johnson, 2003). The lower leaves of the plant may be normal looking, making these symptoms less obvious (Johnson, 2003).

The rolled erect leaflets would be confined to the topmost parts of the stems unless infection took place before the flowering of the plant, in which case the lower leaflets would also become affected. This condition reflects secondary symptoms, but usually does not take place (McKay, 1955). If the plants become infected after their flowering period or towards maturity, no signs of infection occur (McKay, 1955). These plants are by no means free of PLRV, but their tubers will give rise to Secondary Leaf Roll plants the following season (McKay, 1955). Plants that become infected late in the growing season would also not show any signs of infection (Rich, 1983). There would be little if any dwarfing and the infected plant usually remains symptomless (Rich, 1983).

Tubers from plants infected during the current season develop net necrosis in storage (Johnson and Pappu, 2006; Rich, 1983). This appears as a discoloration of the tissue about a centimeter beneath the skin in the form of fine black strands radiating from the stem-end region of the tuber to its circumference (Johnson and Pappu, 2006; McKay, 1955). The degree of net necrosis depends on the potato cultivar, time of infection, length and temperature of potato storage (Johnson and Pappu, 2006). Net necrosis is a variable feature of PLRV because when such infected tubers are planted their progeny seldom show signs of net necrosis (McKay, 1955).

However, once a tuber is infected all its progeny will be diseased. Tubers are usually already infected when Primary Leaf Roll symptoms develop because these symptoms become noticeable until twenty days after infection, but the virus reaches the tubers within ten to fifteen days after infection (McKay, 1955).

Secondary Leaf Roll develops in plants that are grown from infected tubers (Taliansky *et al.*, 2003). This is the commonest phase of the disease, which shows the most easily recognized symptoms scattered indiscriminately throughout the field (McKay, 1955). The symptoms begin to appear when the plants are about a month old, firstly on the lower leaves and gradually progressing upward (Johnson, 2003; McKay, 1955). The leaves of the plants, especially the lower leaves, are stiff, roll upwards, become leathery in texture and rattle when shaken (Johnson, 2003; Johnson and Pappu, 2006). They break easily when crushed and may be chlorotic (Taliansky, 2003). Rolled, thickened lower leaflets are the one invariable symptom of PLRV that are used to distinguish the disease from other potato maladies where rolling of the foliage occurs (McKay, 1955).

The rolling of foliage in PLRV infected plants is due to the accumulation of starch in the leaves. Normally, starch that is manufactured during daylight is converted into sugar at night and carried away through the conducting tissues of the petiole and stem. These sugars are either used for

growth or reconverted to starch in the growing tuber. However, in leaves of PLRV plants, starch manufacture goes on, but the normal translocation process is disrupted. This results in the leaflets becoming congested with starch and curving upward (McKay, 1955).

Seeing that the starch is largely retained in the leaves, it makes sense that tuber formation must suffer in Secondary Leaf Roll plants (McKay, 1955). There are less tubers and they are smaller which results in a marked yield loss (Rich, 1983). In addition, plants emerging from the infected tubers are usually yellow to pale green and may be slightly stunted (Johnson, 2003; Johnson and Pappu, 2006). The diseased plants grow slowly and the tubers set close to the main stem while the old mother tuber also remains firm until digging time (McKay, 1995). It is these chronically infected plants which will provide sources of virus for spread to healthy plants during the current growing season (Johnson and Pappu, 2006).

Due to the economic importance of PLRV, more has been learnt about the infected plants with the use of light microscopy and to some extent with the electron microscope (Shepardson *et al.*, 1980). That enables us to view the pathogenesis of the plant on a smaller scale. Light microscopy studies can reveal the characteristic patterns of phloem necrosis in PLRV infected potato plants while electron microscopy can reveal the pathological effects induced by the virus in the phloem (Shepardson *et al.*, 1980). The focus of these studies of PLRV infected potato plants have been mainly on phloem tissue because the virus is mostly confined to phloem cells (Johnson and Pappu, 2006; Van Regenmortel *et al.*, 2000). As the virus is mostly confined to the phloem, it explains why one of the primary pathological changes in the infected plant is necrosis of the phloem causing death and collapse of cells (Shepardson *et al.*, 1980).

Phloem is a tissue that conducts food materials in vascular plants (plants possessing organized tissue to conduct water and nutrients, consisting of xylem and phloem) from regions where it is produced to regions where they are needed (Beckett *et al.*, 2000). It consists of parenchyma cells, sieve elements and companion cells (Beckett *et al.*, 2000; Raven and Johnson, 2002). The parenchyma cells form part of the phloem tissue, these cells are relatively undifferentiated with air spaces between them frequently (Beckett *et al.*, 2000). Sieve elements are elongated cells that form hollow tubes (sieve tubes), these cells contain little cytoplasm and no nucleus (Beckett *et al.*, 2000). Closely associated with sieve elements are companion cells (Beckett *et al.*, 2000). The function of companion cells is uncertain, though it appears to regulate the activity of the adjacent sieve element (Beckett *et al.*, 2000).

Phloem tissue, mature sieve elements, companion cells and plasmodesmata between sieve elements and companion cells, contain virus-like particles in potato plants infected with PLRV as revealed by electron microscopic studies (Shepardson *et al.*, 1980; Van den Heuvel *et al.*, 1995). Plasmodesmata are fine cytoplasmic strands that connect adjacent plant cells by passing through their cell walls (Beckett *et al.*, 2000). According to Shepardson *et al.* (1980), evidence of infection was localized in sieve elements and companion cells. Phloem parenchyma

cells would only contain virus particles occasionally and in their vacuoles without showing any other cytoplasmic changes (Shepardson *et al.*, 1980).

Viral antigen occurred the most abundantly in companion cells as found by Van den Heuvel *et al.* (1995), and it was quite unevenly distributed within the leaves examined. The uneven distribution of PLRV in phloem tissue suggests that systemic transport of the virus in heavily virus-infested vessels is considerably impaired (Van den Heuvel *et al.*, 1995). Van den Heuvel *et al.* (1995), also found that only sieve elements connected to infected companion cells contained PLRV antigen.

Phloem necrosis in infected potato plants appears before the other external symptoms such as leafrolling is visible (Shepardson *et al.*, 1980). In Primary Leaf Roll, phloem necrosis appears to move only upwards at first, enabling the point of infection to be determined by the lowest level of leafrolling (Shepardson *et al.*, 1980). In Secondary Leaf Roll, necrosis is first noted in primary phloem strands a few inches above the tuber and then spreads both upwards and downward in the plant (Shepardson *et al.*, 1980).

In sieve elements of the necrotic phloem an abnormally large amount of callose is deposited which most probably obstructs phloem transport (Shepardson *et al.*, 1980; Van den Heuvel *et al.*, 1995). The amount of excess callose is used as an indicator of PLRV infection after staining the tissue with resorcin blue or aniline blue (Shepardson *et al.*, 1980). As the translocation process is inhibited from phloem necrosis, plant growth is slowed resulting in dwarfing as seen in Secondary Leaf Roll infected plants (Van Regenmortel *et al.*, 2000).

However, PLRV is not limited exclusively to phloem tissue in infected potato plants, but is also found in mesophyll cells neighbouring minor phloem vessels (Van den Heuvel *et al.*, 1995). Mesophyll cells compose the internal tissue of a leaf blade (Beckett *et al.*, 2000). These infected cells are always found to be immediately adjacent to companion cells in minor veins protruding into the mesophyll tissue (Van den Heuvel *et al.*, 1995). The veins of a plant are made up out of the xylem and phloem, so the infected companion cells of the phloem would infect the rest of the internal tissue of a leaf (Beckett *et al.*, 2000). Even though the highest concentrations of PLRV are detected in phloem cells, viral RNA is also present in almost all cells of infected potatoes (Kaniewski and Thomas, 2004).

Plant viruses would typically, after infection, replicate and spread from the point of infection from cell to cell through the plasmodesmata until they reach the phloem (Taliansky *et al.*, 2003). Once the virions have entered the host vascular system, the different types of phloem-associated cells such as phloem parenchyma and companion cells are invaded (Taliansky *et al.*, 2003). The virions also penetrate into the sieve elements, move through them from leaf to leaf to exit, and infect non-vascular tissues (Taliansky *et al.*, 2003). This is called phloem associated long-distance movement of which less is known than short-distance movement (cell-to-cell movement) (Taliansky *et al.*, 2003). Short-distance movement usually involves one or

more virus-encoded movement proteins as well as certain host components (Taliansky *et al.*, 2003). Long-distance and short-distance movement represents two phases in virus infection.

Short-distance mesophyll cell-to-cell movement is not observed with PLRV (Taliansky *et al.*, 2003; Van den Heuvel *et al.*, 1995). PLRV can only spread via long-distance transport, therefore their accumulation is limited mainly to the phloem cells (Taliansky *et al.*, 2003). Consequently, its spread resembles only one phase of the spread of other plant viruses, namely phloem-associated long-distance movement (Taliansky *et al.*, 2003).

It is in the phloem companion cells where the virus mostly replicates, but mesophyll cells also support PLRV replication to a similar extent (Taliansky *et al.*, 2003; Van den Heuvel *et al.*, 1995). The virus particles associate with various membranes in the companion cells including the membranes of mitochondria and chloroplasts as well as the tonoplast (Shepardson *et al.*, 1980). However, minus-strand RNA has been detected in nearly all cells of infected plants, which shows that PLRV also multiplies there (Kaniewski and Thomas, 2004).

Once PLRV has infected phloem cells, various forms of cell degeneration can be observed. In companion cells, an early indication of virally induced cellular disturbance is dilation of mitochondrial cristae followed by the appearance of vesicles in the parietal cytoplasm (Shepardson *et al.*, 1980). Furthermore, in infected cells the nucleus becomes deficient in chromatin and its nucleolus very dense (Shepardson *et al.*, 1980). The cytoplasmic ribosome content reduces while mitochondria loose much of their matrix and become swollen to several times their normal size (Shepardson *et al.*, 1980). Mitochondrial cristae separate from the envelope and clump while the cell membrane and nuclear envelope degenerates and completely disappears (Shepardson *et al.*, 1980). In the chloroplasts, starch and lipid globules are deposited while chlorophyll content is reduced (Shepardson *et al.*, 1980; Van Regenmortel *et al.*, 2000).

2.1.3 The molecular biology of PLRV

2.1.3.1 Genomic organization

Having discussed PLRV transmission as well as the effects that the virus has on plants when infecting it, attention is directed to the molecular biology of the virus to understand PLRV in greater depth. PLRV, like other plant viruses, are extremely small and cannot be seen without the use of an electron microscope (Johnson, 2003; Johnson and Pappu, 2006). PLRV particles are thought to have 180 subunits arranged in a $T = 3$ icosahedron, creating a hexagonal outline (Van Regenmortel *et al.*, 2000). The isometric particles are without an envelope and measure about 23-25 nm in diameter (Shepardson *et al.*, 1980; Rich, 1983; Garret *et al.*, 1993; Van Regenmortel *et al.*, 2000; Taliansky *et al.*, 2003).

Generally speaking, viruses can be described as consisting of nucleic acid, DNA or RNA, surrounded by a coat protein (Johnson, 2003). Plant viruses may contain any one of the four types of genetic material: single-stranded RNA (ssRNA), double-stranded RNA, single-stranded DNA or double-stranded DNA (Bustamante and Hull, 1998). The vast majority of plant viruses (about 75%) have ssRNA of the (+) or messenger polarity (termed (+) RNA), showing a wide variation in capsid morphology (Bustamante and Hull, 1998). In the case of PLRV, particles are composed of two proteins, one major (*ca.* 23 kDa) and one minor (*ca.* 80 kDa) protein (Van Regenmortel *et al.*, 2000; Reavy and Mayo, 2002; Taliansky *et al.*, 2003) and a positive sense ssRNA of 5882 nucleotides (Mehrad *et al.*, 1979; Van der Wilk *et al.*, 1989; Bahner *et al.*, 1990; Kawchuk *et al.*, 1990; Van Regenmortel *et al.*, 2000; Taliansky *et al.*, 2003).

The genomes of plant viruses may have different terminal structures such as cap structures or genome-linked proteins (virus protein, genome-linked; VPg) at the 5' end and a poly(A)-tail or tRNA-like structure at the 3' end of their RNA (Bustamante and Hull, 1998). The genome of PLRV has neither a 5'-cap nor 3'-terminal poly(A) tract or a 3' tRNA-like structure, but carries a small protein (VPg) at the 5' end (Martin *et al.*, 1990; Taliansky *et al.*, 2003). At the 3' end the genome of PLRV has an –OH as judged by the ability to readily polyadenylate with *E. coli* polymerase (Martin *et al.*, 1990). The VPg has an estimated size M_r 7000 covalently linked to the 5'-end of the genomic RNA (Mayo and Ziegler-Graff, 1996; Van der Wilk, 1997b; Van Regenmortel *et al.*, 2000).

The molecular and structural features of the virion, together with its transmission properties, classify PLRV in the family *Luteoviridae* (Taliansky *et al.*, 2003). The name Luteoviridae is derived from Latin "luteus", which means yellow, since all original members of the group caused yellowing symptoms in their hosts (Martin *et al.*, 1990). The family *Luteoviridae* was created so that molecular differences in replication-related gene sequences could be recognized by separation at the genus level (Mayo, 2002). Within the family *Luteoviridae*, there are two main genera (*Luteovirus* and *Polerovirus*) that differ in certain genome features and type of RNA-dependent RNA polymerase (RdRp) (Van der Wilk *et al.*, 1997b; Van Regenmortel *et al.*, 2000; Taliansky *et al.*, 2003). PLRV is the type species of the genus *Polerovirus* (derived from potato leafroll) (Rouzé-Jouan *et al.*, 2001; Mayo, 2002; Taliansky *et al.*, 2003; Johnson and Pappu, 2006). The genus *Polerovirus* consequently contain viruses with genomes similar to those of PLRV in that they contain a P0 gene, have an extensive overlap between the P1 and P2 genes, and have a 5'-linked VPg (Mayo, 2002).

The RNA genome of PLRV contains six large open reading frames (ORFs) (ORF0 – ORF5), and the coding sequences are separated into two clusters of three genes by an noncoding intergenic region of 200 nucleotides (Fig. 2.2) (Bahner *et al.*, 1990; Mayo and Ziegler-Graff, 1996; Van der Wilk *et al.*, 1997b; Van Regenmortel *et al.*, 2000; Haupt *et al.*, 2005). Common sequences are found in these intergenic regions between different luteoviruses that are likely to

include subgenomic RNA promoter signals (Martin *et al.*, 1990). The ORF0, 1 and 2 region is divergent among luteoviruses while the ORF3, 4 and 5 region is the conserved gene cluster (Martin *et al.*, 1990; Ashoud *et al.*, 1998). The apparent illogicality of naming an ORF as zero, is for the sake of promoting consistency between the two subgroups of luteoviruses as equivalent ORFs have been assigned different numbers in different luteovirus genomes (Mayo and Ziegler-Graff, 1996). The 5'-block of coding sequence, consisting of three ORFs, overlap extensively (Mayo and Ziegler-Graff, 1996). ORF0 overlaps ORF1, while ORF1 overlaps ORF2 by 298 nt (Martin *et al.*, 1990; Van Regenmortel *et al.*, 2000). In the 3'-block of coding sequence there is also some degree of overlap between the other three ORFs; ORF4 is contained completely within ORF3, however ORF5 is positioned directly downstream of and contiguous with ORF3, separated by an amber termination codon (Martin *et al.*, 1990; Van Regenmortel *et al.*, 2000; Taliansky *et al.*, 2003).
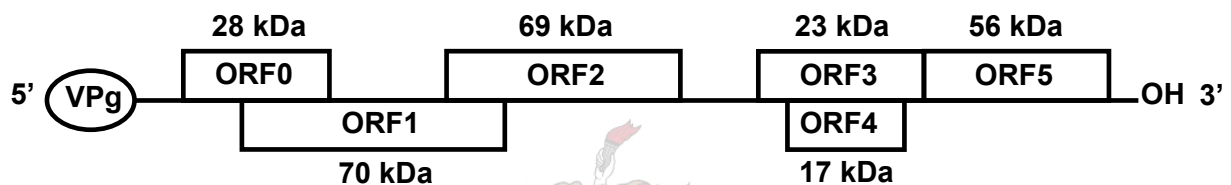
Fig. 2.2: Schematic representation of the genomic organization of PLRV. The diagram shows the arrangement of the open reading frames in the RNA genome of PLRV (5882 nt). Boxes represent the ORFs; solid lines represent untranslated sequences in the RNA and the circle represents the VPg. Numbers outside the boxed areas show the M_r values in kDa of the corresponding proteins encoded by each ORF (figure composed from Van der Wilk *et al.*, 1989; Martin *et al.*, 1990; Rhode *et al.*, 1994; Van der Wilk *et al.*, 1997a; Van der Wilk *et al.*, 1997b; Van Regenmortel *et al.*, 2000).

The 5' half of the luteovirus genomes encode the nonstructural genes presumed to be involved in virus replication within infected plant cells whereas the structural genes are located in the 3' half of the genome (Martin *et al.*, 1990). PLRV 3' structural genes include the coat protein (CP) ORF, an ORF embedded in the CP gene postulated to be the VPg and the CP readthrough also detected on the surface of the virion particle (Martin *et al.*, 1990). These structural genes determine the particle morphology, serological cross activity and possibly virus-vector interaction (Martin *et al.*, 1990).

The predicted sizes of the ORF protein products are as follows: ORF0 = 28 kDa, ORF1 = 70 kDa, ORF2 = 69 kDa, ORF3 = 23 kDa, ORF4 = 17 kDa and ORF5 = 56 kDa (Van der Wilk *et al.*, 1989; Martin *et al.*, 1990). In addition to the 200 nt intergenic noncoding region, there is a 5' noncoding region of 71 nt and a 3' noncoding region of 141 nt (Martin *et al.*, 1990). The non-coding sequences of PLRV are approximately 6.9% of the sequence (Van der Wilk *et al.*, 1989). ORF0, ORF1 and ORF2 are translated from the genomic RNA, while ORF3, ORF4 and ORF5 are expressed from a subgenomic RNA (sgRNA) (Ashoub *et al.*, 1998; Van Regenmortel *et al.*, 2000).

Within the 5'-located ORFs, ORF0 encodes a potential silencing suppressor protein while ORF 1 and ORF2 have motifs characteristic of helicases (ORF1) and polymerases (ORF2) to form part of the viral replicase complex (Ashoub *et al.*, 1998; Haupt *et al.*, 2005). ORF2 is translated by frameshift from ORF1 upstream of the termination of ORF1, therefore the ORF2 product shares an amino acid terminus with the product of ORF1 (Ashoub *et al.*, 1998; Van Regenmortel *et al.*, 2000). Within the 3'-located gene cluster ORF3 encodes the major capsid protein and ORF4, contained within ORF3 in a different frame, encodes a movement protein (Haupt *et al.*, 2005). Initiation of ORF4 transcription takes place at an internally located AUG codon within the CP gene product of ORF3 (Ashoud *et al.*, 1998). ORF5 is expressed by occasional translational readthrough of the amber termination codon to form a minor capsid protein (Haupt *et al.*, 2005). The minor capsid protein expressed by ORF3/ORF5 readthrough is reported to be the aphid transmission factor (Ashoud *et al.*, 1998).

In the case where reference is made to the encoded peptides, it would be assigned the same number as the ORF encoding them, e.g. P0 encoded by ORF0 (Mayo and Ziegler-Graff, 1996). With the exception of the structural proteins, little is known about the functions of the product of the ORFs present on the genome of PLRV, mostly the functions are speculative (Van der Wilk *et al.*, 1997a; Taliansky *et al.*, 2003). For instance it is still obscure which ORF encodes the VPg, and concerning the nonstructural genes only the RdRp (by sequence comparison) and the movement protein have been identified (Van der Wilk *et al.*, 1997a). There are also no data on post-translational modification of the virus proteins (Van Regenmortel *et al.*, 2000). However, the current knowledge of the various ORFs and their products follow as research in this area has produced some insights.
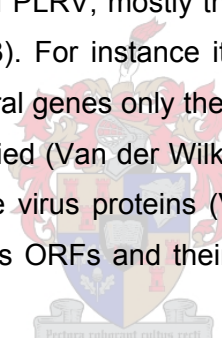
2.1.3.2 ORF0

ORF0 starts at the first AUG codon (position 70) and terminates with a UGA stop codon at position 811 to encode a product of 28 kDa (Van der Wilk *et al.*, 1989). The role of ORF0 has not been resolved, but expression of this ORF in transgenic potato plants has been shown to induce viral disease-like symptoms (Van der Wilk *et al.*, 1997b). P0 has consequently been assigned as being involved in symptom development and as a suppressor of gene silencing (Mayo and Ziegler-Graff, 1996; Taliansky *et al.*, 2003). It is also suggested that P0 plays a role in host recognition or in viral symptom expression (Van der Wilk *et al.*, 1997a). P0 is only expressed in *Poleroviruses* and not in *Luteoviruses*, which means that either its function is not needed by *Luteoviruses* or that it is performed by one of the other proteins as an additional function (Mayo and Ziegler-Graff, 1996).

Homology studies revealed little similarity in the amino acid sequences of luteoviral P0 proteins so not much can be deduced as to the function of P0 from the amino acid sequences (Van der Wilk *et al.*, 1989; Mayo and Ziegler-Graff, 1996). However, the N-terminal regions of P0 are markedly hydrophobic and have shown a weak homology with several membrane associated

proteins (Mayo and Ziegler-Graff, 1996; Van der Wilk *et al.*, 1997a). Analysis of the amino acid sequence of PLRV P0 revealed a putative membrane-binding site between residues 21 and 32 (Van der Wilk *et al.*, 1997a).

Comparisons made between the genomic organization of PLRV and other luteoviruses revealed that PLRV is very similar to BWYV (Van der Wilk *et al.*, 1997a). When the encoded proteins of PLRV and BWYV were compared, it revealed that all the viral proteins share a high homology in amino acid sequence except for the ORF0s (Van der Wilk *et al.*, 1997a). Even though P0 of both viruses is similar in size and position on the genomes, the primary structures are not similar and there is no indication that BWYV P0 is membrane-linked like PLRV P0 (Van der Wilk *et al.*, 1997a). From an ecological point of view, PLRV differs from BWYV mainly in its host ranges, PLRV being able to infect only a limited number of plant species while BWYV can infect many different plant species (Van der Wilk *et al.*, 1997a). It has therefore been suggested that P0 plays a role in host recognition since P0 constitutes the main genetic difference between the viruses (Van der Wilk *et al.*, 1997a).

It has also been suggested that P0 funtions as a protease to cleave the 17 kDa product of ORF4 to give the 7 kDa functional VPg (Martin *et al.*, 1990). However, nucleic acid hybridizations and serological tests have contradicted this proposal for luteoviruses in general (Martin *et al.*, 1990). It could be that this posttranslational processing occurs in some luteoviruses to give a functional VPg (Martin *et al.*, 1990). Another proposal is that P0 is involved in the movement of the virus in conjunction with P4 (Van der Wilk *et al.*, 1997a). The precise function of P0 remains unknown and if the protein is indeed involved in determining the host range of PLRV, it may act as an early gene making it ephemeral and difficult to detect (Mayo and Ziegler-Graff, 1996).

2.1.3.3 ORF1

ORF1 overlaps with ORF0 by a start at position 203 in a different reading frame from ORF0 and stops at the UGA codon present at position 2120 to encode a protein of 70 kDa (Van der Wilk *et al.*, 1989). The ORF1 product P1 contains motifs characteristic of serine-like proteinases and has been classified among the poliovirus 3C-like proteases (Mayo and Ziegler-Graff, 1996; Van der Wilk *et al.*, 1997a; Van der Wilk *et al.*, 1997b). However, the role of this putative protease in the viral infection cycle is still concealed (Van der Wilk *et al.*, 1997a). It has been suggested that ORF1 contains the sequence domains of VPg-protease-polymerase (Mayo and Ziegler-Graff, 1996). This means that ORF1 has proteinase functions and encodes the VPg (Mayo and Ziegler-Graff, 1996; Taliansky *et al.*, 2003). VPg seems to be a product of ORF1 (Mayo and Ziegler-Graff, 1996) rather than that of ORF4, which results from the cleavage of P4 by P0 as previously suggested. Amino acid comparisons revealed that the N-terminal amino acids of VPg are located downstream of the putative protease domain and upstream of the polymerase which comprises part of ORF2 (Van der Wilk *et al.*, 1997b). The VPg N-terminal amino acid sequence

has been mapped to position 400 of the PLRV ORF1 product and its C-terminus approximately at residue 465 of the ORF1 product (Van der Wilk *et al.*,1997b).

Maturation of the VPg requires proteolytic cleavage at both the N- and C-terminus as it is contained in the P1-P2 fusion protein (Van der Wilk *et al.*, 1997b). PLRV VPg is most likely released from P1 by proteolytic activity of its putative protease domain (Van der Wilk *et al.*, 1997b). The VPg N-terminal proteolytic processing site consists of the residues E-S/T as the N-terminal residue of VPg (S/T) is preceded by a glutamic acid residue (Van der Wilk *et al.*, 1997b). Based on the size of the protein, there have been attempts to predict the position of the C-terminal processing site but a sequence similar to the putative N-terminal processing site could not be found in the immediate vicinity of the putative VPg C-terminus (Van der Wilk *et al.*, 1997b).

2.1.3.4 ORF2

ORF2 is proposed to start at position 1540, overlapping ORF1, and to terminate at the UGA stop codon at position 3388 to encode a protein of 69 kDa (Van der Wilk *et al.*, 1989). ORF2 does not encode a separate gene product but is rather expressed as a fusion with the product of ORF1 through a -1 translational frameshift to code the putative RdRp (Martin *et al.*, 1990; Van der Wilk *et al.*, 1997b; Taliansky *et al.*, 2003). As a frameshift takes place, the stop codon at the 3' end of ORF1 is bypassed (Martin *et al.*, 1990).

The P2 protein contains the putative RdRp motif near the C terminus: Gly-xxx-Thr-xxx-Asn(x_{25_40})Gly-Asp-Asp (Mayo and Ziegler-Graff, 1996; Van der Wilk *et al.*, 1997a). This motif has been found in all RNA-dependent polymerases of RNA plant viruses sequenced to date, which makes P2 the most likely candidate to represent the PLRV-encoded RdRp (Van der Wilk *et al.*, 1989). Furthermore, the predicted amino acid sequence of PLRV P2 shows considerable homology with the putative RdRps of other viruses (Van der Wilk *et al.*, 1989). P2, together with P1, has been found to be absolutely necessary for replication in another luteovirus (BWYV), which give further give evidence for the polymerase activity of P2 (Mayo and Ziegler-Graff, 1996).

2.1.3.5 ORF3

ORF3 is separated from ORF2 by a non-coding region of 197 nucleotides (Van der Wilk *et al.*, 1989). It spans from position 3588 to 4212 (UAG), hence encoding 208 amino acids to give a 23 kDa product (Bahner *et al.*, 1990; Kawchuk *et al.*, 1990; Van der Wilk *et al.*, 1989). Only the CP gene has been unequivocally assigned to ORF3 present in the 3'-half of the non-coding sequence, followed in frame by ORF5 (Bahner *et al.*, 1990; Van der Wilk *et al.*, 1997a; Van Regenmortel *et al.*, 2000). P3 was shown to be the PLRV CP largely by immunodetection of a fusion protein with antiserum prepared against whole virions of PLRV (Martin *et al.*, 1990; Mayo and Ziegler-Graff, 1996). There is considerable homology between the ORF3 of PLRV and

other luteoviral ORFs that encode their corresponding CPs that confirm the assignment of P3 as the CP of PLRV (Van der Wilk *et al.*, 1989). The CP of luteoviruses has been reported to be responsible for serological properties and transmission specificity (Kawchuk *et al.*, 1990). The strong homology between the luteoviral CPs can explain the serological cross-reactivity between the different luteoviruses (Van der Wilk *et al.*, 1989).

2.1.3.6 ORF4

ORF4 underlies ORF3 from position 3613 to the UGA codon at position 4081 (Van der Wilk *et al.*, 1989). ORF4 encodes a protein of 17 kDa in a different frame from ORF3 (Van der Wilk *et al.*, 1989; Martin *et al.*, 1990). It has been shown that translational initiation efficiency at the PLRV 17 kDa AUG codon is sevenfold higher than initiation for CP (P3) synthesis (Rhode *et al.*, 1994).

Different functions have been assigned to P4 in the literature; P4 can either be the VPg or a movement protein. The proposal of P4 as the VPg is based on the similarity in size to the VPg isolated from another luteovirus (the PAV serotype of *Barley yellow dwarf virus*; BYDV-PAV) (Martin *et al.*, 1990). Furthermore, the amino acid sequence of PLRV P4 has considerable homology with the products of two other luteoviruses, *Barley yellow dwarf virus* (BYDV) (57%) and BWYV (72%) in analogous ORFs (Van der Wilk *et al.*, 1989). The ORFs in BYDV and BWYV also underly their respective CP ORFs (Van der Wilk *et al.*, 1989). It has been proposed that both of these ORFs code for the respective VPgs (Van der Wilk *et al.*, 1989). However, the VPg of PLRV has been estimated to have a molecular mass of 7 kDa whereas ORF4 has a coding capacity of 17 kDa (Van der Wilk *et al.*, 1989). The possibility exists that ORF4 encodes a VPg-precursor molecule from which the VPg molecule is released at the onset of RNA synthesis (Van der Wilk *et al.*, 1989). This method of VPg synthesis has also been suggested for another plant virus, cowpea mosaic virus.

There is more reference in the literature to the probability that ORF4 encodes a movement protein. P4 most probably constitutes the viral movement protein since it has been shown to be required for long distance movement in poleroviruses and is indispensable for systemic infection of plants (Van der Wilk *et al.*, 1997a; Van Regenmortel *et al.*, 2000). Indirect evidence for PLRV P4 as the movement protein comes from the observation that its analogue in BYDV-PAV was shown by mutational analysis to be a movement protein (Taliansky *et al.*, 2003). Direct evidence for the role for PLRV P4 as the movement protein comes from experiments in which two PLRV mutants with either an untranslateble or modified P4 were able to replicate and accumulate in leaves of potato, but were unable to move into vascular tissues and initiate a systemic infection of the plant (Taliansky *et al.*, 2003). This indicates that P4 is strictly required for virus movement, however the requirement for the P4 movement protein has been shown to be host-dependent and that there is a P4-independent mechanism for PLRV movement that operates at least in some plants (Taliansky *et al.*, 2003).

Furthermore, the biochemical properties of the protein also points to its role as a phloem-specific movement protein (Mayo and Ziegler-Graff, 1996). The protein has been shown to bind single-stranded nucleic acids via its binding domain located in the basic C-proximal part of the protein (Rhode *et al.*, 1994; Mayo and Ziegler-Graff, 1996; Van der Wilk *et al.*, 1997a). In addition, the acidic N terminus of P4 forms an amphipatic α-helix with the negatively charged amino acids located predominantly on one side of the helix (Mayo and Ziegler-Graff, 1996). The amphipatic α-helix exhibits the capacity for homodimer formation to mediate protein-protein interaction, which could explain the tendency of P4 to dimerize (Mayo and Ziegler-Graff, 1996). Indeed, PLRV P4 has been detected in infected and transgenic potato plants mainly as a homodimer (Mayo and Ziegler-Graff, 1996). Furthermore, in PLRV infected potato plants P4 was predominantly associated with membrane-enriched fractions and *in vivo* the protein has been found in a phosphorylated form (Mayo and Ziegler-Graff, 1996). It is these biochemical properties which point to the role of the protein as a movement protein together with the striking similarities in properties with the best studied movement protein, that of *Tobacco mosaic virus* (TMV) (Taliansky *et al.*, 2003). Nevertheless, the movement protein of TMV is involved in cell-to-cell movement in the plant vasculature whereas that of P4 is assumed to mediate virus movement only between cells within the phloem tissues (Taliansky *et al.*, 2003).

2.1.3.7 ORF5

ORF5 is contiguous with the putative CP ORF3, separated only be the amber stop codon of ORF3 (Van der Wilk *et al.*, 1989). This ORF is present in the 3'-half of the PLRV genome to make part of the structural proteins and is about 1400 nts (Van der Wilk *et al.*, 1997b; Van Regenmortel *et al.*, 2000). ORF5 is found in the same translational reading frame as the upstream ORF3 (Martin *et al.*, 1990). Expression of ORF5 is a result of suppression of the CP termination codon to produce a readthrough product of ORF3 (Martin *et al.*, 1990; Van der Wilk *et al.*, 1997a; Van Regenmortel *et al.*, 2000; Reavy and Mayo, 2002). The results of readthrough is a protein, termed readthrough protein, that consists of the CP with a carboxy-terminal extension to give a protein with molecular mass of approximately 80 kDa (Bahner *et al.*, 1990; Reavy and Mayo, 2002). The predicted size of the PLRV readthrough protein is the sum of P3 (23 kDa), a shortened form of P5 (53 kDa), and the 4 kDa encoded by the sequence between ORF3 and ORF5, a total of 80 kDa (Bahner *et al.*, 1990; Mayo and Ziegler-Graff, 1996). Readthrough of amber stop codons to produce a low abundance of high-molecular weight proteins have also been reported for other plant viruses (Martin *et al.*, 1990).

The readthrough product of P5 is also a CP like P3, but only a minor CP as it is only occasionally expressed when readthrough of the amber termination codon occurs (Haupt *et al.*, 2005). This ORF3/ORF5 fusion protein is also assembled into virus particles as a minor component together with the main ORF3 component to contribute to the virus particle surface properties of PLRV (Bahner *et al.*, 1990). However, the readthrough protein is not essential for

virus particle formation (Mayo and Ziegler-Graff, 1996; Reavy and Mayo, 2002). Purified virus particles have about 1% to 5% of P5 derived from P3 with a C-terminal extension (Taliansky *et al.*, 2003).

Downstream of the ORF3 termination codon is a succession of proline residues alternating mainly with serine or threonine residues (Bahner *et al.*, 1990; Mayo and Ziegler-Graff, 1996). It has therefore been suggested that the function of this unusual sequence is to separate the 23 kDa (encoded by ORF3) domain of the readthrough protein from the 56 kDa (encoded by ORF5) domain in the assembly of the capsid (Bahner *et al.*, 1990; Mayo and Ziegler-Graff, 1996). The 23 kDa domain of the readthrough protein could assemble with other 23 kDa polypeptides while the 56 kDa domain would thus protrude from the virus particle (Bahner *et al.*, 1990). Such protrusions have been detected on PLRV surfaces in favourably stained preparations (Bahner *et al.*, 1990). Furthermore, antibodies raised against the readthrough protein detected particles in enzyme-linked immunosorbent assay (ELISA) tests indicating that the readthrough protein is exposed on the surface of virus particles (Reavy and Mayo, 2002). Other luteoviruses also have a succession of proline residues immediately downstream of the ORF3 termination codon and express ORF5 in a similar fashion, which means that it may be characteristic of luteoviruses (Bahner *et al.*, 1990).

PLRV readthrough protein may be associated with aphid transmission or virus particle stability (Van der Wilk *et al.*, 1997a; Van Regenmortel *et al.*, 2000). It has also been suggested that PLRV P5 is involved in virus movement, which could mean that there are different functional domains involved in aphid transmission and systemic spread of PLRV (Taliansky *et al.*, 2003).

Evidence for the role of P5 in virus movement, therefore systemic spread, comes from mutants lacking the 3' parts of the P5 gene that were found to be less infective and to accumulate to a lesser extent than unmodified PLRV (Taliansky *et al.*, 2003). Direct evidence for the role of P5 in systemic spread has been obtained from experiments using PLRV tagged with green fluorescent protein (Taliansky *et al.*, 2003).

Evidence for the aphid transmission role of the P5 domain in the readthrough protein comes from the comparison of amino acid sequences of highly transmissible and poorly aphid transmissible isolates of PLRV (Mayo and Ziegler-Graff, 1996). The only amino acid changes between the sequences were located in the C-terminal part of the readthrough domain (Bahner *et al.*, 1990; Reavy and Mayo, 2002). No consistent amino acid changes have been detected that could be responsible for the differences in transmission efficiency (Mayo and Ziegler-Graff, 1996). However, the P3 CP has also been found to have a region that is important in recognition events at different stages of the transmission process (Reavy and Mayo, 2002).

Furthermore, the N-terminal regions of P5 have considerable sequence conservation among luteoviruses whereas the C-terminal region is more diverse among luteoviruses (Mayo and Ziegler-Graff, 1996; Reavy and Mayo, 2002). However, in the variable C-terminal region there is

a stretch of about 45 amino acids that is similar among PLRV, BWYV and cucurbit aphid-borne yellows virus (Mayo and Ziegler-Graff, 1996). This led to the proposal that this region is involved in the specificity of vector transmission since the three viruses are transmitted by *M. persicae* (Mayo and Ziegler-Graff, 1996; Reavy and Mayo, 2002).

As no replication of virus particles take place in *M. persicae*, virus particles remain intact and the determinants of the specificity and efficiency of transmission must be a part of the P3 and/or P5 sequences on the surface of the particles (Mayo and Ziegler-Graff, 1996). It has been suggested that P5 plays a role in the recognition events in the uptake of PLRV particles because of its surface location and accessibility for proteolytic cleavage (Mayo and Ziegler-Graff, 1996). The *M. persicae*-specific motif in P5 could contribute to the high vector specificity required for the movement into the accessory salivary glands (Mayo and Ziegler-Graff, 1996). On the other hand, the conserved N-terminal P5 domain could contain signals involved in the less selective uptake of PLRV particles at the midgut (Mayo and Ziegler-Graff, 1996).

It has also been suggested that symbionin, produced by the endosymbiotic bacterium (*Buchnera* sp.) of *M. persicae*, that protects the virus particles from proteolytic breakdown could bind to P5 (Taliansky *et al.*, 2003). However, the binding of symbionin to virus particles or to the P5 protein has only been demonstrated *in vitro* (Taliansky *et al.*, 2003). P5 has been proposed to play a role in the interactions between the PLRV particles and receptors in the aphid vector at several stages, but the role of P5 is complex and proposed functions full of contradictions (Taliansky *et al.*, 2003). Clarity only exists surrounding mutants that lack P5, which are not aphid-transmissible (Taliansky *et al.*, 2003).

2.1.3.8 Replication and expression strategies

Replication of positive-stranded RNA viruses takes place in the cytoplasm of infected cells in close association with membrane surfaces whereas replication of several other types of viruses occurs in the nucleus (Carrington *et al.*, 1996; Bustamante and Hull., 1998). For phloem-specific luteoviruses, the phloem has to be made use of during replication to ensure aphid transmissibility of progeny virus particles (Rhode *et al.*, 1994). Reproduction takes place by having the host cell reproduce the virus and can be separated into four overlapping steps: (i) uncoating of the virus, (ii) translation of viral RNA, (iii) replication of the genome and (iv) encapsidation of progeny genomic strands (Bustamante and Hull, 1998). Uncoating of the virus takes place to expose the nucleic acid to the replication processes, translation then follows during which the viral RNA serves as a messenger RNA to produce non-structural and structural proteins (Bustamante and Hull, 1998). The proteins required for replication (e.g. the RdRps) are translated first and those with late functions like the coat protein are translated at a later stage (Bustamante and Hull, 1998). Replication of the genome to yield progeny RNA molecules take place in two stages, catalysed by a RdRp: (1) synthesis of full-length complementary (negative) RNA strand using the positive genomic RNA strand as a template. Initiation of this process

requires binding of the polymerase to a recognition site at the 3' end of the template (Martin *et al.*, 1990; Bustamante and Hull, 1998). (2) The second stage in genome replication is synthesis of progeny genomic RNA and subgenomic RNAs using the negative-strand RNA as a template (Bustamante and Hull, 1998).

Expression strategies as well as understanding of the genome organization of plants viruses have been greatly aided by the development of molecular techniques and the ability to obtain nucleotide sequences of complete viral genomes (Bustamante and Hull, 1998). Viruses with a small genome size are very dependent on the host eukaryotic protein-synthesising system and the viral genomes seem to have evolved to overcome the constraints of the plant host system (Bustamante and Hull, 1998). Small genomes are expected to encode a range of virus proteins and for luteoviruses almost all types of modulation mechanism are used during the expression of the different ORFs (Bustamante and Hull, 1998; Taliansky *et al.*, 2003). This includes the 5.8 kb genome of PLRV for which most of the diverse modes of expression of ssRNA are used (Martin *et al.*, 1990; Mayo and Ziegler-Graff, 1996; Bustamante and Hull, 1998).

The ways that PLRV use to express its genome are (1) translational frameshift between the overlapping ORFs, (2) initiation bypass, (3) readthrough of termination codons, (4) production of subgenomic (sg) RNA and (5) proteolysis of primary translation products to produce more than one protein from an ORF (Mayo and Ziegler-Graff, 1996; Taliansky *et al.*, 2003).

Translational, or ribosomal, frameshifting is a strategy frequently employed by various organisms to produce more than one protein from overlapping reading frames (Bustamante and Hull, 1998; Commandeur *et al.*, 2002). Frameshift may occur in either the +1 (3' direction) or in the -1 direction (5' direction) but the most common frameshift is in the -1 direction (Mayo and Ziegler-Graff, 1996; Bustamante and Hull, 1998; Commandeur *et al.*, 2002). This translational mechanism has been studied in detail and has been demonstrated for luteoviruses as well (Bustamante and Hull, 1998; Commandeur *et al.*, 2002). Very little is known about +1 translational frameshifting in plant viruses (Commandeur *et al.*, 2002).

Translational frameshift occurs between the overlapping ORFs 1 and 2 in PLRV that is required for the expression of the complete RdRp (Martin *et al.*, 1990; Mayo and Ziegler-Graff, 1996; Bustamante and Hull, 1998). ORF2 is only expressed by a -1 frameshift from ORF1 to produce a fusion protein P1 + P2 (Martin *et al.*, 1990; Mayo and Ziegler-Graff, 1996; Bustamante and Hull, 1998). A -1 translational frameshift event allows the ribosomes to bypass the stop codon at the 3' end of ORF1, resulting in a fusion protein (Martin *et al.*, 1990).

The position in which translational frameshifting occurs in PLRV is in the nucleotide overlap of PLRV ORF1-ORF2, towards the 5' end of the overlap (Martin *et al.*, 1990; Rhode *et al.*, 1994). Two characteristic genomic areas are responsible for the frameshift event: the frameshift site, also known as the slippery sequence or heptanucleotide signal, and a stem-loop or pseudoknot structure (Commandeur *et al.*, 2002). For most plant viruses the structural requirement of a

stem-loop or pseudoknot structure has been identified (Commandeur *et al.*, 2002). In PLRV the signal responsible for efficient frameshift is composed of the slippery sequence (UUUAAAU) followed by a sequence that has the potential to adopt either a pseudoknot structure or a simple stem-loop structure (Rhode *et al.*, 1994; Mayo and Ziegler-Graff, 1996; Bustamante and Hull, 1998). Studies on a Polish isolate of PLRV confirmed that the -1 frameshift in the overlap region depends on the slippery site and on a downstream positioned pseudoknot (Bustamante and Hull, 1998; Commandeur *et al*, 2002). On the other hand, studies on the frameshift region in a German isolate of PLRV indicated the presence of a stable stem-loop structure and a pseudoknot structure (Commandeur *et al.*, 2002). If either of these structures is disrupted by mutation, protein expression will still take place as both structures enable expression (Commandeur *et al.*, 2002).

In addition, the stem-loop or pseudoknot structure downstream of the heptanucleotide signal is a weak RNA structure, which explains the low frameshift efficiency of approximately 1% for PLRV (Commandeur *et al.*, 2002). This is low in comparison to frameshift efficiencies of animal viruses that range from 25 to 30% (Commandeur *et al.*, 2002). Mutational analysis studies in which the stem loop or pseudoknot RNA structure was replaced by a more stable stem loop showed a strong increase in frameshift efficiency of up to 15% (Commandeur *et al.*, 2002). The stem loop in the ORF2 sequence following the heptanucleotide signal, codes for a cluster of basic amino acids (Rhode *et al.*, 1994). These basic amino acids is a domain required for nucleic acid binding and may represent the site on the viral replicase for the binding of the PLRV RNA template during replication (Rhode *et al.*, 1994).

Initiation bypass is the result of leaky scanning by ribosomes to translate ORFs downstream of the first AUG in the mRNA (Rhode *et al.*, 1994; Mayo and Ziegler-Graff, 1996). This result in initiation of protein synthesis at internally located translational start codons and has been postulated to operate in the expression of various viral genes (Rhode *et al.*, 1994). *In vitro* translation of PLRV RNA results in the synthesis of two polypeptides that correspond in size to ORF0 and ORF1, the first two ORFs (Martin *et al.*, 1990; Mayo and Ziegler-Graff, 1996). This indicates that initiation of ORF1 occurs at an internal AUG, 133 nt downstream from the initiation site of ORF0, as a result of ribosomes scanning through the 70 nt 5' end of the RNA past the initiation site of ORF0 (Martin *et al.*, 1990). The 28 kDa product (P0) is found predominantly while the 70 kDa polypeptide (P1) is in the minor (Martin *et al.*, 1990).

Most viral RNAs are expressed in a monocistronic fashion with ribosomes scanning the RNA from the 5' end until the first AUG is expressed which make the use of two initiation sites in the same RNA unusual (Martin *et al.*, 1990). However, in the case where ribosomes use two initiation sites, such as for PLRV, the 5' AUG has a suboptimal sequence context to initiate translation at the AUG further downstream (Martin *et al.*, 1990). Only a portion of the ribosomes

initiate translation at the first AUG, while others continue to initiate translation at the AUG further downstream (Martin *et al.*, 1990).

Readthrough of leaky termination is a well-known translation mechanism in plant RNA viruses (Bahner *et al.*, 1990; Commandeur *et al.*, 2002). Readthrough occurs when a stop codon is suppressed by binding a suppressor tRNA, thereby permitting some of the ribosomes to read through into a downstream cistron as a result (Mayo and Ziegler-Graff, 1996; Bustamante and Hull, 1998). The process requires at least two elements: first a suppressor tRNA and secondly, the nucleotide context surrounding the termination codon and in particular the two downstream codons appear important for readthrough (Bustamante and Hull, 1998). The result of readthrough is a lesser proportion of a second functional fusion protein that includes the translation product of the next in-frame ORF (Mayo and Ziegler-Graff, 1996; Bustamante and Hull, 1998).

Readthrough of termination in PLRV occurs when suppression of the amber stop codon separating the PLRV CP gene (ORF3) and the in-frame ORF5 take place (Commandeur *et al.*, 2002). The nucleotide sequence around the termination codon of ORF3 resemble those around several other virus RNA termination codons that are misread by naturally occurring suppressor tRNAs to generate readthrough proteins (Bahner *et al.*, 1990). The sequence context surrounding the leaky stop codon is identical for all luteoviruses: AAA**UAG**GUAGAC (termination codon in bold type) (Martin *et al.*, 1990; Mayo and Ziegler-Graff, 1996). ORF5 seems to be expressed only as an ORF3 + ORF5 fusion protein by translational readthrough of ORF3 UAG termination codon as no polypeptide corresponding the primary translation product of ORF6 has been detected (Van der Wilk *et al.*, 1989; Bahner *et al.*, 1990; Mayo and Ziegler-Graff, 1996).

PLRV UAG suppression occurs at an efficiency of 1% *in vivo*, which is low compared to TMV (Commandeur *et al.*, 2002). This suggests that more information is necessary for the efficient readthrough event (Commandeur *et al.*, 2002). In addition, suppression is mediated by a rare, naturally occurring tRNA and it is present only in minor amounts in uninfected plants (Bahner *et al.*, 1990; Rhode *et al.*, 1994).

Transcription of sgRNA from genomic RNA (gRNA) is one of the ways RNA viruses use to express internally located genes (Commandeur *et al.*, 2002). This process is similar to mRNA transcription from a DNA template and also requires a promoter sequence located on the genomic minus strand (Commandeur *et al.*, 2002). For the formation of sgRNA, a (-) RNA stand is firstly synthesised by the RdRp from the gRNA (Bustamante and Hull, 1998). During the formation of the (-) RNA strand one of two things can happen: either premature termination could lead to the formation of (-) RNA stands of subgenomic length or alternatively the (-) RNA strands can be of genomic length (Bustamante and Hull, 1998). Consequently either subgenomic length (-) RNA strands or genomic length (-) RNA strands can serve as templates

to generate the subgenomic (+) RNA (Bustamante and Hull, 1998). The latter mechanism seems to be favoured, which means that the subgenomic (+) RNA could be synthesised via internal initiation in (-) RNA strands of genomic length (Bustamante and Hull, 1998). Subgenomic RNA is not able to replicate autonomously as the sequence contained in the 3' end required for the production of complementary subgenomic RNA chains is insufficient for replication (Bustamante and Hull, 1998).

For PLRV, there are two sgRNAs: sgRNA1 with a size of ~2.3 kb has been characterized as the mRNA for the 3' clustered ORFs (ORF3, ORF3/5 and ORF4) and the discovery of a second 0.8 kb sgRNA2, which increased the complexity of luteoviral genomes significantly (Ashoub *et al.*, 1998). The first nine nucleotides in the 5'-terminal non-translated leader sequence of sgRNA1 are identical to the respective 5'-terminal sequence of the gRNA, but those of sgRNA2 are different (Mayo and Ziegler-Graff, 1996; Taliansky *et al.*, 2003). Leader sequences have effects on translation of the RNAs *in vitro*, but are not translational enhancers and the non-translated sequence has no effect on translation (Taliansky *et al.*, 2003). The leader sequences of sgRNA1 are 212 nucleotides long and the 5' end are located 12 nucleotides upstream of the termination codon of ORF2 (Mayo and Ziegler-Graff, 1996). The 5' end of the sgRNA for a German PLRV isolate has been mapped to 40 nucleotides upstream of the ORF3 AUG codon (Mayo and Ziegler-Graff, 1996).

PLRV sgRNA1 allows only the expression of ORFs 3-5 while the three 5' ORFs 0-2 are expressed from gRNA (Mayo and Ziegler-Graff, 1996; Taliansky *et al.*, 2003). ORF3 is the 5' most gene of sgRNA1 with the sequence context GUU/GAAUGA surrounding the AUG initiation codon, similar in the luteoviruses PLRV, BYDV-PAV and BWYV (Martin *et al.*, 1990). ORF4, present inside sgRNA1, is proposed to be expressed from the sgRNA by a proportion of ribosomes scanning past the ORF3 AUG initiation codon to initiate translation at the downstream ORF4 AUG initiation codon (Martin *et al.*, 1990; Taliansky *et al.*, 2003). ORF4 initiation therefore happens in a fashion similar to the initiation bypass of ORF0 and ORF1. The ratio between P3 and P4 produced by translation of sgRNA1 has been reported to be either *c.* 1:1 or *c.* 1:7 (Taliansky *et al.*, 2003).

PLRV sgRNA2 is mapped to the 3'-end of the genome in positions 5190-5987 and may code for two additional viral proteins of 7.1 kDa (ORF6) and 14 kDa (ORF7) respectively (Ashoub *et al.*, 1998). *In vivo* experiments demonstrated that sgRNA2 functions as a bicistronic mRNA with high expression of ORF6 and low translational efficiency for synthesis of ORF7 (Ashoub *et al.*, 1998). Both of these ORFs are present in many PLRV isolates sequenced so far (PLRV-Scotland, PLRV-Netherlands, PLRV-Australia, PLRV-Canada and PLRV-Germany) with the single exception that ORF6 protein is truncated in PLRV-Australia due to conversion of a UGG codon for tryptophan to a UAG amber stop codon at position 5263-5265 (Ashoub *et al.*, 1998). PLRV ORF7 corresponds to the C-terminus of the readthrough protein P5 and displays nucleic

acid binding activity that makes it a good candidate for participating in luteoviral transcription regulation, possibly at late stages during the viral life cycle (Ashoub *et al.*, 1998).

Proteolysis of primary translation products might take place in PLRV as a protease consensus sequence has been found in P1 encoded proteins (Mayo and Ziegler-Graff, 1996). However, there is no direct evidence for proteolysis of luteovirus proteins (Mayo and Ziegler-Graff, 1996). Even so, proteolytic processing is implied for PLRV because the VPg (suggested as part of P1) does not correspond to the translational product of an ORF (Mayo and Ziegler-Graff, 1996). By comparison to other viruses, it is predicted that the processing is by a virus-encoded protease to give the VPg (Mayo and Ziegler-Graff, 1996). More recently, a truncated protein representing the C-terminus of ORF1 (P1-C25) could be detected in PLRV-infected plants and it might be a product of protease activity (Commandeur *et al.*, 2002).

Taken together, a large number of canonical and noncanonical strategies are involved in the expression of PLRV genes (Commandeur *et al.*, 2002). Luteoviruses present among the best examples of the versatility of expression strategies and the economy in use of coding sequences in small RNA genomes (Taliansky *et al.*, 2003). It is believed that luteoviruses are still a "translational gold mine" where new insights in plant viral gene expression will soon be unearthed (Commandeur *et al.*, 2002).

2.1.3.9 Characterization of virus coat protein

In studying the variation of PLRV, the CP product of ORF3 is mostly used to identify various isolates of PLRV. Other ORFs can also be used, but for the purpose of this study the variation in PLRV CP is of main importance. ORF3 makes part of the structural genes which determine the particle morphology, serological cross reactivity and possibly virus-vector interactions (Martin *et al.*, 1990). ORF3 therefore is part of the genes that are primarily responsible for the traditional taxonomy of classifying viruses into different groups (Martin *et al.*, 1990). For this reason, more detail is given below on the structure and function of the PLRV CP on a biochemical level.

Little is also known about the specific properties and biologically active domains of the luteovirus particle (Lee *et al.*, 2005). The virion structure of PLRV has not been resolved and there are no crystallographic data available for luteovirus CPs, but the general shape of the particles (consisting of 180 subunits assembled according to T = 3 quasi-symmetry) is known according to X-ray diffraction and molecular mass analysis (Terradot *et al.*, 2001; Lee *et al.*, 2005). A variable but minor number of the of the 180 subunits are readthrough subunits incorporated into the virion via their CP moiety, protruding from the surface of the virion even though virions can be assembled from the CP alone (Lee *et al.*, 2005).

The structures of other icosahedral plant viruses have been resolved by X-ray crystallography to reveal the general architecture of their CPs (Terradot *et al.*, 2001). For those icosahedral plant viruses the CP is based on two domains: the N-terminal arginine rich domain (R-domain) found

in the inner part of the capsid and the shell domain (S-domain) which forms the core of the capsid (Terradot *et al.*, 2001; Lee *et al.*, 2005). As no crystallographic data of PLRV CP is available, attempts have been made to predict the secondary structure from analysis of amino acid sequences (Taliansky *et al.*, 2003). Based on the known structures and structural sequences of these distantly related icosahedral RNA viruses the CP of PLRV can also be divided into an R- and S-domain (Lee *et al.*, 2005).

Luteovirus CPs also have highly basic N-terminal amino acids like the CPs of plant viruses with isometric particles in 58 to 69 amino acids of the N-terminal region (Mayo and Ziegler-Graff, 1996). In this region there is between 19 and 21 arginine or lysine residues separated by relatively nonpolar residues such as glycine and asparagine with no acidic residues (Mayo and Ziegler-Graff, 1996). This very arginine-rich sequence near the N-terminus contributes to the positive charge of P3 (Mayo *et al.*, 1989). According to Mayo *et al.* (1989) the charge of PLRV P3 is +24 when calculated as the difference between the total number of lysine and arginine residues and the total number of aspartic and glutamic acid residues. P3 has a marked positive charge considering that the charge of P1 equals +3, P2 +5, P4 0 and P5 -15 when calculated in the same manner as for P3 (Mayo *et al.*, 1989).

The basic R-domain at the amino terminus is believed to be involved in protein-RNA interaction and in capsid assembly (Kawchuk *et al.*, 1989; Terradot *et al.*, 2001). This basic region, being at the inside of intact particles, forms a close association with the virus RNA; presumably the basic residues neutralize the negative charge on the virus RNA (Torrance, 1992; Mayo and Ziegler-Graff, 1996). Furthermore, the highly basic sequences near the N-termini resemble nuclear localization signals and it is predicted that a stretch of 27 amino acids rich in arginine close to the N-terminus could be the nuclear localization signal (Mayo and Ziegler-Graff, 1996; Mukherjee *et al.*, 2003). Haupt *et al.* (2005) have shown that the PLRV CP region [17]Pro-Arg-Arg-Arg-Arg-Arg-Gln-Ser-Leu-Arg-Arg-Arg-Ala-Asn-Arg[31] operates as a nucleolar localization signal. The nuclear localization signal is probably responsible for transport of the virus particle through the nuclear pore but unlike other luteoviruses that accumulate in the nuclei of infected cells, nuclear localization of intact PLRV particles has not been demonstrated conclusively (Mukherjee *et al.*, 2003).

Furthermore, the nuclear localization signal is not exposed on the surface of the virus particle to facilitate its transport to the nucleus, but is probably located on the inside of the intact particle which means that the role for this signal in nuclear localization of intact virus particles remains to be confirmed (Mukherjee *et al.*, 2003). Thus, the R-domain of PLRV CP is involved in different virus functions as the RNA-binding domain and the nuclear localization signal, and these regions might overlap and compete with each other to modulate the involvement of the CP in different virus functions (Haupt *et al.*, 2005).

The S-domain which forms the virus shell is a wedge-shaped, eight-stranded antiparallel β-barrel often referred to as a jelly roll configuration (Fig. 2.3) (Torrance, 1992; Lee *et al.*, 2005). Two β-sheets are formed by the eight strands with four strands in each sheet (Terradot *et al.*, 2001). These eight β-strands are labeled B to I and it is to the N-terminal side of β-sheet B that the highly basic R-domain is found (Torrance, 1992; Mayo and Ziegler-Graff, 1996). From the formation of two β-sheets, the nomenclature of the strands of the jelly roll is designed as BIDG and CHEF (Terradot *et al.*, 2001). In addition, two α-helices are observed; one connecting strands C and D and another one located between strands E and F (Terradot *et al.*, 2001). In a study by Terradot *et al.* (2001) on a low-transmissible isolate of PLRV (PLRV-14.2), the S-domain have been assumed to span from Thr_{60} to the remaining 149 C-terminal residues. The 59 N-terminal residues which form the R-domain contained more than 33% arginines (Terradot *et al.*, 2001). It is the R-domain-arm-S-domain modules that make up each one of the 180 subunits of the PLRV particle (Torrance, 1992).
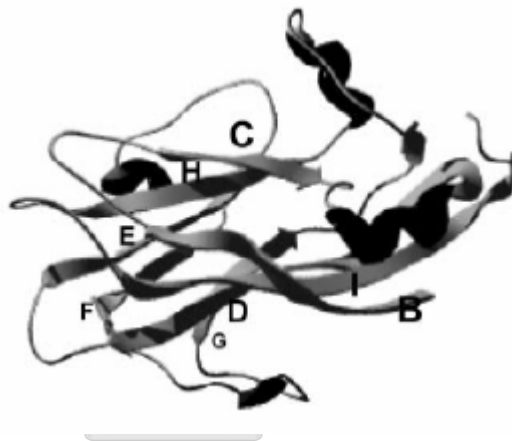
Fig. 2.3: Illustration of the jelly roll scaffold as found for PLRV CP. The jelly roll scaffold is that of subunit A of RYMV CP which illustrates the jelly roll structure that is also found in PLRV CP. Nomenclature BIDG and CHEF is for the β-strands (Terradot *et al.*, 2001).

The S-domain three-dimensional structure is highly conserved among plant viruses as three-dimensional biological structures are generally better conserved than primary protein sequences (Terradot *et al.*, 2001; Lee *et al.*, 2005). From secondary structure predictions, it has been confirmed that the majority of amino acid changes fall under the structurally variable regions, but that the α-helix and β-sheet are more structurally conserved (Mukherjee *et al.*, 2003). Mukherjee *et al.* (2003) found that from a secondary structure prediction of an Indian PLRV isolate, out of 207 amino acids, 6.7% has propensity towards an α-helix, 33.9% will form a β-sheet while the rest will adopt loop conformations.

For the case of poleroviruses, the CP amino acid sequences are strongly conserved but they differ from those of other viruses in the "Luteovirus supergroup" which consist in genus *Sobemovirus,* family *Tombusviridae* and family *Luteoviridae* (Terradot *et al.*, 2001). However, significant alignments have been obtained which suggests that secondary structural elements

are remarkably similar between virus CPs which has been analyzed despite the observed variation in amino acid sequences (Terradot *et al.*, 2001). Members of different virus families share a similar 3D scaffold of eight β-strands (Terradot *et al.*, 2001).

The prediction of the structure of PLRV CP as a basic N-terminal region followed by patches of β-sheet was made shortly after the publication of sequences and this model have been supported from subsequent predictions based on more refined software packages (Taliansky *et al.*, 2003). Recently a model was constructed for PLRV CP from the known crystal structure of *Rice yellow mottle virus* (RYMV) by Terradot *et al.* (2001) (Taliansky *et al.*, 2003). This is by far the most sophisticated of PLRV CP models and have been made following the observation of a weak (17% identity, 33% similarity), but significant resemblance between the CP sequences (Taliansky *et al.*, 2003). Even though such a model cannot be compared to a 3D crystal structure it can give useful insight into the structural shape of the CP up to residue level (Terradot *et al.*, 2001). These models are less accurate at the atomic level but it helps to understand virus properties (Terradot *et al.*, 2001).

The sequence identity of PLRV CP obtained with that of the RYMV template was lower than what is usually considered suitable for reliable model building but comparison of the model with immunological and site-directed mutagenesis data previously reported for PLRV or related viruses has shown that the model was accurate (Terradot *et al.*, 2001). For the identification of structure-function relationships between PLRV CP and RYMV, pairwise comparisons of sequences such as BLAST and FAST were not used as the chance of detecting relationships become increasingly small when the sequence identity is below 30% (Terradot *et al.*, 2001). In the search of shared characteristics in sets of related sequences, Terradot *et al.* (2001) used the hidden Markov model as this method can predict significant alignments of protein sequences with identity as low as 15% and below (Terradot *et al.*, 2001). With the hidden Markov model, the CP of RYMV was identified as a structure model (Terradot *et al.*, 2001).

From the PLRV CP model by Terradot *et al.* (2001) it has been found that the surface of the trimer displays an interesting patch of acidic residues, found at the center of the trimer. It has been suggested that Trp_{171} is located at the center of the trimer, involved in viral encapsidation and a stabilizing effect between the subunits (Terradot *et al.*, 2001). On the other hand, the acidic patch is made of residues Glu_{109}, Glu_{170}, Glu_{176}, Asp_{173} and Asp_{177} on the model (Terradot *et al.*, 2001). Fig. 2.4 illustrates the surface properties of PLRV-14.2 (a low transmissible isolate of PLRV for which the model was constructed) colored according to the properties of amino acid residues (Terradot *et al.*, 2001).

Lee *et al.* (2005) confirmed the PLRV CP model by Terradot *et al.* (2001) in their study of the acidic domains, with only a few changes in the region that includes strands E and F. These changes would not have a large effect on the predicted structure of the acidic patch within a subunit or the interactions of the acidic patch at the center of the CP trimer (Lee *et al.*, 2005).

Previous reports place the acidic patch domain as amino acids in loops C and D and G and H. According to Lee *et al.* (2005) two acidic domains separated by 55 amino acids are formed as βD and βG are split into two domains which results in the acidic patch residues to be partially contained in loops and partially in β sheets (Lee *et al.*, 2005). These domains are predicted to be adjacent surface features on the virion (Lee *et al.*, 2005).

Fig. 2.4: Surface study of PLRV capsid. Surface of the trimeric CP structure is colored according to residue properties: neutral and hydrophobic (Ala, Val, Leu, Ile, Pro, Met, Phe, and Trp) in orange, neutral and polar (Gly, Ser, Thr, Cys, Asn, Gln, and Tyr) in green, acidic (Asp and Glu) in magenta and basic (Arg, His, and Lys) in blue. A white circle indicates the patch of acidic residues at the center of the trimeric assembly (Terradot *et al.*, 2001).

The acidic patch is possibly involved in the retention of parts of P5 on the surface of virus particles or it could be involved in the switch of the N-terminal sequence of the CP (a major epitope) from the inside of the particle to the outside, this change is postulated to be linked with particle disassembly (Taliansky *et al.*, 2003). This region could also be involved in important recognition events at different stages of the transmission process (Taliansky *et al.*, 2003).

In a study by Lee *et al.* (2005) on the acidic domains of PLRV and the surrounding regions, it was found that alanine substitutions of certain amino acids disrupted the ability of the coat protein to assemble stable particles and the ability of the viral RNA to move systemically in four plant species. They were able to identify amino acids in these regions that are critical to virion assembly and stability, host-dependent systemic infection and aphid transmission (Lee *et al.*, 2005). The amino acids critical for virion assembly are located within a depression at the center of the coat protein trimer (residues Glu_{170} and Asp_{177}) while those amino acids that affect plant infection and/or aphid transmission are predicted to be located around the perimeter of the depression (residues Glu_{109}, Asp_{173} and Glu_{176}) (Lee *et al.*, 2005). It became apparent from their study that it is the conserved downstream domain of the acidic patch that is more likely to be the biological active region involved in CP subunit interactions, plant-virus interactions and aphid-virus recognition (Lee *et al.*, 2005).

The aligned sequences of viral CPs have not only been used for secondary structure predictions but also for the prediction of regions of the coat protein that are on the outside of the particles (Reavy and Mayo, 2002). Torrance (1992) used a secondary structure model based on

the multiple sequence alignment of CPs with known secondary structure to interpret the location of epitopes in PLRV particles (Mayo and Ziegler-Graff, 1996). When reactions of antibodies to each of a set of overlapping peptides representing the entire protein sequence was made, Torrance (1992) found that most of the epitopes did not correlate well with antigenic areas predicted by computer algorithms (Torrance, 1992; Mayo and Ziegler-Graff, 1996).

Twelve continuous epitopes were identified; epitopes 1-3 were on the R- or arm-domains while epitopes 4-12 were mapped to different regions of the S-domain (Torrance, 1992). Epitopes 3, 5, 10 and 12 corresponded to areas of antigenicity predicted by computer algorithm, whereas epitopes 1, 2, 6, 7, 8 and 11 have net negative values (Torrance, 1992). Therefore, only four of twelve epitopes correlated well with predictions of antigenicity (Torrance, 1992). It was notable that epitope 3 (amino acid residues 50-60) contained some of the Arg residues in the Arg-repeat region 45-51, and interspersed Arg residues between amino acids 55-66.

Generally, the N-terminal region is thought to be internal while the C-terminal region is likely to be on the exterior of the particle (Reavy and Mayo, 2002). However, a major epitope (epitope 1, amino acid residues 1-7) was located at the N-terminus which indicates that even though the N-terminus is hydrophobic, it is exposed at the surface of the particles (Torrance, 1992). It is suggested that the N-terminal amino acids of PLRV CP are exposed when particles swell because of changes in pH or ionic conditions (Mayo and Ziegler-Graff, 1996). Torrance (1992) concluded from ELISA experiments that epitope 1 is not very abundant nor readily accessible, or both.

There are also epitopes in PLRV which are highly conserved but are found in the particle interior, believed to be involved in RNA protein interactions that are necessary for particle assembly or stability (Martin *et al.*, 1990). These epitopes will only react with antisera when virus particles are denatured for instance in carbonate buffer (Martin *et al.*, 1990). Disrupted virus particles enable serological relationships to be detected between species within a genus (Van Regenmortel *et al.*, 2000). Extensive cross-reactivity is observed when denatured viruses are used as antigen which suggests that the most variable region of the polerovirus capsid is exposed in intact virions (Kawchuk *et al.*, 1989; Van Regenmortel *et al.*, 2000).

Three surface-exposed domains have been identified on the virion structure of PLRV (Reavy and Mayo, 2002; Lee *et al.*, 2005). Epitopes 4 (amino acid residues 72-75), 5 (amino acid residues 83-89) and 10 (amino acid residues 172-178) may be located on the surface of the virus particle (Torrance, 1992). Epitope 4 and 5 could be part of or close to the loop connecting the β-strands B and C (Torrance, 1992; Terradot *et al.*, 2001). On the other hand, epitope 10 can be assigned to the loop connecting β-strands G and H (Torrance, 1992; Terradot *et al.*, 2001). Continuous epitopes (epitopes composed of short linear fragments of amino acid residues contiguous in the primary amino acid sequence) can be found on surface-located N-terminal or C-terminal portions of the sequence, or on the loops and turns connecting the β-

strands of the core regions (β-barrels) of the protein shell as is the case for epitopes 4, 5 and 10 (Torrance, 1992). Epitopes 5 and 10 are particularly antigenic as they are located on loops known to be important antigenic sites in structurally related viruses (Terradot *et al.*, 2001; Lee *et al.*, 2005).

Furthermore, epitopes are regions of sequence that are intrinsically mobile and are therefore candidates for surface regions that contribute to biological properties (Taliansky *et al.*, 2003). The loop connecting β-strands G and H which contains epitope 10 is situated at the outer surface which makes it easily accessible (Terradot *et al.*, 2001). Partially contained in epitope 10 and extending a few residues towards the N-termius is one domain of the acidic patch region (Lee *et al.*, 2005). It has been suggested by Terradot *et al.* (2001) that this region is important for molecular recognition and that it could confer specific properties to each genus and/or species within *Luteoviridae*.

Other studies with monoclonal antibodies (MAbs) specific for certain PLRV isolates have also been conducted to locate various epitopes on the particle surface. One by Van den Heuvel *et al.* (1990) found that there are a minimum number of nine different but overlapping epitopes on a Wageningen isolate of PLRV. These epitopes were located with only ten MAbs which were attributed to the fact that various states of the CP were analyzed (Van den Heuvel *et al.*, 1990). The panel of MAbs could be split into three groups: (1) a group directed against discontinuous epitopes present on a subunit of the capsid which is not degraded under alkaline conditions, (2) another group which did not detect degraded virus particles and are directed to discontinuous epitopes formed by the quaternary protein structure or by a CP-subunit configuration sensitive to alkaline degradation, (3) a group directed to a continuous epitope exposed on the surface of the virus CP subunit which is not accessible when the virus is intact (Van den Heuvel *et al.*, 1990).

Another study with MAbs by Massalski and Harrison (1987) for a British isolate of PLRV found five epitopes on the surface of PLRV particles. Of the ten MAbs that were used by Massalski and Harrison (1987), four did not react with disrupted virus particles, suggesting that those antibodies are specific for epitopes dependent on quaternary structure. Two of the five epitopes located by Massalski and Harrison (1987) were missing in poorly aphid-transmissible isolates of PLRV, which were also apparently dependent on quaternary structure. Of the 30 British isolates of PLRV tested, the predominant impression was one of great antigenic uniformity (Massalski and Harrison, 1987). These results extend those of previous work which showed that PLRV isolates from potatoes in different countries are strongly related (Massalski and Harrison, 1987).

2.1.4 Assessment of PLRV variation

2.1.4.1 Constraints on sequence variation

In addition to MAbs, amino acid sequences of PLRV CPs also show that poleroviruses are strongly conserved (Terradot *et al.*, 2001). Furthermore, sequence data also suggest that PLRV is poorly variable at the genetic level (Guyader *et al.*, 2004). PLRV isolates collected around the world between 1983 and 2000 differ little in sequence from each other with the identity over all open reading frames being 94-98% (Guyader *et al.*, 2004). The CP of luteoviruses is the most highly conserved viral gene unlike that of most other virus groups (Martin *et al.*, 1990). In a study by Mukherjee *et al.* (2003) the nucleotide- and deduced amino acid sequences of the CP of various PLRV isolates showed 97-99% similarity with an Indian isolate of PLRV at both the nucleotide and amino acid sequence level. Faccioli *et al.* (1995) found the sequence of a cloned PLRV CP gene to have 96-98% homology to five other PLRV isolates, which supports the idea that the CP gene is a relatively conserved sequence within the PLRV genome. Likewise, comparisons among isolates obtained in Peru showed that there was little variation in ORF3 with no greater variation when these were compared with published sequences (Haliloglu and Bostan, 2002; Taliansky *et al.*, 2003).

As ORF3 contains a second embedded ORF (ORF4), this may have caused an evolutionary constraint to slow down luteovirus CP-sequence divergence relative to other viruses (Martin *et al.*, 1990). Overlapping reading frames not only maximize the genetic information in smaller genomes, but they constrain the variability of these concurrent genes (Guyader and Ducray, 2002). However, for PLRV, regions where two ORFs overlap do not tend to be less variable than non-overlapping ones (Guyader and Ducray, 2002).

The lack of sequence variation in PLRV and its genetic stability could be an indication of evolutionary stasis (Guyader and Ducray, 2002). This evolutionary stasis appears to be unique for PLRV as other members of the family *Luteoviridae* do not share this nature (Guyader and Ducray, 2002). For instance, other species like those belonging to the beet polerovirus complex and BYDV-PAV are known to be more variable (Guyader and Ducray, 2002). However, since only five complete PLRV sequences are available in the international databases, there are not sufficient information to draw firm conclusions (Guyader and Ducray, 2002). Either PLRV only recently diverged from an ancestral virus by for example acquiring an ability to infect potato, or it has been subject to very strong selection constraints (Taliansky *et al.*, 2003; Guyader *et al.*, 2004).

One of the selection constraints on PLRV variation could be that imposed by its genetic structure as 31% of the genome consists of overlapping coding regions with differential selection applying to ORF products (Guyader and Ducray, 2002). This could result in a poor intrinsic variability of the PLRV genome (Guyader *et al.*, 2004). In addition, the evolution of

luteoviruses involves a trichotomy – the virus, the plant and the aphid vector (Martin *et al.*, 1990). It is for this reason that other constraints on PLRV variation could be both the host plant and the aphid vector, acting on the whole genome (Guyader and Ducray, 2002). The genetic homogeneity of PLRV may be related to the genetic homogeneity of both the host plant and the aphid vector seeing that the range of PLRV plant hosts and aphid vectors is restricted (Guyader and Ducray, 2002). Accordingly, Guyader *et al.* (2004) hypothesized that transmission by aphids and/or multiplication in the hosts plants are key factors structuring virus populations. In general, diversity in viruses is constrained by requirements of transmission and host defenses (Power, 2000).

For the transmission of luteoviruses, a high degree of specificity is required, with each luteovirus being transmitted efficiently by only one or a few aphid species (Martin *et al.*, 1990). This is likely a major factor in the evolution of these viruses as recent studies of insect-transmitted plant viruses demonstrate highly conserved molecular motifs in viral genomes that regulate the specificity of insect transmission (Martin *et al.,* 1990; Power, 2000). For PLRV, *M. persicae* is the main vector considered to be the most efficient aphid in PLRV transmission (Guyader and Ducray, 2002). As a result, insect transmission probably constitutes a severe genetic bottleneck which may lower both the number of transmitted viruses and their genetic variability through selection by aphids (Guyader *et al.*, 2004).

Insect transmission has a strong selection pressure also on the antigenic conservation of PLRV particle proteins, implying a strong selection pressure on P3 and P5 (Massalski and Harrison, 1987). This is because viral particle proteins have a crucial function in transmission of luteoviruses by their vectors, which is supported by the considerable amount of antigenic uniformity found among PLRV isolates (Massalski and Harrison, 1987). For example, Massalski and Harrison (1987) found that the only examples in their study on the antigenic variation of PLRV were isolates characterized by poor aphid transmissibility or aphid non-transmissibility (Massalski and Harrison, 1987). In addition, Guyader *et al.* (2004) found evidence that vector transmission limits PLRV population diversity in the CP coding region through a mechanism of stabilizing (purifying) selection. Generally, there is a pattern of highly conserved motifs that are responsible for insect transmissibility that suggest that the requirements of vector transmission exert significant selection pressure to limit the diversity of CP sequences (Power, 2000).

A high degree of specificity is required between the virus and vector even for viruses that are simply carried on the mouthparts of vectors and appear to have the least specific relations with their vectors (Power, 2000). Thus, variation is constrained by the need to retain specific interactions with the aphid vector so much the more for PLRV, which circulates through the aphid vector (Power, 2000). As insect transmission mode is such a consistent evolutionary feature and constraint on the variability of different viruses, it can be used to assign viruses to different genera (Power, 2000). Furthermore, most viruses are able to take advantage of only a

relatively narrow set of vectors that adds to the degree of specificity required for transmission (Power, 2000). Moreover, plant hosts are immobile so most plant viruses are dependent on insect vectors for transmission that indicate the significant need to retain requirements for vector compatibility (Power, 2000).

On the other hand, the level of variation in plant viruses is dependent on virus-host interactions (Roossinck, 2003). For three plant viruses it has been reported that population diversity is conditioned by host-virus interactions and is positively correlated to their host range (Guyader *et al.*, 2004). Host plants could constrain plant virus evolution as viral RNAs and proteins must interact with host cell components to allow translation, genome replication and both cell-to-cell and long distance movement of virus particles (Guyader *et al.*, 2004).

Furthermore, there is a correlation between mutation frequency and virus host range that suggests diverse populations constitute an advantage for RNA plant viruses (Roossinck, 2003). In natural settings plants are rarely found in monoculture as seen in agricultural settings, which means that insect feeding is likely to transmit a virus to a variety of plant hosts (Roossinck, 2003). Consequently, for the survival of a plant virus in natural settings, host adaptability is essential (Roossinck, 2003). For PLRV most isolates originated from a single plant species, the potato, and together with the narrow genetic base of potatoes in cultivation may restrict PLRV variation (Guyader and Ducray, 2002; Taliansky *et al.*, 2003). Other poleroviruses, like the beet polerovirus complex and BYDV-PAV, need to infect alternative hosts to complete their epidemiological cycle (Guyader and Ducray, 2002). PLRV is subject to strong constraints within host plants which slow down the process of adaptation to new hosts as diversity is hardly found in the course of infection of different host plant species (Guyader *et al.*, 2004).

Virus distribution is constrained more by the specificity of virus-vector interaction than by the specificity of virus-host-plant interactions as expansion of the host range of insect vectors has been shown to increase the host range of the viruses that these vectors transmit (Power, 2000). Vector-borne plant viruses demonstrate a greater specificity with their vectors than with their hosts as revealed in a quantitative comparison of specificity in virus-host and virus-vector relationships among these viruses (Power, 2000).

The host range of the virus is largely determined by the host range of the vector, which suggests that viruses can adapt to new hosts quite readily (Power, 2000). The selection imposed by a requirement for efficient vectors may be more severe than that imposed by host plant defenses as suggested by the consistency of transmission mode within a virus genus and the generally greater specificity of vector interactions compared to host interactions (Power, 2000). It is not clear why it should be significantly more difficult for a virus to increase its range of efficient vectors than to increase its host range, this question should be addressed in future research (Power, 2000).

Thus, the separation of a virus in a different host with a different vector can allow for independent evolution as effectively as geographical separation (Marin *et al.*, 1990). If a virus is introduced into a new host that is not a preferred host of its vector, it must be able to multiply in that host and be vectored by an aphid that feeds on that host before it can establish and evolve independently (Martin *et al.*, 1990). If the new host is perennial, the virus can evolve over considerable time within the host before it needs to become transmissible by a second aphid (Martin *et al.*, 1990). On the other hand, in annual hosts the time for evolution within a single host plant is limited because luteoviruses are not seed-transmitted (Martin *et al.*, 1990). In the case of PLRV, the host plant cellular environment is a key parameter constraining virus population diversity as it was shown that by forcing the virus to explore a new cellular environment, the genome is allowed to vary more (Guyader *et al.*, 2004). PLRV is consequently an interesting model to study the nature and the influence of selective constraints on virus evolution (Guyader and Ducray, 2002).

2.1.4.2 Diversification of PLRV isolates

Despite the low variability of PLRV at the sequence level, there is diversity in the biological properties of the virus (Guyader and Ducray, 2002). Several isolates of PLRV can be distinguished by differences in either the severity of symptoms induced in potato or in their ease of transmission by the aphid, *M. persicae* (Massalski and Harrison, 1987). The transmissibility by aphid vectors range from low to high levels, according to both the isolate and aphid clone used to transmit the virus (Guyader and Ducray, 2002). In addition, some PLRV isolates differ in their ability to infect some plant species, for instance a certain PLRV isolate can infect tomato plants while typical potato isolates have no effects on the plants (Guyader and Ducray, 2002). On the other hand, this PLRV isolate causes only weak or no symptoms on potato plants (Guyader and Ducray, 2002).

An example of a PLRV isolate which is differentially transmitted by various clones of aphids is PLRV-14.2 (Terradot *et al.*, 2001). This isolate is not transmitted efficiently due to inefficient interaction of the capsid with the gut membrane of the insect (Terradot *et al.*, 2001). There are several changes at specific positions of the CP and readthrough protein amino acid sequences of PLRV-14.2 which could be involved in its poor transmissibility (Terradot *et al.*, 2001). PLRV isolates 15 and V are also known to be poorly aphid transmissible (Jolly and Mayo, 1994). These particles are different from transmissible isolates of PLRV in that their particles lack an epitope present in transmissible isolates of PLRV (Massalski and Harrison, 1987; Jolly and Mayo, 1994). In a comparison of the amino acid sequences of transmissible and poorly aphid transmissible isolates, most differences were in the C-terminal half of the readthrough protein (Jolly and Mayo, 1994). Two sites in the C-terminal half of the readthrough protein has been located at which changes could have resulted in the loss of an epitope and in a diminished efficiency of transmission (Jolly and Mayo, 1994). However, in a study on Dutch transmissible

and Dutch poorly aphid transmissible PLRV isolates, transmissibility was shown to correlate with the presence of a conformation-insensitive epitope in the CP (Jolly and Mayo, 1994). These results suggest that a part of the coat protein and a part of the readthrough protein are involved in transmission (Jolly and Mayo, 1994).

However, even though there is variation in the pathogenicity and transmissibility by *M. persicae*, PLRV isolates seem in general to be antigenically very similar (Tamada *et al.*, 1984; Massalski and Harrison, 1987). For example, a polyclonal antiserum to a Japanese isolate of PLRV reacted with PLRV isolates from several other countries (Massalski and Harrison, 1987). Furthermore, when British PLRV isolates differing in virulence or aphid transmissibility was compared using a polyclonal antiserum to the British stock isolate of PLRV, no differences could be detected (Massalski and Harrison, 1987). However, it could be that the antigenic differences between the isolates differing in aphid transmissibility are too subtle to be detected in tests with polyclonal antisera (Massalski and Harrison, 1987).

Although there are mainly differences in the biological properties of PLRV, with very few differences at the sequence level, PLRV can be resolved into different groups (Taliansky *et al.*, 2003). The 5' ORFs show marked diversity between luteoviruses and relationships with different viruses of other groups (Martin *et al.*, 1990). In an assessment of the distribution variability along the genome of PLRV, it was found that the most variable sites concentrated in ORF1 (including the overlapping regions with ORF0 and ORF2), the 3' third of ORF2, the central non-coding region, the 5' overlapping regions between ORF3 and ORF4 and in ORF5 (Guyader and Ducray, 2002). However, the 5'-proximal five to 20 residues of viral genomes are generally strongly conserved within a virus groups, presumably because these are important for initiation of translation and/or replication (Keese *et al.*, 1990).

The region of ORF0 and the 5' part of ORF1 resolve PLRV isolates into three groups (Taliansky *et al.*, 2003; Guyader *et al.*, 2004). Such a study in comparisons among 19 PLRV isolates in the region of ORF0 and the 5' part of ORF1 resolved them into one group of exclusively Australian isolates, one containing Peruvian and European isolates and a third group containing isolates from diverse countries and continents such as Australia, Cuba and Europe (Guyader and Ducray, 2002; Taliansky *et al.*, 2003). No geographical correlation was found with sequence variation (Guyader and Ducray, 2002; Taliansky *et al.*, 2003). Comparison of nucleotide diversity in each group showed that isolates in the exclusively Australian group were more diverse than those in either of the other two groups (Guyader and Ducray, 2002). In nearly all ORFs of PLRV the Australian isolate is the most divergent particularly in ORF1, but in ORF0 a Peruvian isolate was more divergent (Guyader and Ducray, 2002). It is the 600 nucleotide region (residues 935 to 1557) in the polymerase-coding region (ORF1) that shows marked diversity in the Australian PLRV isolate compared to the rest of the PLRV isolates (Keese *et al.*,

1990). This region has 22% sequence variation whereas the CP gene (ORF3) is the most conserved coding region (Keese *et al.*, 1990).

Martin *et al.* (1990) also found that sequence variation confined to ORF1 resolved an Australian PLRV isolate as the one with the most divergent genome when compared to three isolates from other countries. However, Martin *et al.* (1990) established that the ORF0 sequences are markedly conserved among the different PLRV isolates, but poorly conserved when compared to the equivalent ORF of BWYV, a closely related luteovirus (Martin *et al.*, 1990). In their study, the four isolates compared were from widespread geographical locations; the Netherlands, Scotland, Australia and Canada (Martin *et al.*, 1990). These sequences were found to be closely related with more than 93% sequence homology, excluding a possible extensive 5' terminal sequence difference in the Scottish isolate (Martin *et al.*, 1990). In addition, within the Australian isolate it was found that there was little sequence heterogeneity, since multiple independent clones for 95% of the genome revealed less than 10 nucleotide residue differences (Martin *et al.*, 1990).

The diversification of Australian PLRV isolates has been found in a number of studies. Sequence comparison of the CP gene of an Italian isolate with five other PLRV isolates in a study by Faccioli *et al.* (1995) revealed a close similarity to three European and a Canadian isolate and a more distant relationship with an Australian one. The PLRV CP nucleotide sequences of the Italian isolate showed homology of up to 98% with those of European isolates (from Scotland, Germany and the Netherlands) and the isolate from Canada (Faccioli *et al.*, 1995). The Australian isolate had a significantly higher degree of variability (3.7%). However, not all of the differences at the nucleotide level of CP genes resulted in an amino acid change due to the degenerate character of the genetic code (Faccioli *et al.*, 1995). For instance in the Australian isolate there were as many as 23 nucleotide shifts, but only seven amino acids were modified as a result (Faccioli *et al.*, 1995). These results corroborate findings by Martin *et al.* (1990) who revealed 98-99% homology in the ORF3 region of several European PLRV isolates and a somewhat lower degree of similarity (96-97%) between these isolates and an Australian one (Faccioli *et al.*, 1995; Mukherjee *et al.*, 2003).

Similarly, comparisons of complete sequences of PLRV isolates originating from Canada, Scotland, the Netherlands, Poland and Australia showed that four of these isolates shared about 98% of their residues whereas the Australian isolate was more divergent showing an overall identity of 93% with the other isolates as found by Guyader *et al.* (2004). Keese *et al.* (1990) also found that an Australian isolate of PLRV is more divergent from complete genome sequences of Canadian, Scottish and Dutch isolates of PLRV. Canadian, Dutch and Scottish PLRV sequences differed in pairwise comparisons by about 2% whereas each of these three isolates differed from the Australian isolate by 6.6 to 6.8% (Keese *et al.*, 1990). The results of Keese *et al.* (1990) therefore strengthens the argument of Guyader *et al.* (2004) that PLRV

isolates from Canada, Scotland and the Netherlands share about 98% of the residues and that the Australian PLRV isolate is the most divergent with an overall identity of about 93% with other isolates (Guyader and Ducray, 2002).

Other than the Australian isolate, an isolate from the United States is also more diverse than sequences from other countries. De Souza-Dias *et al*. (1999b) compared Brazilian PLRV sequences with isolates from the United States, Canada, Poland, Australia, Scotland and the Netherlands. In these isolates, the most variable regions of the PLRV genome were compared, which constitute 21.6% of the genome (De Souza-Dias *et al*., 1999b). It was found that at the nucleotide level the United States and Australian isolates display the least amount of relatedness when compared to the other PLRVs (De Souza-Dias *et al*., 1999b). Three Brazilian isolates were the most closely related with approximately 99% homology in the variable region sequences, these were approximately 97% homologous to PLRVs from Europe and Canada (De Souza-Dias *et al*., 1999b). On the other hand, the Brazilian isolates showed only 95% homology to the isolates from the United States and Australia (De Souza-Dias *et al*., 1999b).

The most common variations were transitions attributed to replication errors, i.e., C to U (44.6%) and A to G (30%) (De Souza-Dias *et al*., 1999b). Most of these changes (approximately 63%) did not result in changes at the amino acid level (De Souza-Dias *et al*., 1999b). From the nucleotide variations that did cause amino acid changes, 34 of the 51 amino acid variations were unique mutations (De Souza-Dias *et al*., 1999b). However, many of the amino acid variations probably have no or little effect on function because one nonpolar amino acid would for instance be exchanged for another (De Souza-Dias *et al*., 1999b). Therefore, the PLRVs isolated in Brazil were more closely related to PLRVs from Europe (the Netherlands, Poland and Scotland) and Canada than to isolates from either Australia or the United States (De Souza-Dias *et al*., 1999b).

Mukherjee *et al*. (2003) compared an Indian isolate of PLRV with PLRV isolates from UK, Canada and Australia to find a somewhat distant relationship with the Australian isolate. The highest sequence similarity was found with isolates from UK and Canada (99%) and the least with the Australian isolate (97%). Only two nucleotide shifts were observed among Indian isolates of PLRV in comparison with the Canadian and UK isolates but in comparison with an Australian isolate, a total of eighteen nucleotide shifts were noticed (Mukherjee *et al*., 2003). The Indian isolate is consequently closer to the European and Canadian isolates than to the Australian isolate probably due to the introduction of PLRV infected potatoes from Europe in India (Mukherjee *et al*., 2003). On the other hand, some of the nucleotide changes may be because of the natural variability (Mukherjee *et al*., 2003).

A phylogenetic study by Haliloglu and Bostan (2002) found that an Australian isolate was very distinct from other isolates in the phylogenetic relationship. Seventeen PLRV CP nucleotide sequences of published GenBank PLRV isolates were used to analyze the phylogenetic

relationships (Haliloglu and Bostan, 2002). The phylogenetic tree based on the nucleotide sequences showed two major clusters originating from a common branch in the phylogram (Haliloglu and Bostan, 2002). The first cluster includes isolates from the Netherlands, USA and Poland with isolates from Korea, Scotland and Canada in a sub-group in the second cluster (Haliloglu and Bostan, 2002). The Australian isolate was found to be very distinct and were sub-grouped in the second cluster (Haliloglu and Bostan, 2002).

Finally, in a comparative study by Jolly and Mayo (1994) of P3 and P5 amino acid sequences from Scottish isolates, it was found that these sequences were as divergent from each other (ca. 1%) as they are from overseas isolates, except for one from Australia. (Jolly and Mayo, 1994; Guyader and Ducray, 2002; Haliloglu and Bostan, 2002). The P3 and P5 sequences among isolates obtained in Scotland were 96 to 99% identical, whereas the number of amino acid differences with a reference Scottish isolate was 24 with Dutch PLRV, 24 with Canadian and 17 with American PLRV but 31 with Australian PLRV (Jolly and Mayo, 1994; Mayo and Ziegler-Graff, 1996). Amino acid changes in P3 or P5 of Australian isolates might influence its structure as Massalski and Harrison (1987) found from a study that included two MAbs, which might detect similar epitopes, that those epitopes were lacking in some Australian isolates of PLRV.

The reason for the differences in Australian isolates of PLRV compared with those from other countries and continents are not clear (Keese *et al.*, 1990). It may be the result of the geographical or evolutionary isolation of Australian isolates from those in the northern hemisphere (Keese *et al.*, 1990; Guyader and Ducray, 2002). Furthermore, the physical isolation and unique environmental selection pressures in Australia could contribute to its divergence (Faccioli *et al.*, 1995). Isolates of PLRV presumably all spread in potato material derived from the Andean region of South America (Taliansky *et al.*, 2003). Consequently, an examination of the genomic diversity of PLRV isolates in the Andean highlands may shed light on the evolution of this virus (Keese *et al.*, 1990).

2.1.4.3 Evolutionary events of PLRV

PLRV can evolve not only in response to unique environmental selection pressures as could be the case for isolates in Australia, but also due to the high mutation rate of viruses. A common property of plant RNA viruses is their high mutation rate due to their error-prone RNA polymerase that usually range between 10^{-3} and 10^{-5} misincorporation per nucleotide copied (Power, 2000; Guyader *et al.*, 2004). The high mutation rate of viruses results in their remarkable genetic diversity both within and between species (Power, 2000). Other than high mutation rates, RNA viruses can exchange genetic material through recombination and reassortment mechanisms (Power, 2000; Guyader *et al.*, 2004). This allows both the elimination of deleterious mutations and the creation or spread of beneficial combinations of changes (Guyader and Ducray, 2002).

Alignment results of complete sequences suggest that major events such as deletion e.g. in the ORF1 of isolate Noir, mutation at a stop codon e.g. a single substitution which abolished the consensus ORF0 stop codon of isolate 14.2 and intraspecific homologous recombination events e.g. an intraspecific recombination event in the ORF0 in isolate Au16, have occurred in the evolutionary history of PLRV (Guyader and Ducray, 2002; Guyader *et al.*, 2004). This suggests that PLRV has reached an equilibrium although it seems to be able to readily adapt when subjected to environmental changes (Guyader *et al.*, 2004).

2.2   Diagnostic techniques for the detection of viruses in potatoes

With the increase in the use of molecular techniques in recent years, many methods can be used for laboratory diagnosis of potato viruses and specifically PLRV, but currently mainly ELISA and RT-PCR are used (Klerks *et al.*, 2001; Bystricka *et al.*, 2005).

ELISA was introduced in the late 1970s and greatly facilitated the detection of plant viruses (Robert *et al.*, 2000). This was possible after the development of a method of purifying milligram quantities of particles of luteoviruses and subsequent preparation of luteovirus antisera suitable for use in ELISA (Tamada *et al.*, 1984). ELISA allows specific, sensitive and quick detection of most potato viruses and is suitable for large-scale routine testing as required for certification (Robert *et al.*, 2000). PLRV is readily detected by ELISA in leaves of infected plants (Spiegel and Martin, 1993).

One of the disadvantages of ELISA is that the direct double-antibody sandwich (DAS) form is sensitive to differences in the antigenic specificity of isolates, and in some conditions it can fail to detect antigenic variants that are detected by other kinds of serological tests (Tamada *et al.*, 1984). However, very little variation in the antigenic specificity of isolates of PLRV was found, indicating that the problems encountered with DAS-ELISA detection of other viruses are not a problem in the detection of PLRV. Torrance (1992) found that as long as polyclonal antisera which recognized a number of antigenic determinants, were used in DAS-ELISA, the virus could be readily detected using this method.

Another disadvantage of ELISA is that it can only be reliably performed on sprouted tubers, for ELISA is not able to detect low-level positive samples in dormant tubers (Spiegel and Martin, 1993; Robert *et al.*, 2000; Mumford *et al.*, 2004). Hence, most potato certification programs rely on winter "grow-out" tests to ascertain the virus status of seed lots (Spiegel and Martin, 1993). This causes a delay in the length of time required to complete testing (i.e. 6 to 8 weeks) (Mumford *et al.*, 2004).

Detection of viruses in dormant tubers was not possible until the 1990's with the introduction of RT-PCR (Singh *et al.*, 2004). RT-PCR assays were developed for PLRV, *Potato virus S* and

Potato virus X (Singh *et al.*, 2004). Sensitive RT-PCR assays for PLRV were developed with primers that bind to areas in ORF4 (Singh *et al.*, 2000). With the development of RT-PCR, PLRV could be detected in dormant tubers, thus reducing the time for completion of the diagnostic test, which was not possible with ELISA (Klerks *et al.*, 2001).

RT-PCR is generally 10^2 to 10^5 times more sensitive than ELISA and this sensitivity is of particular importance when viruses occur in low concentrations such as in dormant tubers with concentrations too low for detection by ELISA (Spiegel and Martin, 1993; Dietzgen, 2002; Mumford *et al.*, 2004). The potential power and sensitivity of RT-PCR is due to the exponential amplification of small amounts of specific target DNA sequences *in vitro* (Dietzgen, 2002). For this reason RT-PCR is the method of choice for applications such as diagnosis of plant viruses present at levels below the detection limit of ELISA (Dietzgen, 2002). Specific RT-PCR tests have been developed for the most economically important plant viruses (Dietzgen, 2002).

There are many variations and applications of RT-PCR to attest to its enormous versatility (Dietzgen, 2002). For instance, detection can be made more sensitive and more specific by using immune-capture PCR in which virus particles are bound to a plate coated with a specific antibody and the RNA is extracted from the bound virus and subjected to RT-PCR (Mayo and Ziegler-Graff, 1996). This method is also effective for detecting PLRV (Mayo and Ziegler-Graff, 1996). Immune-capture PCR appears to the method of choice when specific antisera are available and the highest sensitivity is required (10^4 to 10^6 times more sensitive than ELISA) in applications such as certification of virus-free planting material (Dietzgen, 2002).

With the use of a combination of several primer pairs in the same RT-PCR reaction (multiplex PCR, or multiplex RT-PCR) more than one virus can be detected simultaneously in the same specimen (Dietzgen, 2002; Singh *et al.*, 2004). This technique saves time and reagents and each PCR product has a unique size by design, so it can be differentiated from the others by gel electrophoresis (Dietzgen, 2002). Currently, more rapid and more sensitive detection techniques (e.g. real-time PCR) are being developed with molecular beacon probes to simultaneous amplify and detect amplicons in a sealed tube (Klerks *et al.*, 2001; Dietzgen, 2002). In 2004, Mumford *et al.* developed a direct tuber testing method based upon real-time PCR for the detection of potato viruses (e.g. PLRV) that allows extremely sensitive detection and overcomes many of the problems associated with conventional PCR, in particular the need for post-PCR analysis. Further advances in this method are underway (Mumford *et al.*, 2004).

However, even though RT-PCR can detect one or a few (multiplex RT-PCR) pathogens in one reaction, this method is too costly and not suitable for routine detection (Klerks *et al.*, 2001; Bystricka *et al.*, 2005). The high sensitivity of RT-PCR requires special care in setting up and analyzing RT-PCR reactions to avoid contamination (Dietzgen, 2002). This is a disadvantage as significant investment of skilled operators, expensive equipment and expensive reagents may make RT-PCR technology cost-prohibitive in some countries or in some regions of a country

(Singh *et al.*, 2004). Another disadvantage of RT-PCR technology as a diagnostic tool for PLRV detection is the difficulty associated with the visualization method. The DNA fragments produced during RT-PCR are most commonly visualized in the research laboratory with gel electrophoresis followed by staining with ethidium bromide (Dietzgen, 2002). However, when testing many samples in a diagnostic laboratory it is quite laborious, expensive and time-consuming and does not easily lend itself to automation (Dietzgen, 2002; Mukherjee *et al.*, 2003).

In conclusion, although RT-PCR technology has the potential for sensitive detection of potato viruses, it is labour intensive when many samples have to be processed and therefore has not yet been applied for large-scale, routine testing (Robert *et al.*, 2000). Real-time PCR developments are underway, but cost versus benefit aspects need to be better defined for seed potato certification in relation to RT-PCR and ELISA (Klerks *et al.*, 2001). Although ELISA is not as sensitive as RT-PCR, its sensitivity is still acceptable, it is cost effective and therefore presently the method of choice for routine PLRV detection. Currently, for large scale testing of potato seed tubers ELISA is the most cost effective. In cases where the ELISA results are ambiguous, sample results could be verified with the more expensive and labour intensive RT-PCR based techniques.

2.3 Production of antibodies using conventional methods

Crucial for successful ELISA development is the production of antibodies that have high specificity and avidity. The conventional method to produce antibodies to PLRV is to purify PLRV from potato plants for immunization. There are, however, many disadvantages and complications associated with this approach to antibody production.

As PLRV is confined to the phloem and is present in very low amounts, purification is difficult and virus yield extremely low (Faccioli *et al.*, 1995; Johnson and Pappu, 2006). Furthermore, the PLRV purification procedure is time-consuming and complex (Faccioli *et al.*, 1995; Mukherjee *et al.*, 2003). When antiserum is prepared for use in ELISA, a major limitation is that antisera with relatively high titres to virus particles without measurable titres to other plant materials are desirable (Tamada and Harrison, 1980). In spite of this, ELISA is used for regular control of viral contamination of potatoes in propagating programs due to its specificity and sensitivity (Joerdens-Roettger, 1987). This method is routinely used all over the world in seed potato production schemes (Joerdens-Roettger, 1987). However, because of the problems associated with antibody production to PLRV, i.e. antibody production to co-isolated plant material, other means of producing antisera suitable for ELISA have been investigated by workers from around the world.

2.4 Production of antibodies to recombinant viral CPs

The widespread use of PCR and the increasing number of available viral sequences has enabled many laboratories to amplify and clone viral genes into expression vectors, purify the recombinant proteins and produce antibodies (Vaira *et al.*, 1996). It is especially the cloning and sequencing of viral CP genes that gave the opportunity to utilize molecular biology methods as an alternative approach to obtaining viral CP antigens for antiserum production (Nikolaeva *et al.*, 1995; Ling *et al.*, 2000).

When virus particles are used for immunization, antibodies are elicited against the antigenic determinants of the viral CPs (Hélias *et al.*, 2003). Recombinant CPs produced in bacterial cells are therefore suitable alternatives to the use of purified virus for antiserum production for detection purposes (Hélias *et al.*, 2003). The expression of viral CPs in prokaryotic systems followed by purification for polyclonal antiserum production for use in western blots and different types of ELISA, has been reported for a number of different plant viruses (Jelkmann and Keim-Konrad, 1997).

This use of recombinant CPs as an alternative for viral antigens in the production of diagnostic antibodies is often sought for plant viruses that are present in the host in low concentration and that are difficult to culture in host plants and to purify (Korimbocus *et al.*, 2002; Abou-Jawdah *et al.*, 2004). For example, closteroviruses are also phloem limited with an overall low titer in infected plants like PLRV, making it difficult to purify sufficient quantities for production of high quality antibodies (Hourani and Abou-Jawdah, 2003). Preparations of one such closterovirus, *Citrus tristeza virus* (CTV), are still contaminated with host components even with the best methods to produce CTV preparations resulting in polyclonal antisera that are unsatisfactory for the accurate diagnosis of CTV (Nikolaeva *et al.*, 1995). Recombinant DNA technology allowed the successful production of specific antibodies to several closteroviruses enabling these problems to be overcome (Hourani and Abou-Jawdah, 2003).

Recombinant DNA technology used for production of the immunogen was also used for polyclonal antiserum production to *Grapevine leafroll associated closterovirus*-3 (GLRaV-3) (Ling *et al.*, 2000). Yields of purified closteroviruses are generally low and it is complicated to produce true-to-type polyclonal antisera because mixed infections of different types of grapevine leafroll associated closteroviruses in a field collected sample are common, and most of these viruses cannot be separated biologically (Ling *et al.*, 2000). Thus, molecular cloning obviated virus purification (Ling *et al.*, 2000).

Recombinant DNA technology was also sought as an alternative for antigen preparation for antiserum production in the case of *Prune dwarf virus* (PDV) (Abou-Jawdah *et al.*, 2004). These viral particles are labile and extraction and purification in sufficient quantities with high purity are especially difficult (Abou-Jawdah *et al.*, 2004). This technique was also sought to reduce

problems that may arise from mixed infections of plant tissues from which the virus is purified (Abou-Jawdah *et al.*, 2004).

For tuberose mild mosaic virus, it has been shown that antiserum can be prepared by immunizing the viral coat protein expressed in bacteria (Chen *et al.*, 2002). The bacterially expressed protein was preserved in an antigenically indistinguishable form from the native viral CP, although some antibodies were formed against the amino acid sequences originating from the expression vector (Chen *et al.*, 2002).

For the production of antibodies to viruses using recombinant viral CPs formed in bacteria, various strategies have been used. Recombinant viral CPs were either produced on their own, as fusion or with a 6 histidine-tag attached to it, as summarized in Table 1. Not all of the antibodies produced against these recombinant viral CPs could be used successfully for the various applications as listed in Table 1.

Table 1 a, b, c. Literature summary of various approaches for the production of antibodies to recombinant viral CPs. All of the recombinant viral CPs were used for the production of antibodies and applied in either DAS-ELISA, indirect-ELISA and/or western blot. Some of the antibodies formed to recombinant viral CPs were effective for the detection of native viruses in the above mentioned applications, others were not.

Table 1 a. The production of antibodies to free recombinant viral CPs.

Free recombinant viral CPs			
Virus name	**Expected size of expressed protein**	**Application**	**Reference**
Potato virus A (PVA)	32 kDa	Double-antibody sandwich (DAS) enzyme-linked immunosorbent assay (ELISA) (coating with antibody): not effective for detection of PVA. Indirect-ELISA (coating with antigen): effective for detection of PVA. Western-blot: effective for detection of PVA.	Čeřovská *et al.* (2002)

Table 1 b. The production of antibodies to 6 histidine-tagged recombinant viral CPs.

6 Histidine-tagged recombinant viral CPs			
Virus name	**Expected size of expressed protein**	**Application**	**Reference**
Strawberry mild yellow edge associated potexvirus (SMYEAV)	26 kDa	DAS-ELISA: effective for detection of SMYEAV.	Kaden-Kreuziger *et al.* (1995)
Prune dwarf virus (PDV)	25-30 kDa	DAS-ELISA: effective for detection of PDV. Indirect-ELISA: effective for detection of PDV. Western-blot: effective for detection of PDV.	Abou-Jawdah *et al.* (2004)
Cucurbit yellow stunting disorder virus (CYSDV)	28.5 kDa	DAS-ELISA: medium effectivity for detection of CYSDV. Indirect-ELISA: effective for detection of CYSDV.	Hourani and Abou-Jawdah (2003)
Grapevine rupestris stem pitting associated virus (GRSPaV)	28 kDa	DAS-ELISA: not effective for detection of GRSPaV. Indirect-ELISA: not effective for detection of GRSPaV. Western-blot: effective for detection of GRSPaV.	Minafra *et al.* (2000)

| Apple stem pitting virus (ASPV) | 44 kDa | DAS-ELISA: not effective for detection of ASPV. Indirect-ELISA: medium effectivity for detection of ASPV. | Jelkmann and Keim-Konrad (1997) |
| Faba bean nectrotic yellow virus (FBNYV) | 45 kDa | DAS-ELISA: not effective for detection of FBNYV. Western-blot: effective for detection of FBNYV. | Kumari et al. (2001) |

Table 1 c. The production of antibodies to fusion protein recombinant viral CPs.

Fusion protein recombinant viral CPs				
Virus name	**Fusion protein**	**Expected size of expressed protein**	**Application**	**Reference**
Tomato spotted wilt tospovirus (TSWV)	GST at N-terminal of CP.	55 kDa	DAS-ELISA: effective for detection of TSWV. Indirect-ELISA: effective for detection of TSWV. Western-blot: effective for detection of TSWV.	Vaira et al. (1996)
Grapevine leafroll associated closterovirus-3 (GLRaV-3)	β-galactosidase at C-terminal of CP.	43 kDa	DAS-ELISA: effective for detection of GLRaV-3. Western-blot: effective for detection of GLRaV-3.	Ling et al. (2000)
Citrus tristeza virus (CTV)	Fragment of maltose-binding protein (MBP) at N-terminal of CP.	67 kDa	Indirect-ELISA: effective for detection of CTV. Western-blot: effective for detection of CTV.	Nikolaeva et al. (1995)
Rupestris stem pitting associated virus (RSPaV)	MBP at N-terminal of CP.	71 kDa	DAS-ELISA: not effective for detection of RSPaV. Indirect-ELISA: effective for detection of RSPaV. Western-blot: effective for detection of RSPaV.	Meng et al. (2003)
SMYEAV	GST at N-terminal of CP.	53 kDa	DAS-ELISA: not effective for detection of SMYEAV.	Kaden-Kreuziger et al. (1995)

Long term storage of recombinant DNAs for the use of recombinant proteins as a pure and stable source of antigen, is easier than storage of viable virus isolates (Vaira et al., 1996; Abou-Jawdah et al., 2004). Recombinant clones can be preserved in expression vectors and stored indefinitely at -80°C and reactivated when needed (Kumari et al., 2001). In contrast to the conventional method of antigen preparation (virion purification following virus propagation in a greenhouse), which can easily take 2 or more months, recombinant viral CPs can be purified in an average of 3 days and is thus far more economical (Kumari et al., 2001; Čeřovská et al., 2002; Čeřovská et al., 2003). One liter of bacterial culture fluid typically yields an amount sufficient for a repeated immunization of laboratory animals for antisera preparation (Kumari et al., 2001; Čeřovská et al., 2002; Čeřovská et al., 2003). Thus, in comparison with virus purification, recombinant DNA technology is by far superior in many respects.

2.4.1 Protein expression using the bacterial system

2.4.1.1 The pET expression system

The bacteriophage T7 RNA polymerase-based pET expression system (plasmid for expression by T7 RNA polymerase) is one of the most powerful and widely used expression systems (Studier and Moffatt, 1986; Rosenberg *et al.*, 1987; Pan and Malcolm, 2000). The plasmid vector has been developed for cloning and expression of target DNAs under control of a T7 promoter (Studier *et al.*, 1990).

Bacteriophage T7 RNA polymerase is highly selective for its own promoters, which does not occur naturally in *E. coli* (Studier and Moffatt, 1986; Studier *et al.*, 1990). The conserved sequence required to make an active promoter for T7 RNA polymerase is long enough that it is unlikely to occur by chance in any DNA unrelated to T7 DNA (Studier and Moffatt, 1986). Furthermore, this enzyme is very active and elongates chains about five times faster than *E. coli* RNA polymerase does (Studier *et al.*, 1990). T7 RNA polymerase also terminates transcription less frequently (Tabor, 1990). The result of the selectivity and activity of T7 RNA polymerase is that the enzyme is able to produce complete transcripts from almost any DNA that is linked to a T7 promoter (Studier and Moffatt, 1986).

Only small amounts of T7 RNA polymerase are needed to generate very large amounts of RNA from DNA that is linked to a T7 promoter (Studier and Moffatt, 1986). T7 RNA polymerase seems to interfere with transcription by *E. coli* RNA polymerase and in favourable circumstances essentially all of the resources of the cell can become concentrated on the production of an individual protein (Studier and Moffatt, 1986). As a result this protein can accumulate rapidly to reach 50% or more of the total cell protein (Studier and Moffatt, 1986). However, the polymerase is so active that even a small basal level of expression in the uninduced cell can prevent relatively toxic target genes from being established and even innocuous proteins can cause the cell to stop growing and dividing as the level of transcription of target DNA becomes so high (Studier, 1991). For this reason bacteriophage T7 lysozyme, which inhibits T7 RNA polymerase, has been used to reduce leaky expression (Studier, 1991; Kelly *et al.*, 1995).

Bacteriophage T7 lysozyme is a natural, selective inhibitor of T7 RNA polymerase by binding to it, a feed-back mechanism that ensures a controlled burst of transcription during T7 infection (Studier, 1991). This protein can be supplied in low level by a pLysS plasmid, which is compatible with the pET vectors for expressing genes from a T7 promoter (Studier, 1991). These low levels of T7 lysozyme are sufficient to stabilize many target plasmids and yet allow high levels of target proteins to be produced upon production of T7 RNA polymerase (Studier, 1991). (It is important to note that T4 lysozyme and vertebrate lysozymes are glycoside hydrolases that cleave glycosidic bonds between N-acetyl muraminic acid and the fourth carbon

of N-acetylglucosamine in peptidoglycans in the cell wall of bacteria but that T7 lysozyme cleaves peptide bonds to cleave the lipoprotein from these peptidoglycans). In addition, the presence of T7 lysozyme has the advantage of facilitating the lysis of cells in preparing extracts for purification of target gene products as this protein also cuts a peptide bond in the peptidoglycan layer in the cell wall of *E. coli* (Inouye *et al.*, 1973; Studier, 1991). However, the accumulation of T7 lysozyme does not prevent growth of *E. coli* cells, apparently because the lysozyme cannot penetrate the inner membrane to reach its substrate in the peptidoglycan layer (Moffatt and Studier, 1987). The enzyme only facilitates the preparation of cell extracts for purification of target proteins when cells are treated to disrupt their membranes (Studier, 1991).

2.4.1.2 Rare codons in protein expression

As PLRV CPs have an N-terminal arginine rich domain, expression of the viral CP can be problematic since the codons for arginine are rare in *E. coli*. The occurrence of codons in genes that are used infrequently in *E. coli* (so-called rare codons) can be one of the reasons for a lack of expression in recombinant protein expression systems (Schendel, 1998). Within *E. coli* and other species there is a strong positive correlation between the frequency of amino acids encoded in its genes and the amount of their respective tRNAs (Ikemura, 1981). For this reason a bias exists among the 61 amino acid codons for the production of most mRNA molecules in *E. coli* (Kane, 1995; Novy *et al.*, 2001). Codons that occur in highly expressed genes are called major codons and those that occur in genes expressed at a low level are called minor or rare codons (Kane, 1995). Rare codons appear to cause problems mostly in the translation of proteins (Kane, 1995). One of the negative effects on translation could be by causing ribosomes to pause, which in effect can uncouple transcription from translation, leading to premature termination of the message (Schendel, 1998). Even if transcription would proceed normally after pausing, the mRNA 3' to the stalled ribosomes can be exposed to degradation by host ribonucleases, reducing its stability (Schendel, 1998). During the overexpression of heterologous target genes in *E. coli*, insufficient tRNA pools can also lead to premature translation termination, translational frameshifting and amino acid misincorporation (Novy *et al.*, 2001).

Arginine codons account for four of the top six rare codons (Kane, 1995). In particular, the arginine codons AGA and AGG, the product of the *dnaY* gene, are the least used codons in *E. coli* (Chen and Inouye, 1990). In *E. coli* mRNA these rare arginine codons (AGA and AGG) occur at a frequency of ~0.21% and 0.14%, respectively (Kane, 1995). These codons were the first rare codons demonstrated to have a detrimental effect on protein expression (Kane, 1995). Clusters of AGA/AGG codons can reduce the quantity and quality of the synthesized protein and if there is an excess of these codons, even without clusters, it could create translational problems (Kane, 1995).

The magnitude of the effect of these rare aginine codons depends not only upon their number but also their position within the mRNA (Kane, 1995). It has been found that clusters of two to five AGG codons have a greater effect on decreasing protein expression as the number of AGG codons increased (Kane, 1995). Furthermore, effects on protein expression decrease as the AGG clusters approach the carboxyl terminus (Kane, 1995). Chen and Inouye (1990) found that the negative effect of AGG codons on gene expression was dependent upon the distance between the site of the AGG codons and the initiation codon. As the distance between the initiation codon and the AGG codons increased, protein production increased almost linearly (Chen and Inouye, 1990). Gonzalez de Valdivia and Isaksson (2004) reported that consecutive AGA or AGG codons near the 5' end of the message give significantly lowered gene expression and if these codon sequences are further down in the gene there is no apparent effect on gene expression. A codon combination of AGA AGG at an early position or a consecutive row of AGG codons downstream in a gene can give a negative effect on gene expression (Gonzalez de Valdivia and Isaksson, 2004). Even a single AGG codon, if located in the early coding region, is enough to give a negative effect on gene expression (Gonzalez de Valdivia and Isaksson, 2004).

Brinkmann *et al*. (1989) found that supplementation with tRNA $^{Arg}_{AGA/AGG}$ by cotransfection with the *dnaY* gene, which supplies this minor tRNA, results in high-level production of proteins that are not readily expressed in *E. coli*. The *dnaY* gene has been inserted into compatible plasmids pRARE and pRARE2 by Novagen to supply the tRNA that recognizes AGA/AGG codons for use in recombinant protein production (Novy *et al*., 2001). Rosetta-gami B(DE3)pLysS strains contain the pRARE plasmid while Rosetta-2(DE3)pLysS strains contain the pRARE2 plasmid (Anon, 2006).

2.4.1.3 The pGEX expression system

For the expression of genes in *E. coli* using recombinant DNA technology, it was thought that a strong promoter and a start codon at the beginning of the gene would be sufficient for good expression (Riggs and La Vallie, 1994). However, the requirements for good expression are often more complicated (Riggs and La Vallie, 1994). For one, the level of expression is affected by codon preferences and may be affected by the coding sequence in other ways that are not yet well understood (Riggs and La Vallie, 1994). Problems with expression at the beginning of the gene can be solved by a variety of ways and the quickest is by introducing the cloned gene into an expression vector 3' to a carrier sequence coding for the amino terminus of a highly expressed protein (carrier protein) (Riggs and La Vallie, 1994). The carrier sequence can be from an *E. coli* gene or any gene that is strongly expressed in *E. coli* (Riggs and La Vallie, 1994). This sequence provides the necessary signals for good expression resulting in a fusion protein that contains an N-terminal region encoded by the carrier (Riggs and La Vallie, 1994).

The carrier sequence can code for an entire protein such as glutathione-*S*-transferase (GST) (Riggs and La Vallie, 1994).

The GST gene fusion system is one of the more extensively used systems for the expression and purification of recombinant proteins in *E. coli* (Saluta and Bell, 1998). Foreign polypeptides are expressed as fusion proteins linked to the C-terminus of GST using pGEX vectors (Smith and Corcoran, 1994). The pGEX vectors are designed so that foreign polypeptides can be expressed in *E. coli* in a form that allows them to be purified rapidly under nondenaturing conditions (Smith and Corcoran, 1994). The GST gene used to generate the pGEX vectors was originally cloned from the parasitic helminth, *Schistosoma japonicum* and the GST protein is a common 26 kDa cytoplasmic protein of eukaryotes (Smith and Corcoran, 1994).

The main advantage of this system is that most fusion proteins remain soluble and stable even at high levels of expression (Smith and Corcoran, 1994; Smith, 2000; Terpe, 2003). Furthermore, purification can be performed under non-denaturing conditions by affinity chromatography (Smith and Corcoran, 1994; Smith, 2000). Affinity chromoatography with immobilized glutathione for the purification of fusion proteins from crude lysate is used because of the affinity of the GST moiety for glutathione (Smith and Corcoran, 1994; Terpe, 2003). Thereafter, the bound fusion proteins can be eluted with free reduced glutathione at neutral pH (Smith and Corcoran, 1994; Terpe, 2003).

As the GST protein is also antigenic, it can be disadvantageous for the production of antigens to the fusion protein (Yahalom and Chamovitz, 2002). This can be side-stepped by cleaving the GST moiety from the fusion protein with site-specific proteases (Smith and Corcoran, 1994; Yahalom and Chamovitz, 2002). The pGEX-6P-1, 2 and 3 vectors (Amersham) include a peptide sequence that is optimally cleaved by the human rhinovirus 3C protease (Smith, 2000). The PreScission protease (Amersham) contains the human rhinovirus 3C protease including the GST-tag so that both the GST carrier and the protease can be removed after proteolysis by affinity chromatography on glutathione-agarose (Smith, 2000; Terpe, 2003).

Uses of the GST fusion proteins produced from pGEX vectors include antigens for immunological or vaccination studies or as a way of providing functionally active enzymes for biochemical or structural studies (Terpe, 2003; Smith, 2000).

2.5 References

Abou-Jawdah, Y., Sobh, H., Cordahi, N., Kawtharani, H., Nemer, G., Maxwell, D.P. and Nakhla, M.K. (2004). Immunodiagnosis of prune dwarf virus using antiserum produced to its recombinant coat protein. *J. Virol. Methods*, **121**: 31-38.

Anon. (2004a). Seed potato industry in dire straits. *Chips (mouthpiece of the South African potato industry)*, May/June 2004, **18**(3): 5.

Anon. (2004b). Beheer van rolblad. *Chips (mouthpiece of the South African potato industry)*, May/June 2004, **18**(3): 44-46.

Anon. (2005). Die bestuur van die virusprobleem. *Chips (mouthpiece of the South African potato industry)*, May/June 2005, **19**(3): 42-46.

Anon. (2006). Novagen competent cells: what a difference a strain makes. *EMB Biosciences*, pp. 1-12. Available: www.emdbiosciences.com/novagen (Accessed on 17 January 2007).

Ashoub, A., Rohde, W. and Prüfer, D. (1998). *In planta* transcription of a second subgenomic RNA increases the complexity of the subgroup 2 luteovirus genome. *Nucleic Acids Res.*, **26**(2): 420-426.

Bahner, I., Lamb, J., Mayo, M.A. and Hay, R.T. (1990). Expression of the genome of potato leafroll virus: readthrough of the coat protein termination codon *in vivo. J. Gen. Virol.*, **71**: 2251-2256.

Beckett, B.S., Hands, R.A., Lewis, M., Phillips, W.D., Beardsley, T., Bender, L., Cupid, B., Clark, J., Clarke, H.M., Daintith, E.K., Hart, M., Hine, R.S., Holmes, E., Lockwood, A., Neal, J.V., Prince, R.A., Ruse, M., Stratton, B. and Tootill, E. (2000). *Oxford Dictionary of Biology.* Martin, E. and Hine, R.S. (eds.), Oxford University Press, New York, pp. 137, 371, 440, 454, 468, 545, 615.

Bonthuys, J. (2005). Weer en virusse kan boere in moerbedryf op die knieë bring. *Die Burger*, 14 May 2005, pp. 10.

Brinkmann, U., Mattes, R.E. and Buckel, P. (1989). High-level expression of recombinant genes in *Escherichia coli* is dependent on the availability of the *dnaY* gene product. *Gene*, **85**: 109-114.

Bustamante, P.I. and Hull, R. (1998). Plant virus gene expression strategies. *EJB*, **1**(2): 65-82.

Bystricka, D., Lenz, O., Mraz, I., Piherova, L., Kmoch, S. and Sip, M. (2005). Oligonucleotide-based microarray: A new improvement in microarray detection of plant viruses. *J. Virol. Methods*, **128**: 179-182.

Carrington, J.C., Kasschau, K.D., Mahajan, S.K. and Schaad, M.C. (1996). Cell-to-cell and long-distance transport of viruses in plants. *The Plant Cell*, **8**: 1669-1681.

Čeřovská, N., Moravec, T. and Velemínský, J. (2002). Polyclonal antibodies to a recombinant coat protein of potato virus A. *Acta Virol.*, **46**: 147-151.

Čeřovská, N., Moravec, T., Rosecká, P., Dědič, P. and Filigarova, M. (2003). Production of polyclonal antibodies to a recombinant coat protein of *Potato mop-top virus. J. Phytopathol.*, **151**: 195-200.

Chen, G.T. and Inouye, M. (1990). Suppression of the negative effect of minor arginine codons on gene expression; preferential usage of minor codons within the first 25 codons of the *Escherichia coli* genes. *Nucleic Acids Res.*, **18**(6): 1465-1473.

Chen, C., Hsaing, T., Chaing, F. and Chang, C. (2002). Molecular characterization of *Tuberose mild mosaic virus* and preparation of its antiserum to the coat protein expressed in bacteria. *Bot. Bull. Acad. Sin.*, **43**: 13-20.

Coetsee, J. (2004). Rolblad maai onder aartappels. *Landbouweekblad*, 12 Nov: 26-27.

Coetsee, J. (2005). Virusse bedreig hele aartappelbedryf. *Landbouweekblad*, 9 Sept: 44-45.

Commandeur, U., Rohde, W., Fischer, R. and Prüfer, D. (2002). Gene expression strategies of RNA viruses. In: *Plant viruses as molecular pathogens.* Khan, J.A. and Dijkstra, J. (eds.), The Haworth Press, United States of America, pp. 175-201.

De Souza-Dias, J.A.C., Russo, P., Betti, J.A., Miller, L. and Slack, S.A. (1999a). Simplified extraction method for ELISA and PCR detection of potato leafroll luteovirus primary infection in dormant potato tubers. *Am. J. Potato Res.*, **76**: 209-213.

De Souza-Dias, J.A.C., Russo, P., Miller, L. and Slack, S.A. (1999b). Comparison of nucleotide sequences from three potato leafroll virus (PLRV) isolates collected in Brazil. *Am. J. Potato Res.*, **76**: 17-24.

Dietzgen, R.G. (2002). Application of PCR in plant virology. In: *Plant viruses as molecular pathogens*. Khan, J.A. and Dijkstra, J. (eds.), The Haworth Press, United States of America, pp. 471-500.

Faccioli, G., Rosner, A. and Forni, M. (1995). Use of the polymerase chain reaction to clone the potato leafroll virus coat protein gene directly from the total RNA of infected plants. *Potato Res.*, **38**: 211-218.

Fisken, A.G. (1959). Factors affecting the spread of aphid-borne viruses in potato in eastern Scotland. *Ann. Appl. Biol.*, **47**(2): 264-273.

Gabriel, W. (1965). The influence of temperature in the spread of aphid-borne potato virus diseases. *Ann. Appl. Biol.*, **56**: 461-475.

Garret A., Kerlan, C. and Thomas, D. (1993). The intestine is a site of passage for potato leafroll virus from the gut lumen into the haemocoel in the aphid vector, *Myzus persicae* Sulz. *Arch. Virol.*, **131**: 377-392.

Gonzalez de Valdivia, E.I. and Isaksson, L.A. (2004). A codon window in mRNA downstream of the initiation codon where NGG codons give strongly reduced gene expression in *Escherichia coli*. *Nucleic Acids Res.*, **32**(17): 5198-5205.

Guyader, S. and Ducray, D. G. (2002). Sequence analysis of potato leafroll virus isolates reveals genetic stability, major evolutionary events and differential selection pressure between overlapping reading frame products. *J. Gen. Virol.*, **83**: 1799-1807.

Guyader, S., Tanguy, S., Andrieu, A., Pasquer, F. and Ducray, D. G. (2004). Is *potato leafroll virus* (PLRV) a slowly evolving RNA virus? *12th EAPR Virology Section Meeting Rennes, France, 2004: abstracts of oral presentations and poster presentations.* Available: http://www.rennes.inra.fr/eapr2004/abstracts.htm (Accessed on 24 February 2005).

Haliloglu, K. and Bostan, H. (2002). Nucleotide sequence analysis for assessment of variability of potato leafroll virus and phylogenetic comparisons. *OnLine Journal of Biological Sciences*, **2**(9): 582-586.

Haupt, S., Stroganova, T., Ryabov, E., Kim, S.H., Fraser, G., Duncan, G., Mayo, M.A., Barker, H. and Taliansky, M. (2005). Nucleolar localization of potato leafroll virus capsid proteins. *J. Gen. Virol.*, **86**: 2891-2896.

Hélias, V., Jacquot, E., Guillet, M., Le Hingrat, Y. and Giblot-Ducray, D. (2003). Production of recombinant *Potato mop-top virus* coat protein in *Escherichia coli* and generation of antisera recognizing native virus protein. *J. Virol. Methods*, **110**: 91-97.

Herrbach, E. (1992). Alarm pheromones and allelochemicals as a means of aphid control. *Neth. J. Pl. Path.*, **98**(2): 63-71.

Hille Ris Lambers, D. (1955). Potato aphids and virus diseases in the Netherlands. *Ann. Appl. Biol.*, **42**: 355-360.

Hourani, H. and Abou-Jawdah, Y. (2003). Immunodiagnosis of cucurbit yellow stunting disorder virus using polyclonal antibodies developed against recombinant coat protein. *J. Plant Pathol.*, **85**(3): 197-204.

Ikemura, T. (1981). Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes. *J. Mol. Biol.*, **146**: 1-21.

Inouye, M., Arnheim, N. and Sternglanz, R. (1973). Bacteriophage T7 lysozyme is an *N*-acetylmuramyl-L-alanine amidase. *J. Biol. Chem.*, **248**(20): 7247-7252.

Jelkmann, W. and Keim-Konrad, R. (1997). Immuno-capture polymerase chain reaction and plate-trapped ELISA for the detection of apple stem pitting virus. *J. Phytopathol.*, **145**: 499-503.

Joerdens-Roettger, D. (1987). *Seed potato production in the Philippines: in vitro culture and rapid multiplication techniques.* Deutsche Gesellschaft für Technische Zusammenarbeit (GTZ) GmbH, Germany, pp. 25-26.

Johnson, D.A. and Pappu, H. (2006). Potato leaf roll. Available: http://www.wsu.edu/~potatoes/leafroll.htm (Accessed on 11 August 2006).

Johnson, S.B. (2003). Potato facts: potato diseases caused by PVY and PLRV. Available: http://www.umext.maine.edu/onlinepubs/PDFpubs/2492.pdf (Accessed on 18 February 2006).

Jolly, C.A. and Mayo, M.A. (1994). Changes in amino acid sequence of the coat protein readthrough domain of potato leafroll luteovirus affect the formation of an epitope and aphid transmission. *Virology*, **201**: 182-185.

Kane, J.F. (1995). Effects of rare codon clusters on high-level expression of heterologous proteins in *Escherichia coli. Curr. Opin. Biotechnol.*, **6**: 494-500.

Kaniewski, W.K. and Thomas, P.E. (2004). The potato story. *AgBioForum*, **7**(1&2): 41-46.

Kassanis, B. (1952). Some factors affecting the transmission of leaf-roll virus by aphids. *App. Biol.*, **39**(2): 157-167.

Kawchuk, L.M., Martin, R.R., Rochon, D.M. and McPherson, J. (1989). Identification and characterization of the potato leafroll virus putative coat protein gene. *J. Gen. Virol.*, **70**: 783-788.

Kawchuk, L.M., Martin, R.R. and McPherson, J. (1990). Resistance in transgenic potato expressing the potato leafroll virus coat protein gene. *Mol. Plant-Microb. Interact.*, **3**(5): 301-307.

Keese, P., Martin, R.R., Kawchuk, L.M., Waterhouse, P.M. and Gerlach, W.L. (1990). Nucleotide sequences of an Australian and a Canadian isolate of potato leafroll luteovirus and their relationships with two European isolates. *J. Gen. Virol.*, **71**: 719-724.

Kelly, K.C., Huestis, K.J., Austen, D.A., Sanderson, C.T., Donoghue, M.A., Stickel, S.K., Kawasaki, E.S. and Osburne, M.S. (1995). Regulation of *sCD4-183* gene expression from phage-T7-based vectors in *Escherichia coli. Gene*, **156**: 33-36.

Klerks, M.M., Leone, G.O.M., Verbeek, M., van den Heuvel, J.F.J.M and Schoen, C.D. (2001). Development of a multiplex AmpliDet RNA for the simultaneous detection of *Potato leafroll virus* and *Potato virus Y* in potato tubers. *J. Virol. Methods,* **93**: 115-125.

Korimbocus, J., Preston, S., Danks, C., Barker, I., Coates, D. and Boonham, N. (2002). Production of monoclonal antibodies to sugarcane yellow leaf virus using recombinant readthrough protein. *J. Phytopathol.*, **150**: 488-494.

Kumari, S.G., Makkouk, K.M., Katul, L. and Vetten, H.J. (2001). Polyclonal antibodies to the bacterially expressed coat protein of faba bean nectrotic yellows virus. *J. Phytopathol.*, **149**: 543-550.

Laubscher, J.M. (2006). Aktiwiteite van virusvektore in die Sandveldgebied (2004). *Chips (mouthpiece of the South African potato industry)*, May/June 2006, **20**(3): 41.

Lee, L., Kaplan, I.B., Ripoll, D.R., Laing, D., Palukaitis, P. and Gray, S.M. (2005). A surface loop of the potato leafroll virus coat protein is involved in virion assembly, systemic movement, and aphid transmission. *J. Virol.*, **79**(2): 1207-1214.

Ling, K., Zhu, H., Jaing, Z. and Gonsalves, D. (2000). Effective application of DAS-ELISA for detection of grapevine leafroll associated closterovirus-3 using a polyclonal antiserum developed from recombinant coat protein. *Eur. J. Plant Pathol.*, **106**: 301-309.

López, L., Muller, R., Balmori, E., De La Riva, G., Ramírez, N., Doreste, V., López, M., Peréz, S., Oramas, P. and Selman-Housein, G. (1994). Molecular cloning and nucleotide sequence of the coat protein gene of a Cuban isolate of potato leafroll virus and its expression in *Escherichia coli. Virus Genes*, **9**(1): 77-83.

Martin, R.R., Keese, P.K., Young, M.J., Waterhouse, P.M. and Gerlach, W.L. (1990). Evolution and molecular biology of luteoviruses. *Annu. Rev. Phytopathol.*, **28**: 341-363.

Massalski, P.R. and Harrison, B.D. (1987). Properties of monoclonal antibodies to potato leafroll luteovirus and their use to distinguish virus isolates differing in aphid transmissibility. *J. Gen. Virol.*, **68**: 1813-1821.

Matzopoulos, M. (2005). The development of enzyme-linked immunosorbent assays to detect potato virus Y and potato leaf roll virus using recombinant viral coat proteins as antigens. M.Sc. thesis, Stellenbosch University.

Mayo, M.A. (2002). The principles and current practice of plant virus taxonomy. In: *Plant viruses as molecular pathogens.* Khan, J.A. and Dijkstra, J. (eds.), The Haworth Press, United States of America, pp. 3-23.

Mayo, M.A. and Ziegler-Graff, V. (1996). Molecular biology of Luteoviruses. *Adv. Virus Res.*, **46**: 413-460.

Mayo, M.A., Robinson, D.J., Jolly, C.A. and Hyman, L. (1989). Nucleotide sequence of potato leafroll luteovirus RNA. *J. Gen. Virol.*, **70**: 1037-1051.

Mayo, M., Ryabov, E., Fraser, G. and Taliansky, M. (2000). Mechanical transmission of *potato leafroll virus. J. Gen. Virol.*, **81**: 2791-2795.

McKay, R. (1955). Potato diseases. Irish Potato Marketing Company, Dublin, pp 74, 76-80, 83-84.

Mehrad, M., Lapierre, H. and Yot, P. (1979). RNA in pototo leafroll virus. *FEBS Letters*, **101**(1): 169-174.

Meng, B., Credi, R., Petrovic, N., Tomazic, I. and Gonsalves, D. (2003). Antiserum to recombinant virus coat protein detects rupestris stem pitting associated virus in grapevines. *Plant Dis.*, **87**(5): 515-522.

Minafra, A., Casati, P., Elicio, V., Rowhani, A., Saldarelli, P., Savino, V. and Martelli, G.P. (2000). Serological detection of grapevine rupestris stem pitting-associated virus (GRSPaV) by a polyclonal antiserum to recombinant virus coat protein. *Vitis*, **39**(3): 115-118.

Moffatt, B.A. and Studier, F.W. (1987). T7 lysozyme inhibits transcription by T7 RNA polymerase. *Cell*, **49**: 221-227.

Mukherjee, K., Verma, Y., Chakrabarti, S.K., Singh, M.N. and Khurana, S.M.P. (2003). Cloning and sequencing of coat protein gene of an Indian potato leaf roll virus (PLRV) isolate and its similarity with other members of luteoviridae. *Virus Genes*, **26**(3): 247-253.

Mumford, R.A., Fisher, T., Elmore, J., Vickers, D., Swan, H., Walsh, K., Barker, I. and Boonham, N. (2004). The development of a routine direct tuber testing method as a rapid and reliable alternative to the traditional growing-on test. *12th EAPR Virology Section Meeting Rennes, France, 2004: abstracts of oral presentations and poster presentations.* Available: http://www.rennes.inra.fr/eapr2004/abstracts.htm (Accessed on 24 February 2005).

Nikolaeva, O.V., Karasev, A.V., Gumph, D.J., Lee, R.F. and Garnsey, S.M. (1995). Production of polyclonal antisera to the coat protein of citrus tristeza virus expressed in *Escherichia coli*: application for immunodiagnosis. *Phytopathology*, **85**(6): 691-694.

Novy, R., Drott, D., Yaeger, K. and Mierendorf, R. (2001). Overcoming the codon bias of *E. coli* for enhanced protein expression. *inNovations*, **12**: 1-3.

Pan, S. and Malcolm, B.A. (2000). Reduced background expression and improved plasmid stability with pET vectors in BL21(DE3). *BioTechniques*, **29**(6): 1234-1238.

Power, A.G. (2000). Insect transmission of plant viruses: a constraint on virus variablility. *Curr. Opin. Plant Biol.*, **3**: 336-340.

Raven, P.H. and Johnson, G.B. (2002). *Biology*, 6th edn., McGraw-Hill, New York, pp. 760, 762, 790.

Reavy, B. and Mayo, M.A. (2002). Persistent transmission of luteoviruses by aphids. *Adv. Bot. Res.,* **36**: 21-46.

Rhode, W., Gramstat, A., Schmitz, J., Tacke, E. and Prüfer, D. (1994). Plant viruses as model systems for the study of non-canonical translation mechanisms in higher plants. *J. Gen. Virol.*, **75**: 2141-2149.

Rich, A.E. (1983). *Potato Diseases.* Academic Press, New York, pp. 1, 2, 95, 113-118.

Riggs, P. and La Vallie, E.R. (1994). Introduction to expression by fusion protein vectors. *Current Protocols in Molecular Biology*, **Supplement 28**: 16.4.1-16.4.4.

Robert, Y., Woodford, J.A.T. and Ducray-Bourdin, D.G. (2000). Some epidemiological approaches to the control of aphid-borne virus diseases in seed potato crops in northern Europe. *Virus Res.*, **71**: 33-47.

Roossinck, M.J. (2003). Plant RNA virus evolution. *Curr. Opin. Microbiol.*, **6**: 406-409.

Rosenberg, A.H., Lade, B.N., Chui, D., Lin, S., Dunn, J.J. and Studier, F.W. (1987). Vectors for selective expression of cloned DNAs by T7 RNA polymerase. *Gene*, **56**: 125-135.

Rouzé-Jouan, J., Terradot, L., Pasquer, F., Tanguy, S. and Ducray-Bourdin, D.G. (2001). The passage of potato leafroll virus through *Myzus persicae* gut membrane regulates transmission efficiency. *J. Gen. Virol.*, **82**: 17-23.

Rowhani, A. and Stace-Smith, R. (1979). Purification and characterization of potato leafroll virus. *Virology*, **98**: 45-54.

Saluta, M. and Bell, P.A. (1998). Troubleshooting GST fusion protein expression in *E. coli. Life Science News I (Amersham Biosciences)*, pp. 1-3.

Schendel, P.F. (1998). Expression of proteins in *Escherichia coli. Current Protocols in Molecular Biology*, **Supplement 41**: 16.1.1.-16.1.3.

Shepardson, S., Esau, K. and McCrum, R. (1980). Ultrastructure of potato leaf phloem infected with potato leafroll virus. *Virology*, **105**: 379-392.

Singh, R.P., Nie, X., Singh, M. (2000). Duplex RT-PCR: reagent concentrations at reverse transcription stage affect the PCR performance. *J. Virol. Methods*, **86**(2): 121-129.

Singh, R.P., Dilworth, A.D., Singh, M. and McLaren, D.L. (2004). Evaluation of a simple membrane-based nucleic acid preparation protocol for RT-PCR detection of potato viruses from aphid and plant tissues. *J. Virol. Methods*, **121**: 163-170.

Smith, D.B. (2000). Generating fusions to glutathione *S*-transferase for protein studies. *Method. Enzymol.*, **326**: 254-270.

Smith, D.B. and Corcoran, L.M. (1994). Expression and purification of glutathione-*S*-transferase fusion proteins. *Current Protocols in Molecular Biology*, **Supplement 28**: 16.7.1-16.7.7.

Spiegel, S. and Martin, R.R. (1993). Improved detection of potato leafroll virus in dormant potato tubers and microtubers by the polymerase chain reaction and ELISA. *Ann. Appl. Biol.*, **122**: 493-500.

Syller, J. (1980). Transmission of potato leaf roll virus by *Myzus persicae* (Sulz.) from leaves and plants of different ages. *Potato Res.*, **23**: 453-456.

Syller, J. (1987). The influence of temperature on transmission of potato leaf roll virus by *Myzus persicae* Sulz.. *Potato Res.*, **30**: 47-58.

Studier, F.W. (1991). Use of bacteriophage T7 lysozyme to improve an inducible T7 expression system. *J. Mol. Biol.*, **219**: 37-44.

Studier, F.W. and Moffatt, B.A. (1986). Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J. Mol. Biol.*, **189**: 113-130.

Studier, W., Rosenberg, A.H., Dunn, J.J. and Dubendorff, J.W. (1990). Use of T7 RNA polymerase to direct expression of cloned genes. *Method. Enzymol.*, **185**: 60-89.

Tabor, S. (1990). Expression using the T7 RNA polymerase/promoter system. *Current Protocols in Molecular Biology*, **Supplement 11**: 16.2.1-16.2.11.

Taliansky, M., Mayo, M.A. and Barker, H. (2003). *Potato leafroll virus*: a classic pathogen shows some new tricks. *Mol. Plant Pathol.*, **4**(2): 81-89.

Tamada, T. and Harrison, B.D. (1980). Factors affecting the detection of potato leafroll virus in potato foliage by enzyme-linked immunosorbent assay. *Ann. Appl. Biol.*, **95**: 209-219.

Tamada, T. and Harrison, B.D. (1981). Quantitative studies on the uptake and retention of potato leafroll virus by aphids in laboratory and field conditions. *Ann. Appl. Biol.*, **98**: 261-276.

Tamada, T., Harrison, B.D. and Roberts, I.M. (1984). Variation among British isolates of potato leafroll virus. *Ann. Appl. Biol.*, **104**: 107-116.

Terpe, K. (2003). Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems. *Appl. Microbiol. Biotechnol.*, **60**: 523-533.

Terradot, L., Souchet, M., Tran, V. and Giblot Ducray-Bourdin, D. (2001). Analysis of a three-dimensional structure of potato leafroll virus coat protein obtained by homology modeling. *Virology*, **286**: 72-82.

Thomas, P.E., Kaniewski, W.K. and Lawson, E.C. (1997). Reduced field spread of potato leafroll virus in potatoes transformed with the potato leafroll virus coat protein gene. *Plant Dis.*, **81**(12): 1447-1453.

Torrance, L. (1992). Analysis of epitopes on potato leafroll virus capsid protein. *Virology*, **191**: 485-489.

Treder, K. and Lewosz, J. (2001). Production of antibodies against potato leafroll virus coat protein expressed in *Escherichia coli* as an antigen. *Proceedings of 'The 11th EAPR Virology Section Meeting', Triest, Czech Republic*, pp. 62-64.

Vaira, A.M., Vecchiati, M., Masenga, V. and Accotto, G.P. (1996). A polyclonal antiserum against a recombinant viral protein combines specificity with versatility. *J. Virol. Methods*, **56**: 209-219.

Van den Heuvel, J.F.J.M., De Blank, C.M., Goldbach, R.W. and Peters, D. (1990). A characterization of epitopes on potato leafroll virus coat protein. *Arch. Virol.*, **115**: 185-197.

Van den Heuvel, J.F.J.M., De Blank, C.M., Peters, D. and Van Lent, J.W.M. (1995). Localization of potato leafroll virus in leaves of secondarily-infected potato plants. *Eur. J. Plant Path.*, **101**: 567-571.

Van den Heuvel, J.F.J.M., Verbeek, M. and Van der Wilk, F. (1994). Endosymbiotic bacteria associated with circulative transmission of potato leafroll virus by *Myzus persicae*. *J. Gen. Virol.*, **75**: 2559-2565.

Van der Wilk, F., Houterman, P., Molthoff, J., Hans, F., Dekker, B., Van den Heuvel, J., Huttinga, H. and Goldbach, R. (1997a). Expression of the potato leafroll virus ORF0 induces viral-disease-like symptoms in transgenic potato plants. *Mol. Plant-Microb. Interact.*, **10**(2): 153-159.

Van der Wilk, F., Verbeek, M., Dullemans, A. M. and Van den Heuvel, J.F.J.M. (1997b). The genome-linked protein of potato leafroll virus is located downstream of the putative protease domain of the ORF1 product. *Virology*, **234**: 300-303.

Van der Wilk, F., Huisman, M.J., Cornelissen, B.J.C., Huttinga, H. and Goldbach, R. (1989). Nucleotide sequence and organization of potato leafroll virus genomic RNA. *FEB*, **245**(1, 2): 51-56.

Van Regenmortel, M.H.V., Fauquet, C.M., Bishop, D.H.L., Carstens, E.B., Estes, M.K., Lemon, S.M., Maniloff, J., Mayo, M.A., McGeoch, D.J., Pringle, C.R. and Wickner, R.B. (2000). Virus Taxonomy Online. Available: http://www.virustaxonomyonline.com (Accessed on 18 February 2006).

Yahalom, A. and Chamovitz, D.A. (2002). Expression of recombinant proteins. In: *Molecular Plant Biology: A Practival Approach*, Volume 2, Gilmartin, P.M and Bowler, C. (eds.), Oxford University Press, United States of America, pp. 111-122.

Chapter 3:

An assessment of the variation in South African isolates of *Potato leafroll virus* based on coat protein gene sequences

Adri H. Rothmann, Mark Matzopoulos, Dirk U. Bellstedt

Department of Biochemistry, Stellenbosch University, Private Bag X1, Stellenbosch 7602, South Africa.

Summary. Presently, the observed variation in symptoms of *Potato leafroll virus* (PLRV) infection in potato cultivars of South Africa cannot be reconciled with PLRV symptoms obtained 10-15 years ago, even if the different interactions between the pathogen and the cultivar are taken into account. In an effort to analyze this variation, the mutation patterns in the coat protein gene of South African isolates of PLRV were assessed. The coat protein gene of PLRV isolates from different areas within South Africa was amplified by reverse transcription polymerase chain reaction (RT-PCR), cloned and sequenced. Significant sequence variation in the coat protein genes was found within the analyzed South African isolates of PLRV. Phylogenetic analysis revealed two major clades with most South African isolates and an Australian and North American isolate grouped together and the remainder grouped with isolates from diverse countries worldwide. The deduced amino acid sequences from representatives of these two clades indicated differences in coat protein three-dimensional structure.

Introduction

PLRV is one of a number of viral pathogens that can impact on seed potato (*Solanum tuberosum* L.) stocks and is considered a major cause for seed potato declassification (De Souza-Dias *et al.*, 1999a; De Souza-Dias *et al.*, 1999b). PLRV is the type species of the genus *Polerovirus*, belonging to the family *Luteoviridae* (Van Regenmortel *et al.*, 2000). The virus is approximately 25 to 30 nm in diameter with an icosahedral shape. The genome of PLRV consists of a single stranded messenger-sense RNA-molecule, of 5882 nucleotides, containing six open reading frames (ORFs) (Mehrad *et al.*, 1979; Van der Wilk *et al.*, 1989; Bahner *et al.*, 1990; Kawchuk *et al.*, 1990; Van Regenmortel *et al.*, 2000; Taliansky *et al.*, 2003). The coding sequences of the ORFs are separated into two clusters of three genes by a noncoding intergenic region of 200 nucleotides (Bahner *et al.*, 1990; Mayo and Ziegler-Graff, 1996; Van der Wilk *et al.*, 1997; Van Regenmortel *et al.*, 2000; Haupt *et al.*, 2005). The 5' half of the luteovirus genome encodes the nonstructural genes (ORF0, ORF1 and ORF2) presumed to be involved in virus replication within infected plant cells whereas the structural genes are located in the 3' half of the genome (Martin *et al.*, 1990). PLRV 3' structural genes include the ORF encoding the coat protein (CP) (ORF3), an ORF embedded in the CP gene postulated to be the VPg (virus protein, genome-linked) (ORF4) and the CP readthrough also detected on the surface of the virion particle (ORF5) (Martin *et al.*, 1990).

In previous examinations of the sequence variation found in PLRV, Keese *et al.* (1990) found that the complete genome of an Australian PLRV isolate differed more (93% similar) from those of Canadian, Scottish and Dutch isolates than they differed among each other (98% similar). In a study by Faccioli *et al.* (1995) sequence comparisons of the CP gene of an Italian isolate with those of five other PLRV isolates revealed a close similarity (up to 98%) to three European (from Scotland, Germany and the Netherlands) and a Canadian isolate, and a greater nucleotide sequence diversity between the Italian and an Australian isolate. Guyader and Ducray (2002) compared the complete genomic sequences of seven isolates collected worldwide with five sequences available in GenBank. Thereafter a restricted polymorphic region of the genome within the ORF0 was analyzed phylogenetically and allowed the detection of three groups of isolates. A first group encompassed three European isolates with a Peruvian isolate. The second group contained exclusively Australian isolates, whereas the third group contained isolates originating from diverse countries and continents. Guyader and Ducray (2002) speculated that the reasons for this diversification could have been the geographical isolation of Australia, since tuber exchanges were presumably reduced with the other continents.

Haliloglu and Bostan (2002) compared the nucleotide sequences of 17 PLRV CP genes and also found that they grouped in two major clusters in a phylogenetic tree. The first cluster included isolates from the Netherlands, USA and Poland. In the second cluster one sub-group

contained isolates from Korea, Scotland and Canada while a second sub-group contained an Australian isolate. Mukherjee *et al.* (2003), in a comparison of the PLRV CP nucleotide sequences of an Indian isolate with those collected worldwide, found that an Austalian CP nucleotide sequence showed the least identity (identities of 97%) whereas the highest nucleotide sequence similarity was found with isolates from UK and Canada (identities of 99%). Furthermore, five completely sequenced PLRV isolates originating from Canada, Scotland, the Netherlands, Poland and Australia were compared by Guyader *et al.* (2004) who found that four isolates shared about 98% of their residues whereas the Australian isolate was more divergent showing an overall identity of 93% with other isolates.

All of the studies of PLRV isolate variation, whether based on whole genome sequences, ORF0 sequences or CP gene sequences reveal that the Australian isolates showed significant divergence from all other isolates of PLRV analysed to date.

To assess the variation of PLRV in South Africa the CP genes of 39 PLRV isolates from all potato growing areas in South Africa were amplified by RT-PCR, cloned and sequenced. Nucleotide sequences obtained of regions throughout South Africa were subsequently aligned with PLRV CP nucleotide sequences of overseas isolates and analyzed phylogenetically. Alpha-helix and beta-sheet structures of translated amino acid sequences of one overseas and one South African isolate were predicted and compared. The effect of altered amino acid sequence on the antigenicity of their respective proteins was assessed.

Materials and methods

PLRV infected leaf sources

Potato leaves infected with PLRV were collected from various potato growing regions in South Africa by officials of Potatoes South Africa, the organization representing South African commercial potato farmers, these are listed in Table 1.

Sample preparation

Leaf samples were stored at – 80°C until used. Infected plant material (0.1 g) was ground with a mortar and pestle in 2 ml grinding buffer [15 mM Na_2CO_3; 35 mM $NaHCO_3$; 2% PVP40; 0.2% bovine serum albumin (BSA); 0.05% Tween 20; 1% (w/v) sodium meta bisulphide; pH 9.6] to a homogenous solution. Homogenized solution (4 µl) was added to 25 µl GES-buffer (0.1 M glycine-NaOH, pH 9.0; 50 mM NaCl; 1 mM EDTA, pH 8.0; 5% (v/v) Triton X-100) and boiled for 10 min at 95°C. The solution was placed on ice for 5 min before RT-PCR.

Table 1. CP nucleotide sequences of South African PLRV isolates established in this study. The sequences from a single leaf sample were designated by the name of the sample (e.g. Sand1) followed by a hyphen to indicate the sequence number (e.g. Sand1-1). The following abbreviations are used for isolates: WC = Western Cape, FS = Free State, EC = Eastern Cape. These sequences are yet to be submitted to GenBank.

Sample name and sequence number	Area of origin
CC5-1	Wittenberg Vallei, WC, SA
CC5-2	Wittenberg Vallei, WC, SA
CC5-3	Wittenberg Vallei, WC, SA
CC6-6; CC6-8	Koue Bokkeveld, Ceres, WC, SA
CC6-7	Koue Bokkeveld, Ceres, WC, SA
DD020-1	Modderrivier, Douglas, NC, SA
DD020-2	Modderrivier, Douglas, NC, SA
DD020-3	Modderrivier, Douglas, NC, SA
NN3-1; NN3-3	KwaZulu-Natal, SA
NN3-2	KwaZulu-Natal, SA
NN334-6; NN334-8	KwaZulu-Natal, SA
NN334-7	KwaZulu-Natal, SA
PP0-6; PP0-7	Bethlehem, OFS, SA
PP003-11	Reitz, FS, SA
PP003-13	Reitz, FS, SA
SandA-1	Sandveld, WC, SA
SandA-3	Sandveld, WC, SA
SandB-1	Sandveld, WC, SA
SandB-2	Sandveld, WC, SA
SandB-3	Sandveld, WC, SA
SandF-1	Sandveld, WC, SA
SandF-2	Sandveld, WC, SA
SandF-3	Sandveld, WC, SA
SandH-7	Sandveld, WC, SA
SandH-10	Sandveld, WC, SA
Sand1-1; Sand1-2	Sandveld, WC, SA
Sand2-1	Sandveld, WC, SA
Sand2-2	Sandveld, WC, SA
Sand2-3	Sandveld, WC, SA
TT-11	Bethal, Mpumalanga, SA
TT-12	Bethal, Mpumalanga, SA
TT026-14; TT026-15; TT026-16	Middelburg, Mpumalanga, SA
WW154-19; WW154-21	Christiana, FS, SA
WW154-20	Christiana, FS, SA
Z25-14	Cradock, EC, SA
Z25-15	Cradock, EC, SA
Z26-23	Cradock, EC, SA
Z26-24	Cradock, EC, SA
Z26-25	Cradock, EC, SA

RT-PCR amplification of the PLRV coat protein gene

The cDNA of the full PLRV CP gene of each leaf sample was amplified using primers that were designed to bind to the 20 nucleotide bases on the 5' and 3' ends of the CP genes based on PLRV CP nucleotide sequences deposited in the GenBank™ data base (http://www.ncbi.nlm.nih.gov/Genbank) (see Table 2). Primers were designed using the computer software package, Primer Designer Version 2.0. Suitable additional bases were added to enable subcloning with either *Nco*1 and *Bam*H1, *Nco*1 and *Nde*1, *Nco*1 and *Xho*1 or

*Eco*R1 and *Xho*1. The sequence of the forward primer was either 5´-GCACGCCATGGGTACGGTCGTGGTTAAAGG-3´ (*Nco*1-cut site underlined) or 5´-GAATTCAGATGGGTACGGTCGTGGTTAAAGG-3´ (*Eco*R1-cut site underlined) and that of the reverse primer either 5´-GCGGATCCCTATTTGGGGTTCTGCAAAGC-3´ (*Bam*H1-cut site underlined), 5´-CTGCTGCCATATGCTATTTGGGGTTCTGCAAAGC-3´ (*Nde*1-cut site underlined) or 5´-CTCGAGCTATTTGGGGTTCTGCAAAGC-3´ (*Xho*1-cut site underlined). All primers were synthesized by the DNA Synthesis Laboratory, Department of Molecular & Cell Biology, University of Cape Town, South Africa.

Sample solutions (2 µl) were subjected to first strand cDNA synthesis in a standard 25 µl reaction mixture consisting of 2.5 µl 10x PCR Buffer (Southern Cross Biotechnology), 5 mM DTT, 1.5 mM $MgCl_2$, 0.5 µM forward primer, 0.5 µM reverse primer, 200 µM dNTPs, 0.25 µl Super-Therm Polymerase (5 U/µl, Southern Cross Biotechnology) and 0.125 µl M-MLV SuperScript[TM] III Reverse Transcriptase (200 U/µl, Invitrogen).

Subsequently amplifications were performed using the Hybaid Px2 Thermal Cycler programmed as follows: reverse transcription at 48°C for 30 min, followed by 35 cycles of PCR amplification at 94°C for 30 sec, 64°C for 45 sec, and 72°C for 1 minute, final extension was done at 72°C for 7 min. Samples were kept at 15°C until use.

RT-PCR product confirmation and purification

RT-PCR product formation was confirmed by agarose gel electrophoresis. The RT-PCR products were separated on an agarose gel (1% w/v), in 1x TAE electrophoresis buffer (0.48% w/v Tris-base; 0.11% glacial acetic acid; 0.5 M EDTA, pH 8.0, Maniatis *et al.*, 1982). Electrophoresis was performed at 110 V for 45 min. Gels were soaked in deionized water containing 1 µg/ml ethidium bromide for 20 min and stained DNA was visualized using an UV transilluminator. PCR products were excised from the gel and the DNA purified using a Wizard® SV Gel and PCR Clean-Up System kit (Promega) according to manufacturer's instructions.

Cloning of RT-PCR products

The purified RT-PCR products were cloned into pGEM-T Easy vectors using a pGEM-T Easy Vector Systems kit according to manufacturer's instructions (Promega). The RT-PCR products were ligated into the pGEM-T Easy vector in a 1:1 molar ratio as follows; 5 µl 2x rapid ligation buffer, 1 µl pGEM-T Easy Vector (50 ng), 3 µl RT-PCR product (1:1 molar ratio), 1 µl T4 DNA Ligase for a total volume of 10 µl. For each ligation reaction Luria-Bertani (LB) (10 g Bacto-tryptone, 5 g Bacto-yeast extract, 5 g NaCl, 1 L deinonized water, pH 7.0) plates (15 g agar/L of LB medium) with 100 µg/ml ampicillin, 160 µl IPTG (0.1 M) / 100 mL LB medium and 80 µl X-Gal (50 mg/mL, 100 mg 5-bromo-4-chloro-3-indolyl-B-D-galactoside, 2 ml N,N'-dimethyl-formamide, 2 ml dH_2O) / 100 mL LB medium were prepared. The plasmids were transformed into *Escherichia coli* strain JM109 (Promega) by mixing 2 µl of each ligation reaction with high

efficiency competent cells and placed on ice for 20 min, heat shocked for 50 sec at 42°C and returned to ice for 2 min. LB medium (950 µl) was added to each tube, which was incubated at 37°C for 1.5 h with shaking at 200 rpm. The content of each ligation reaction was transferred to an Eppendorf tube and centrifuged at 10 000 x g for 2 min, and 700 µl of the cleared LB medium removed. The pellet was resuspended in the remaining LB medium and plated out in duplicate onto the LB/ampicillin/IPTG/X-Gal plates and incubated overnight at 37°C. A successful transformant was identified as a white colony.

White colonies were subjected to direct PCR using T7 and SP6 promoter primers flanking the insert, and a toothpick scrape of the colonies as template DNA. Colony scrapings were subjected to PCR in a standard 10 µl reaction mixture consisting of 1 µl 10x PCR Buffer (Southern Cross Biotechnology), 1.5 mM MgCl$_2$, 0.1 µl Super-Therm DNA polymerase (5 U/µl, Southern Cross Biotechnology), 20 pmol/µl forward T7 primer, 20 pmol/µl reverse SP6 primer, 200 mM dNTPs and deionized water to a final volume of 10 µl. Subsequently amplifications were performed using the Hybaid Px2 Thermal Cycler programmed as follows: 1 cycle of PCR amplification at 94°C for 5 min followed by 25 cycles of 94°C for 30 sec, 55°C for 30 sec and 72°C for 30 sec, final extension was done at 72°C for 7 min. Samples were kept at 15°C until used. The products were electrophoresed on a 1% agarose gel as previously described to identify which colonies contained insert DNA. For each leaf sample, three colonies which contained insert DNA were purified using a Plasmix Miniprep kit (Talent) according to the manufacturer's instructions.

Sequencing PCR

Inserts of three cloned plasmids of each of the South African PLRV isolates were sequenced by cycle sequencing using the T7 and SP6 primers. The T7 and SP6 primers were synthesized by the DNA Synthesis Laboratory, Department of Molecular & Cell Biology, University of Cape Town, South Africa. Sequencing reactions consisted of 4 µl Terminator mix (BigDye® Terminator v3.1 Cycle Sequencing Kit, Applied Biosystems), 3 µl T7 or SP6 pGEM Primer (3.3 pmol/µl) and approximately 300 ng plasmid DNA in a final volume of 10 µl for each sequencing reaction. PCR amplifications for sequencing was performed using the Hybaid Px2 Thermal Cycler programmed as follows; 35 cycles of 96°C for 10 sec, 52°C for 30 sec and 60°C for 4 min followed by 1 cycle of 60°C for 10 min. Samples were stored at 4°C until analyzed. The PCR products were analyzed with an ABI PRISM® 373 DNA sequencer by the University of Stellenbosch Core DNA Cycle Sequencing Facility.

Nucleotide sequence analysis

Electropherograms were edited in Chromas v1.45 (Technelysium Pty., Tewantin, Australia). The PLRV CP nucleotide sequences of South African PLRV isolates were aligned with other PLRV CP nucleotide sequences and an outgroup sequence, the CP of *Sweet potato leaf speckling*

virus (SPLSV), obtained from the GenBank^TM data base (Table 2) using the computer program Bioedit v7.0.1 (Hall, 1999) and the alignment function of Clustal X v1.8 (Thompson *et al.*, 1997). Further alignment refinements were done by eye.

Phylogenetic analysis

Phylogenetic analysis was performed using the parsimony option in PAUP 4.0b10 (Swofford, 2002) and all substitutions were weighted equally. All missing or gap characters were treated as missing data. A heuristic search with 1000 replicates was used to find shortest trees using TBR branch swapping, holding ten trees with MulTrees on. Clade support was calculated with 1000 bootstrap replicates using TBR branch swapping and Multrees off. Bootstrap values above 75% were considered as well supported, and bootstrap values between 60 and 75% as weakly supported. No deductions were made where clade support was below 60%. Sequence divergence within phylogenetic clades was established using PAUP.

Amino acid sequence analysis and three dimensional protein structure prediction

The influence of the PLRV CP nucleotide sequence variation on subsequent amino acid sequences was investigated by deducing the amino acid sequences using Bioedit. To view the relevance of certain amino acid substitutions on the three-dimensional structure of CPs, graphs of the alpha-helix and beta-sheet content of one South African (Sand2-2) and one European (Neth31) PLRV CP deduced amino acid sequence were drawn using the computer program Prograph, Windows '95 version (Hofman, 1990). Antigenicity plots of these two sequences were drawn using the same computer program Prograph, Windows '95 version (Hofman, 1990).

Table 2. CP nucleotide sequences of PLRV isolates from GenBank. Code names for some of the GenBank sequences are according to Guyader and Ducray (2002). SPLSV was used as the outgroup.

Sample name	Area of origin	GenBank Accession number/ Reference
SPLSV	Peru	DQ655700
Rood	Roodeplaat, Gauteng, SA	AF022782
NALR7	North America	M89926
Aus	Australia	D13953
NethV	The Netherlands	X77321
Neth31	The Netherlands	X77326
NethV4	The Netherlands	X77325
Neth11	The Netherlands	X77322
Neth30	The Netherlands	X77324
Pak	Pakistan	AY307123
OP	Spain	AF453389
SKorRB	South Korea	U74377
NethWa	The Netherlands	Y07496
Cuba	Cuba	S77421
India	India	AF539791
Italy	Italy	Faccioli *et al.* (1995)
14.2	France	AF453394
FrPT	France	AY007727
Fr1	France	AF453391
Zim13	Zimbabwe	AF453388
Egp	Egypt	AY138970
Pol	Poland	X74789
SKor777	South Korea	U73777
Can	Canada	D13954
SKor	South Korea	AF296280
Neth15	The Netherlands	X77323
Sco	Scotland	D00530
CIP01	Peru	AF453392
Ger	Germany	X13906
Noir	France	AF453390
Chi	China	AY079210
Cu87	Cuba	AF453393
FrCU87	France	AF271215

Results and discussion

RT-PCR amplification of the PLRV coat protein gene

The CP gene of PLRV contained in the plant samples was readily amplified (Fig. 1) following the sample preparation technique described in the methods section and by using the optimized RT-PCR to give a 650 bp product.

Cloning and sequencing of RT-PCR products

Successful cloning could be confirmed using the confirmatory PCR to identify the colonies of positive clones (results not shown). Sequence analysis of all of the cloned RT-PCR products confirmed that the RT-PCR product was the PLRV CP gene, and these sequences are shown in addendum A of this thesis.
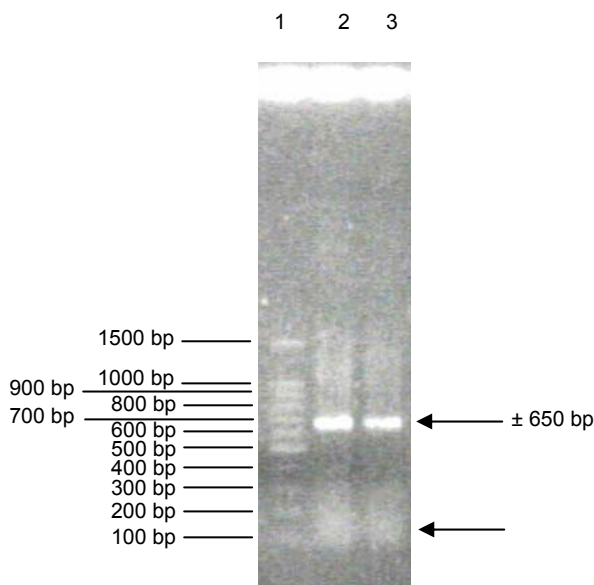
Fig. 1. RT-PCR products of the PLRV CP. Lane 1: 100 bp DNA ladder (Promega); lane 2-3: RT-PCR products of the PLRV CP (± 650 bp), isolate CC6. Bottom arrow indicates primer dimers as a result of the RT-PCR.

Nucleotide sequence and phylogenetic analysis

The nucleotide sequences were aligned (see addendum A of this thesis), but a few gaps had to be introduced into the outgroup sequence. The length of the aligned matrix was 631.

In the parsimony analysis, the last four characters were excluded from the analysis as they were only present in the outgroup. The analysis revealed that the number of parsimony informative characters was 67 and the number of parsimony uninformative characters was 152. The tree statistics of the parsimony analysis gave a consistency index value (CI) of 0.798 and a retention index value (RI) of 0.914. The shortest tree length was 317, and the number of trees that were retained was 271. One of the shortest trees retrieved is shown in Fig. 2. The trees generated by this analysis were used to generate a strict consensus tree that is shown in Fig. 3. The bootstrap values generated by the bootstrap analysis were plotted onto the strict consensus tree (Fig. 3).

The parsimony analysis retrieved two major clades of which one was strongly supported. One clade contained all of the overseas isolates with the exception of one Australian and one North American isolate. Five South African isolates were also retrieved in this clade. Branch lengths in this clade were short in the majority of cases. The second clade contained the remaining South African isolates and one Australian and one North American isolate. In the second clade, branch lengths were longer in most cases.
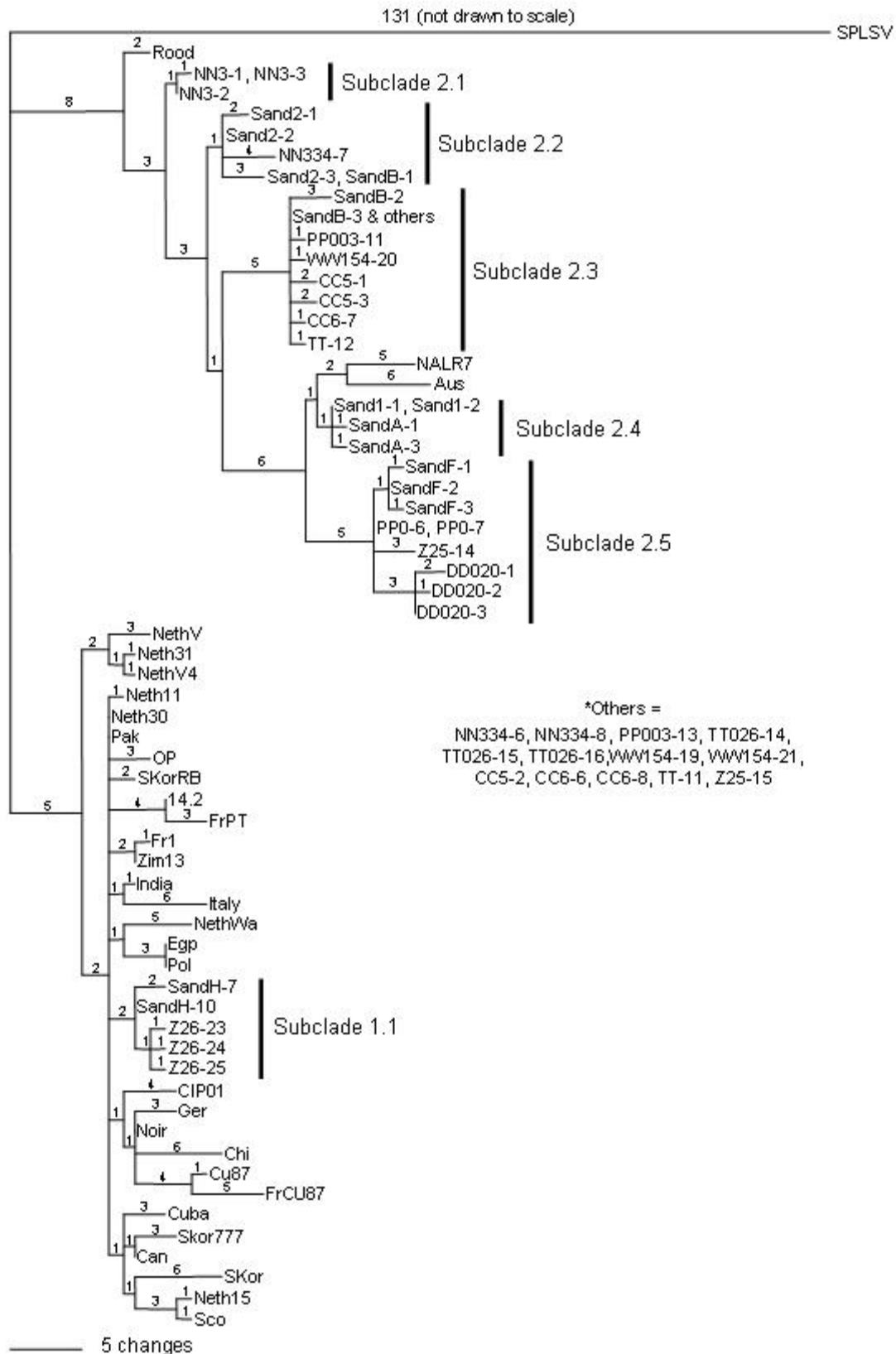
Fig. 2. One of the shortest trees of the parsimony analysis of the PLRV CP nucleotide sequence data. Branch lengths are shown above branches.
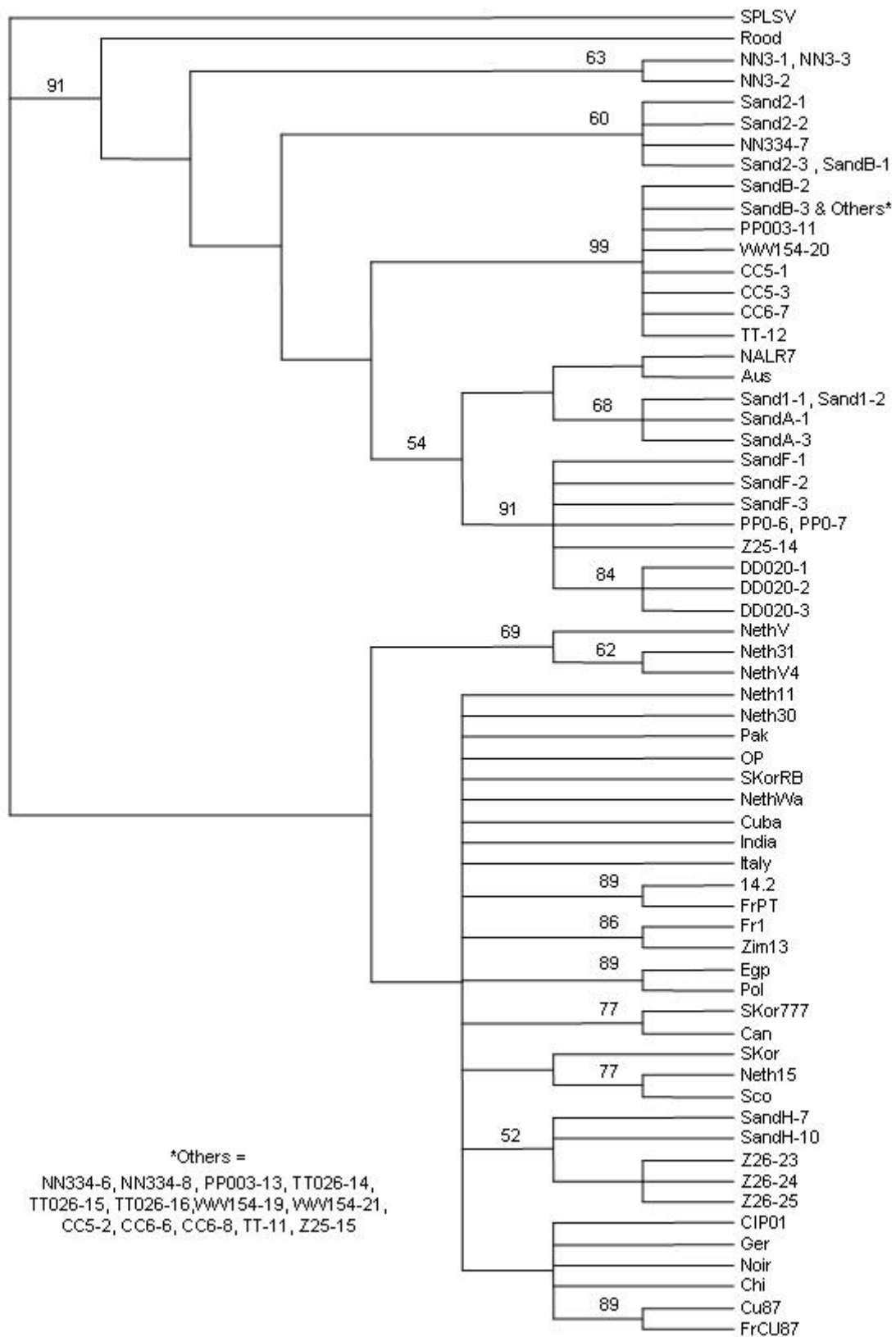
Fig. 3. The strict consensus tree of the parsimony analysis of the PLRV CP nucleotide sequence data. Bootstrap values are shown above branches.

Sequence divergence was assessed over all PLRV nucleotide sequences and within both clades. The maximum sequence divergence over all PLRV nucleotide sequences was 5.344%. Within clade 1, the highest sequence divergence was 3.030% whilst in clade 2, the highest sequence divergence was 3.668%. Although divergence within the two clades as calculated from the sequence divergence does not appear to be very different, divergence in clade 1 between most isolates is much lower than between isolates in clade 2. This is evident from the single tree presented in Fig. 3 in which it can be seen that branch lengths in clade 1 are much shorter in the majority of cases than branch lengths in clade 2.

Within clade 1, the South African isolates group in a single subclade (shown as subclade 1.1 in Fig. 3). A number of other strongly supported subclades are also retrieved in clade 1, but cannot be grouped geographically, i.e. some members of a subclade come from areas that are separated by vast distances. Clade 2 can be further subdivided into five subclades, three of which are weakly supported and two of which is strongly supported, with the Roodeplaat isolate appearing in a basal position.

Members of subclade 2.1 belong to South African isolates collected from KwaZulu-Natal. Subclade 2.2 contains South African isolates that have been collected in the Sandveld region and KwaZulu-Natal. Subclade 2.3 contains South African isolates only that have been collected in the Sandveld region, KwaZulu-Natal, Free State, Mpumalanga, Western Cape and the Eastern Cape. Subclade 2.4 contains South African isolates that were collected in the Sandveld region sister to isolates from the GenBank database that have been collected in North America and Australia. Subclade 2.5 contains isolates from the Sandveld region, Free State, Eastern Cape and Northern Cape.

In two instances, isolates from different areas have identical CP nucleotide sequences. Examples are: Sand2-3 and SandB-1 (subclade 2.2); SandB-3, NN334-6, PP003-13, TT026-14, WW154-19, CC5-2, CC6-6, TT11 and Z25-15 (subclade 2.3). Therefore, many leaf samples were found to be infected with the same viral isolate.

In most instances in clade 2, isolates from the same leaf sample fall in the same subclade. Examples of this include NN3-1 and NN3-2 (subclade 2.1); Sand2-1, Sand2-2 and Sand2-3 (subclade 2.2); SandB-2 and SandB-3 (subclade 2.3); PP003-11 and PP003-13 (subclade 2.3); WW154-19 and WW154-20 (subclade 2.3); CC5-1, CC5-2 and CC5-3 (subclade 2.3); CC6-6 and CC6-7 (subclade 2.3); TT-11 and TT-12 (subclade 2.3); SandA-1 and SandA-3 (subclade 2.4); SandF-1, SandF-2 and SandF-3 (subclade 2.5) as well as DD020-1, DD020-2 and DD020-3 (subclade 2.5). Therefore, each of the leaf samples from which these isolates were isolated contained two or three very similar isolates.

In other instances isolates from the same leaf sample fall into different subclades. There are isolates from the same leaf sample in clade 2 that fall into different subclades. Examples of this include: NN334-7 (subclade 2.2) and NN334-6 (subclade 2.3); SandB-1 (subclade 2.2) and SandB-2, SandB-3 (subclade 2.3) as well as Z25-15 (subclade 2.3) and Z25-14 (subclade 2.5). Therefore, these leaf samples contained two or three more diverse isolates.

Amino acid sequence analysis and three dimensional protein structure prediction

Translation of one South African (Sand2-2) and one European (Neth31) PLRV CP nucleotide sequence revealed that there was also variation in the amino acid sequences of these two PLRV CPs. These amino acid sequences show 97.1% homology. Three of the amino acid substitutions are conservative, found in amino acid positions 7 (Arg/Lys), 18 (Arg/Lys) and 80 (Met/Val) (see addendum B of this thesis). Non-conservative amino acid substitutions are found in amino acid positions 2 (Ser/Gly), 153 (Thr/Ile) and 192 (Thr/Pro) (see addendum B of this thesis).

From Fig. 4 it can be seen that the alpha-helix and beta-sheet content of the South African PLRV CP differs from that of the European PLRV CP. There is a reduced alpha-helixes and beta-sheet frequency in the South African PLRV CP in comparison with the European PLRV CP. The reduction in alpha-helixes and beta-sheets frequency can be attributed to the proline substitution in position 192.

The hydrophilicity plots (Fig. 5) of the South African and European isolates showed minor changes in the C-terminal region of the CP as a result of the proline substitution in position 192 whereas the N-terminal regions were found to be identical. This is the result of a proline substitution in the South African isolate, which falls within a potential antigenic region of the CP (hydrophilicity value greater than zero).
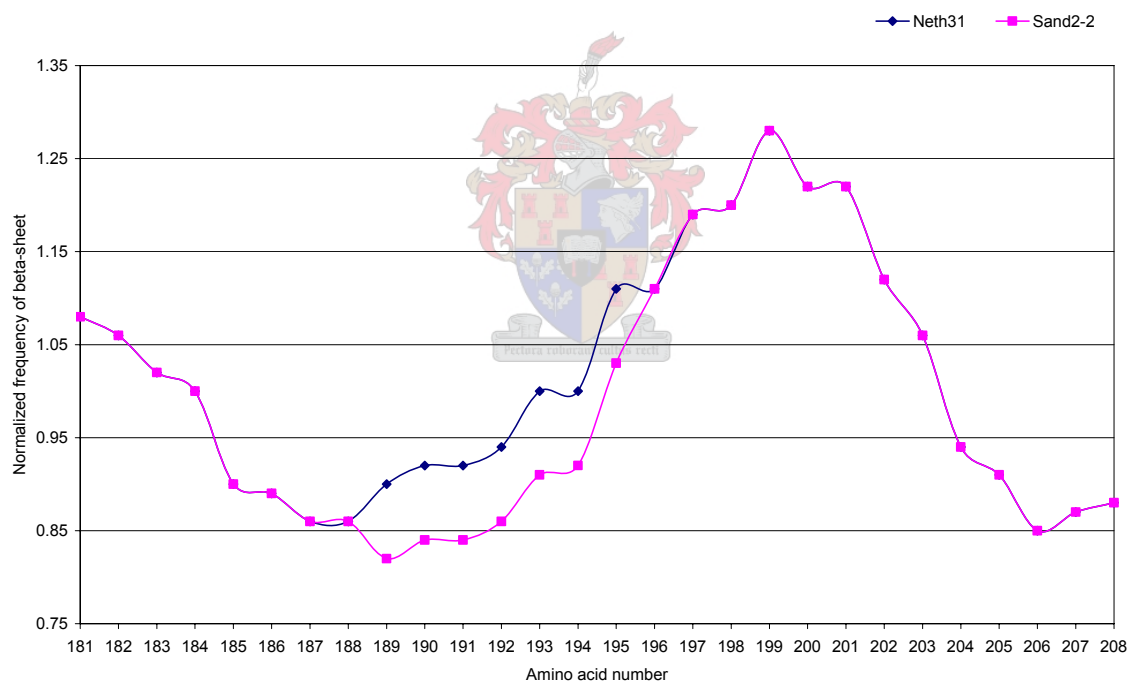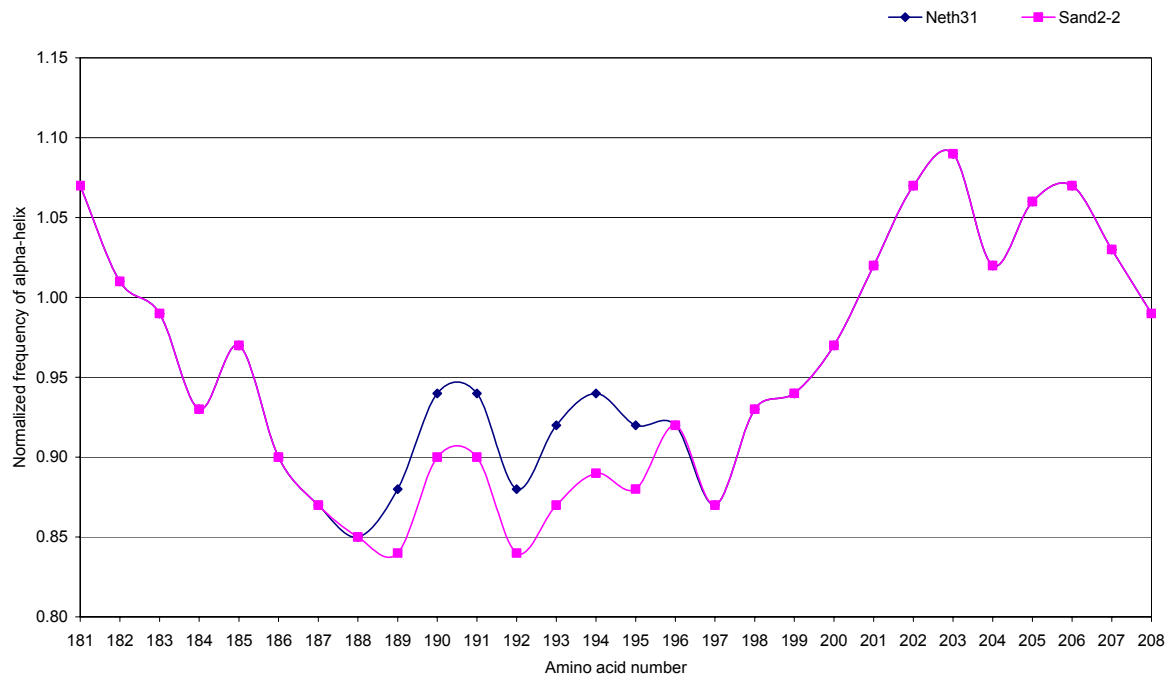
Fig. 4: (a). Alpha-helix content of amino acid residues 181-208 from Sand2-2 in comparison with Neth31. The alpha-helix curves were drawn for the amino acid region 181-208 from one South African PLRV CP (Sand2-2) to compare with the same amino acid region from one European PLRV CP (Neth31) **(b). Beta-sheet content of amino acid residues 181-208 from Sand2-2 in comparison with Neth31.** The beta-sheet curves were drawn for the amino acid region 181-208 from one South African PLRV CP (Sand2-2) to compare with the same amino acid region from one European PLRV CP (Neth31).
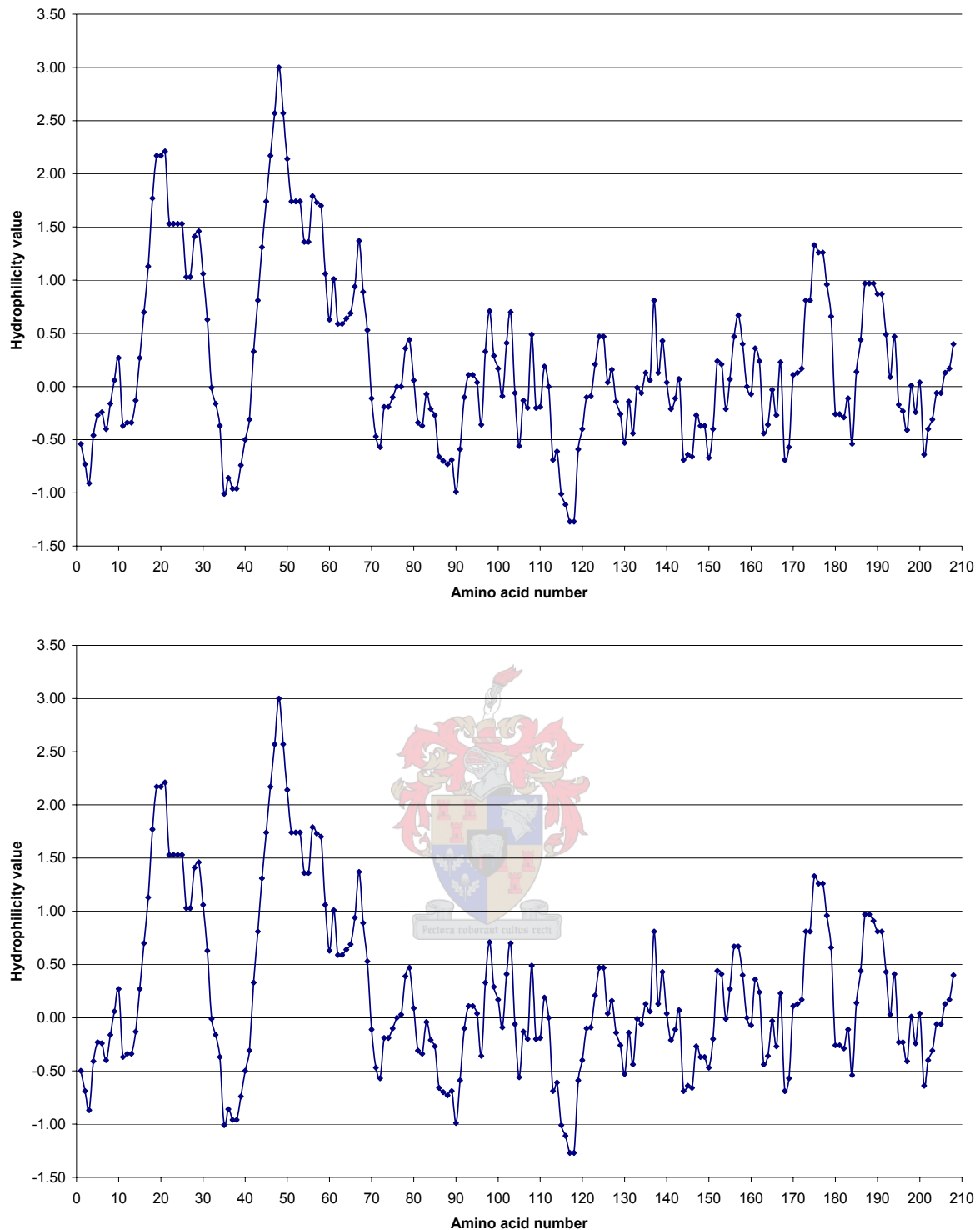
Fig. 5: (a). Hydrophilicity plot of a South African PLRV CP deduced amino acid sequence (Sand2-2). Amino acid 192 has a hydrophilicity value of 0.49. **(b). Hydrophilicity plot of a European PLRV CP deduced amino acid sequence (Neth31).** Amino acid 192 has a hydrophilicity value of 0.43. The complete deduced amino acid sequence of the South African or European PLRV CP was used to generate the graph using the computer program Prograph (Hofman, 1990). Hydrophilicity values of the deduced amino acids were used as an indication of their antigenicity, values above zero indicate potential antigenic regions in the protein.

Conclusions

In this study, it was found that there is significant sequence variation in the CP genes of South African PLRV isolates. In the interpretation of the variation found in South African PLRV CP nucleotide sequences it is, however, important to take into consideration that the reverse transcriptase and *Taq* polymerase used in the RT-PCR can also introduce sequence errors. Smaller variation between isolates may therefore be the result of such RT-PCR errors and not as a result of true PLRV CP mutations. However, the larger the sequence differences are, the less likely it is that they are the result of RT-PCR errors. Thus sequence variation between the subclades of clade 2 may therefore be ascribed to real differences between isolates. Furthermore, the fact that the same isolates occurred in different geographic regions in South Africa can largely be ascribed to the movement of tubers from one area to another. Larger variation between isolates in single leaf samples, which was found in a limited number of leaf samples, indicates multiple infections and also occurs in South African potato stocks. To our knowledge this has not been reported elsewhere in the world. This has the potential of increasing the chances for formation of recombinants, which poses as additional threat for new strain evolution. A careful inspection of the CP nucleotide sequences of South African isolates did not reveal any recombination of strains at present.

The phylogenetic analysis showed that a limited number of South African PLRV isolates grouped with isolates from Europe and other countries in one major phylogenetic clade. The rest of the South African isolates were found to group with isolates from North America and Australia, but many of these isolates occupied more intermediate positions between these major groupings in a second major clade. Keese *et al.* (1990), Faccioli *et al.* (1995), Haliloglu and Bostan (2002), Guyader and Ducray (2002), Mukherjee *et al.* (2003) and Guyader *et al.* (2004) all found that the Australian PLRV isolates had diverged significantly from PLRV isolates found elsewhere in the world. The intermediate position of the isolates identified in this study in South Africa and their diversity indicates that the isolates in the second clade may have evolved independently, and that the Australian isolates may have originated here in South Africa. In this regard, the Roodeplaat isolate is of interest, as it was sequenced in 1997, and therefore represents a viral isolate that was present in South Africa ten years ago. Its basal position indicates that present virus isolates may have diverged from it subsequently.

It appears that the South African PLRV isolates have diverged to a larger degree in comparison to those found elsewhere. Guyader and Ducray (2002) found limited variation in European isolates in ORF0 whereas Australian isolates were quite divergent. Furthermore, Guyader *et al.* (2004) concluded that the limited variation in European PLRV isolates indicated genetic stability and slow evolution of these isolates (many of which are included in clade 1 in this study). In contrast this analysis reveals that the isolates in clade 2 are evolving at a faster rate and have diversified to a greater extent than the isolates in clade 1. Guyader *et al.* (2004) concluded that

the faster PLRV mutation rate in isolates in Australia may be the result of unique environmental selection pressures. However, it may also be due to the inherent high mutation rate of viruses. A common property of plant RNA viruses is their high mutation rate due to their error-prone RNA polymerase which lacks proofreading functions resulting in between 10^{-3} and 10^{-5} misincorporations per nucleotide copied (Power, 2000; Guyader *et al.*, 2004). Or stated differently, because of the error-prone RNA polymerase, mutations are introduced at a rate close to one substitution per genome per replication cycle (Guyader and Ducray, 2002). Thus the South African and Australian isolates could show a faster rate of mutation due to more than one crop cycle per year (a common practice in South Africa) versus one crop cycle in Europe and other countries.

Additionally, the greater isolate variation in South Africa may be the result of more aphid hosts in South Africa in which PLRV can develop (Dr Ben Pieterse, Potatoes South Africa, personal communication). PLRV can also infect and multiply in different plant hosts in South Africa, for example in a variety of indigenous *Solanum* species as well as cultivated peppers and chillies. Therefore the vector-host selection pressure is potentially released in South Africa because of different carriers. The lack of selective intermediary (aphid) hosts together with different plant hosts may therefore result in unique South African mutations.

South Africa has been isolated in terms of the influx of potatoes as new imports of potatoes have been closely regulated and only limited mother stocks were imported in the past 20 years. The South African viral isolates in clade 1 may possibly have entered the country in this way through recent imports. However, the variation in clade 2 as well as the intermediate position of South African isolates supports that these isolates evolved through mutation in South Africa, and that these isolates were not recently imported. The isolation of South Africa as a reason for unique divergence of PLRV is in line with the geographical isolation of Australia that might have influenced the divergence of Australian PLRV isolates from the rest of the PLRV isolates from other countries (Mukherjee *et al.*, 2003, Guyader *et al.*, 2004).

Variation occurred not only in the nucleotide sequences of the various PLRV isolates but also in the deduced amino acid sequences. Non-conservative as well as conservative substitutions were found in the South African CPs. The non-conservative proline substitution in amino acid position 192 of Sand2-2 CP, shared by most of the South African CPs, was found to influence its alpha-helix and beta-sheet content. According to a prediction by Mayo and Ziegler-Graff (1996), amino acid 192 falls in a potential beta-sheet structure right next to a region probably exposed on the surface of the virus particle and five amino acids from a region thought to be, or to contribute, to epitopes. Since proline is known to disrupt beta-sheets its substitution in South African PLRV CPs could have a significant influence on the three-dimensional structure of this epitope. This proline substitution is also found in PLRV CP amino acid sequences found in

South Korea (Skor777), Canada (Can), North America (NALR7) and Australia (Aus). However, the influence that this may have on the antigenicity of the protein as a whole is not known, as this region is closer to the C-terminal region of the CP, whilst it is known that the N-terminal region is the most important antigenic region of the CP (Prof M.H.V. van Regenmortel, Strasbourg, personal communication). The outcome of continued unique mutations in the South African PLRV CP nucleotide sequences and subsequently the amino acid sequences would be an increasing lack of detection of South African PLRV isolates by European ELISA kits. The results therefore indicate a need to develop highly specific ELISA kits aimed at detecting South African isolates of PLRV in potato leaves and tubers.

The identification of diversified and different PLRV isolates in South Africa indicates there is an ongoing threat that PLRV isolates with altered and also more severe pathogenicity may evolve. Thus, the sequencing of CP genes of South African PLRV isolates on an ongoing basis is important to monitor the evolution of PLRV in this country. Additionally, the introduction of new PLRV isolates via newly imported potato material into South Africa, holds a threat to the South African potato industry as new PLRV introductions could form new and more pathogenic recombinants with the already unique PLRV isolates present in South Africa.

References

Bahner, I., Lamb, J., Mayo, M.A. and Hay, R.T. (1990). Expression of the genome of potato leafroll virus: readthrough of the coat protein termination codon *in vivo*. *J. Gen. Virol.*, **71**: 2251-2256.

De Souza-Dias, J.A.C., Russo, P., Betti, J.A., Miller, L. and Slack, S.A. (1999a). Simplified extraction method for ELISA and PCR detection of potato leafroll luteovirus primary infection in dormant potato tubers. *Am. J. Potato Res.*, **76**: 209-213.

De Souza-Dias, J.A.C., Russo, P., Miller, L. and Slack, S.A. (1999b). Comparison of nucleotide sequences from three potato leafroll virus (PLRV) isolates collected in Brazil. *Am. J. Potato Res.*, **76**: 17-24.

Faccioli, G., Rosner, A. and Forni, M. (1995). Use of the polymerase chain reaction to clone the potato leafroll virus coat protein gene directly from the total RNA of infected plants. *Potato Res.*, **38**: 211-218.

Guyader, S. and Ducray, D. G. (2002). Sequence analysis of *potato leafroll virus* isolates reveals genetic stability, major evolutionary events and differential selection pressure between overlapping reading frame products. *J. Gen. Virol.*, **83**: 1799-1807.

Guyader, S., Tanguy, S., Andrieu, A., Pasquer, F. and Ducray, D. G. (2004). Is *potato leafroll virus* (PLRV) a slowly evolving RNA virus? *12th EAPR Virology Section Meeting Rennes, France, 2004: abstracts of oral presentations and poster presentations*. Available: http://www.rennes.inra.fr/eapr2004/abstracts.htm. (Accessed on 24 February 2005).

Haliloglu, K. and Bostan, H. (2002). Nucleotide sequence analysis for assessment of variability of potato leafroll virus and phylogenetic comparisons. *OnLine Journal of Biological Sciences*, **2**(9): 582-586.

Hall, T.A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl. Acids Symp. Ser.,* **41**: 95-98.

Haupt, S., Stroganova, T., Ryabov, E., Kim, S.H., Fraser, G., Duncan, G., Mayo, M.A., Barker, H. and Taliansky, M. (2005). Nucleolar localization of potato leafroll virus capsid proteins. *J. Gen. Virol.*, **86**: 2891-2896.

Hofman, K. (1990) Prograph, Windows '95 version. Institut Fuer Biochemie, Universitaet Koeln, Germany.

Kawchuk, L.M., Martin, R.R. and McPherson, J. (1990). Resistance in transgenic potato expressing the potato leafroll virus coat protein gene. *Mol. Plant-Microb. Interact.*, **3**(5): 301-307.

Keese, P., Martin, R.R., Kawchuk, L.M., Waterhouse, P.M. and Gerlach, W.L. (1990). Nucleotide sequences of an Australian and a Canadian isolate of potato leafroll luteovirus and their relationships with two European isolates. *J. Gen. Virol.*, **71**: 719-724.

Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982). *Molecular Cloning, A Laboratory Manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, pp. 156.

Martin, R.R., Keese, P.K., Young, M.J., Waterhouse, P.M. and Gerlach, W.L. (1990). Evolution and molecular biology of luteoviruses. *Annu. Rev. Phytopathol.*, **28**: 341-363.

Mayo, M.A. and Ziegler-Graff, V. (1996). Molecular biology of Luteoviruses. *Adv. Virus Res.*, **46**: 413-460.

Mehrad, M., Lapierre, H. and Yot, P. (1979). RNA in pototo leafroll virus. *FEBS Letters*, **101**(1): 169-174.

Mukherjee, K., Verma, Y., Chakrabarti, S.K., Singh, M.N. and Khurana, S.M.P. (2003). Cloning and sequencing of coat protein gene of an Indian potato leaf roll virus (PLRV) isolate and its similarity with other members of luteoviridae. *Virus Genes*, **26**(3): 247-253.

Swofford, D.L. (2002) Phylogenetic Analysis Using Parsimony PAUP, version 4.0b1. Washington, DC: Smithsonian Institution.

Taliansky, M., Mayo, M.A. and Barker, H. (2003). *Potato leafroll virus*: a classic pathogen shows some new tricks. *Mol. Plant Pathol.*, **4**(2): 81-89.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., Higgins, D.G. (1997). The ClustalX window interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucl. Acids. Res.*, **24**: 4876-4882.

Van der Wilk, F., Huisman, M.J., Cornelissen, B.J.C., Huttinga, H. and Goldbach, R. (1989). Nucleotide sequence and organization of potato leafroll virus genomic RNA. *FEB*, **245**(1, 2): 51-56.

Van der Wilk, F., Verbeek, M., Dullemans, A. M. and Van den Heuvel, J.F.J.M. (1997). The genome-linked protein of potato leafroll virus is located downstream of the putative protease domain of the ORF1 product. *Virology*, **234**: 300-303.

Van Regenmortel, M.H.V., Fauquet, C.M., Bishop, D.H.L., Carstens, E.B., Estes, M.K., Lemon, S.M., Maniloff, J., Mayo, M.A., McGeoch, D.J., Pringle, C.R. and Wickner, R.B. (2000). Virus Taxonomy Online. Available: http://www.virustaxonomyonline.com (Accessed on 18 February 2006).

Chapter 4:

The production of recombinant potato leafroll virus coat protein in *Escherichia coli*

Adri H. Rothmann, Mark Matzopoulos, Johan T. Burger, Dirk U. Bellstedt

Department of Biochemistry, University of Stellenbosch, Private Bag X1, Stellenbosch 7602, South Africa.

Summary. The coat protein (CP) gene of a South African isolate of potato leafroll virus (PLRV) was amplified by reverse transcription–polymerase chain reaction (RT-PCR). The amplicon was cloned in pGEM-T Easy, sequenced and subcloned into a bacterial expression vector (pET14-b) with various restriction enzyme combinations. Expression of full length recombinant PLRV CP was attempted in *Escherichia coli* strains BL21(DE3)pLysS, Rosetta-gami B(DE3)pLysS and Rosetta-2(DE3)pLysS. As this was not successful, the PLRV CP gene was subcloned in another expression vector (pGEX) for expression as an N-terminal fusion protein with glutathione-*S*-transferase (GST) in *E. coli* strains BL21(DE3)pLysS and Rosetta-2(DE3)pLysS. The recombinant GST-PLRV CP fusion protein was purified with affinity chromatography yielding recombinant GST-PLRV CP fusion protein and used for antibody production in rabbits. Cleavage of the recombinant GST-PLRV CP fusion protein was successful, but the cleaved PLRV CP could not be purified and hence used for immunization. The effectiveness of antibodies produced to recombinant GST-PLRV CP fusion protein for PLRV CP recognition was assessed in western blots with these antibodies as well as commercial antibodies against PLRV CP and GST. It was found that antibodies to the recombinant GST-PLRV CP fusion protein were more effective for the detection of GST than PLRV CP and that production of antibodies to the cleaved PLRV CP product would be necessary if antibodies with a higher titre against native PLRV CP are required.

Introduction

PLRV can be purified from infected potato (*Solanum tuberosum* L.) with yields of 0.4-0.6 mg/kg foliage (Rohwani and Stace-Smith, 1979) although it is technically difficult. This low yield is due to the fact that the virus is phloem limited. For antibody production a relatively large amount of the virus is required for immunization (approximately 3 mg – 4 mg, Tamada and Harrison, 1980). Consequently, these procedures are time-consuming and expensive. Furthermore, during virus purification, plant proteins are often co-isolated and when this material is used for immunization, antibodies against contaminating plant proteins are also elicited. This results in antisera that have the potential of also detecting plant proteins in potato material over and above PLRV. Such antisera can give false positive results in ELISA which would then result in inconsistent and unreliable results.

For the production of antibodies to viruses, recombinant CPs formed in bacteria has been used by various workers in the field. These recombinant CPs were either expressed on their own (Čeřovská *et al.*, 2002), as fusion proteins (Kaden-Kreuziger *et al.*, 1995; Nikolaeva *et al.*, 1995; Vaira *et al.*, 1996; Ling *et al.*, 2000; Meng *et al.*, 2003) or with 6 histidine-tag attached (Kaden-Kreuziger *et al.*, 1995; Jelkmann and Keim-Konrad, 1997; Minafra *et al.*, 2000; Kumari *et al.*, 2001; Hourani and Abou-Jawdah, 2003; Abou-Jawdah *et al.*, 2004). In most instances where a fusion protein or 6-histidine tag was used, it was attached at the N-terminal region of the recombinant protein (Kaden-Kreuziger *et al.*, 1995; Nikolaeva *et al.*, 1995; Vaira *et al.*, 1996; Minafra *et al.*, 2000; Kumari *et al.*, 2001; Hourani and Abou-Jawdah, 2003; Meng *et al.*, 2003; Abou-Jawdah *et al.*, 2004), but instances where the fusion protein or 6-hisitidine tag was attached at the C-terminal of the recombinant protein were also used (Jelkmann and Keim-Konrad, 1997; Ling *et al.*, 2000). Recombinant viral CP expression for antibody production is summarized in Table 1. Not all of the antibodies produced against these recombinant viral CPs could be used successfully for various applications as listed in Table 1.

Table 1 a, b, c. Literature summary of various approaches for the production of antibodies to recombinant viral CPs. All of the recombinant viral CPs were used for the production of antibodies and applied in either DAS-ELISA, indirect-ELISA and/or western blot. Some of the antibodies formed to recombinant viral CPs were effective for the detection of native viruses in the above mentioned applications, others were not.

Table 1 a. The production of antibodies to free recombinant viral CPs.

Free recombinant viral CPs			
Virus name	**Expected size of expressed protein**	**Application**	**Reference**
Potato virus A (PVA)	32 kDa	Double-antibody sandwich (DAS) enzyme-linked immunosorbent assay (ELISA) (coating with antibody): not effective for detection of PVA. Indirect-ELISA (coating with antigen): effective for detection of PVA. Western-blot: effective for detection of PVA.	Čeřovská *et al.* (2002)

Table 1 b. The production of antibodies to 6 histidine-tagged recombinant viral CPs.

6 Histidine-tagged recombinant viral CPs			
Virus name	**Expected size of expressed protein**	**Application**	**Reference**
Strawberry mild yellow edge associated potexvirus (SMYEAV)	26 kDa	DAS-ELISA: effective for detection of SMYEAV.	Kaden-Kreuziger *et al.* (1995)
Prune dwarf virus (PDV)	25-30 kDa	DAS-ELISA: effective for detection of PDV. Indirect-ELISA: effective for detection of PDV. Western-blot: effective for detection of PDV.	Abou-Jawdah *et al.* (2004)
Cucurbit yellow stunting disorder virus (CYSDV)	28.5 kDa	DAS-ELISA: medium effectivity for detection of CYSDV. Indirect-ELISA: effective for detection of CYSDV.	Hourani and Abou-Jawdah (2003)
Grapevine rupestris stem pitting associated virus (GRSPaV)	28 kDa	DAS-ELISA: not effective for detection of GRSPaV. Indirect-ELISA: not effective for detection of GRSPaV. Western-blot: effective for detection of GRSPaV.	Minafra *et al.* (2000)
Apple stem pitting virus (ASPV)	44 kDa	DAS-ELISA: not effective for detection of ASPV. Indirect-ELISA: medium effectivity for detection of ASPV.	Jelkmann and Keim-Konrad (1997)
Faba bean nectrotic yellow virus (FBNYV)	45 kDa	DAS-ELISA: not effective for detection of FBNYV. Western-blot: effective for detection of FBNYV.	Kumari *et al.* (2001)

Table 1 c. The production of antibodies to fusion protein recombinant viral CPs.

Fusion protein recombinant viral CPs				
Virus name	**Fusion protein**	**Expected size of expressed protein**	**Application**	**Reference**
Tomato spotted wilt tospovirus (TSWV)	GST at N-terminal of CP.	55 kDa	DAS-ELISA: effective for detection of TSWV. Indirect-ELISA: effective for detection of TSWV. Western-blot: effective for detection of TSWV.	Vaira *et al.* (1996)
Grapevine leafroll associated closterovirus-3 (GLRaV-3)	β-galactosidase at C-terminal of CP.	43 kDa	DAS-ELISA: effective for detection of GLRaV-3. Western-blot: effective for detection of GLRaV-3.	Ling *et al.* (2000)
Citrus tristeza virus (CTV)	Fragment of maltose-binding protein (MBP) at N-terminal of CP.	67 kDa	Indirect-ELISA: effective for detection of CTV. Western-blot: effective for detection of CTV.	Nikolaeva *et al.* (1995)
Rupestris stem pitting associated virus (RSPaV)	MBP at N-terminal of CP.	71 kDa	DAS-ELISA: not effective for detection of RSPaV. Indirect-ELISA: effective for detection of RSPaV. Western-blot: effective for detection of RSPaV.	Meng *et al.* (2003)
SMYEAV	GST at N-terminal of CP.	53 kDa	DAS-ELISA: not effective for detection of SMYEAV.	Kaden-Kreuziger *et al.* (1995)

For the production of recombinant proteins in *E. coli*, the pET expression system (Novagen) is an efficient protein expression system that has been used successfully in numerous applications (Kelly *et al.*, 1995). The pET system uses the bacteriophage T7 promoter in hosts bearing the T7 RNA polymerase gene (λDE3 lysogen) to direct the expression of target genes (Anon, 1994). Hosts carrying the pLysS plasmid encode T7 lysozyme, a natural inhibitor of T7 RNA polymerase, and this reduces its ability to transcribe target genes in uninduced cells. Furthermore, the pLysS plasmid helps with the purification of proteins from the host as it produces T7 lysozyme that assists in the disruption of the cell wall of *E. coli* (Studier *et al.*, 1990). The *E. coli* BL21 strain is the standard expression host for pET constructs (Anon, 1994). If rare codons are present in the protein to be expressed, there are various other strains that can be considered for expression of the protein without altering the expression construct. These strains supplement the production of tRNAs for rare codons and include *E. coli* strains Rosetta-2 and Rosetta-gami B. Rosetta-2 host strains are BL21 derivatives designed to enhance the expression of eukaryotic proteins that contain codons rarely used in *E. coli* (Anon, 2006a). Rosetta-gami B strains combine the key features of BL21 and Rosetta strains to enhance both the expression of eukaryotic proteins and the formation of target protein disulphide bonds in the bacterial cytoplasm (Anon, 2006b). By supplying these rare tRNAs, the Rosetta strains enable the translation of proteins, which would otherwise be limited by the codon usage of *E. coli* (Anon, 2006a).

In the case of abundant rare codons close to the N-terminus of the protein to be expressed, a fusion could be added to the N-terminus. This can be done with the pET expression system or the GST gene fusion system. The GST gene fusion system makes use of N-terminal *Schistosoma japonicum* GST (26 kDa) fusions for expression of fusion proteins in *E. coli* (Anon, 2002) resulting in the protein of interest to be closer to the carboxyl terminus, which in turn results in less hindered translation and thereby improved expression (Chen and Inouye, 1990). GST fusion proteins can be purified from bacterial lysates by affinity chromatography using immobilized glutathione (Anon, 2002).

In this study, the PLRV CP gene from a South African isolate of PLRV was amplified and cloned into pGEM-T Easy after which it was subcloned into the expression vector pET14-b with various restriction enzyme combinations to test their efficiency. Expression of full length recombinant PLRV CP using this construct was attempted in *E. coli* strains BL21(DE3)pLysS, Rosetta-gami B(DE3)pLysS and Rosetta-2(DE3)pLysS. Subsequently, the PLRV CP gene was subcloned in the expression vector pGEX and used to express the PLRV CP as a fusion protein with GST in *E. coli* strains BL21(DE3)pLysS and Rosetta-2(DE3)pLysS. Purified recombinant GST-PLRV CP fusion protein was used for antibody production in rabbits. Commercial antibodies to PLRV CP and GST as well as those produced against recombinant GST-PLRV CP fusion protein were subsequently used in western blots analyses to assess the effectiveness of PLRV CP

recognition by the antibodies to the recombinant GST-PLRV CP fusion protein and their potential use in ELISA.

Materials and methods

Amplification and cloning of the PLRV CP gene into pGEM-T Easy

Primers for the amplification of the PLRV CP gene were designed using the computer software package, Primer Designer Version 2.0. Suitable additional bases were added to enable subcloning with *Nco*1 and *Nde*1. The sequence of the forward primer was 5´-GCACG<u>CCATGG</u>GTACGGTCGTGGTTAAAGG-3´ (*Nco*1-cut site underlined) and that of the reverse primer, 5´-CTGCTGC<u>CATATG</u>CTATTTGGGGTTCTGCAAAGC-3´ (*Nde*1-cut site underlined). All primers were synthesized by the DNA Synthesis Laboratory, Department of Molecular & Cell Biology, University of Cape Town, South Africa. RT-PCR was performed on PLRV contaminated potato leaves from the Sandveld region (Sand-B) as described (see chapter 3 of this thesis). The RT-PCR product of the PLRV CP gene was cloned into pGEM-T Easy (Promega) and transformed into *E. coli* JM109 cells as described (see chapter 3 of this thesis). Subsequently, this PLRV CP construct was sequenced as described (see chapter 3 of this thesis).

Nco1 and Nde1 subcloning into pET14-b

The PLRV CP clone that was used for subcloning into pET14-b was SandB-3 (see addendum A of this thesis). The PLRV CP gene was digested from the recombinant PLRV CP pGEM-T Easy construct using *Nco*1 and *Nde*1 (Roche) at their restriction enzyme cut sites that were added at the 5' ends of the forward and reverse primer, respectively. Subsequently, the PLRV CP gene was subcloned into the *Nco*1 and *Nde*1 site of pET14-b (Novagen). The ligation reaction for subcloning was performed as described above. Before transformation of the PLRV CP pET14-b construct into the respective expression strains, the construct was transformed into non-expression host *E. coli* strain JM109 (Promega) as described (see chapter 3 of this thesis) and purified as described above.

Nco1 and Xho1 subcloning into pET14-b

Before the restriction enzymes *Nco*1 and *Xho*1 was used for sub-cloning, a new set of primers was used with these restriction enzyme cut sites to amplify the PLRV CP by PCR from the pGEM-T Easy plasmid used for the *Nco*1 and *Nde*1 restriction enzyme digestion. The additional bases of the *Nde*1 reverse primer were altered to enable subcloning with *Xho*1. The sequence of the forward primer (for *Nco*1 subcloning) was the same as above and that of the reverse primer 5´-<u>CTCGAG</u>CTATTTGGGGTTCTGCAAAGC-3´ (*Xho1*-cut site underlined). The PCR reaction that was used for the different primer sets was a standard 10 µl reaction mixture consisting of 1 µl 10x PCR Buffer (Southern Cross Biotechnology), 1.5 mM MgCl$_2$, 0.1 µl Taq$^{\text{TM}}$

DNA polymerase (5 U/µl, Southern Cross Biotechnology), 1 pmol/µl forward primer, 1 pmol/µl reverse primer, 0.2 mM dNTPs, approximately 85 ng plasmid DNA (the pGEM-T Easy plasmid used for the *Nco*1 and *Nde*1 restriction enzyme digestion) and deionized water to a final volume of 10 µl. Subsequently amplifications were performed using the Hybaid Px2 Thermal Cycler programmed as follows: 1 cycle of PCR amplification at 94°C for 5 min followed by 25 cycles of 94°C for 30 sec, 55°C for 30 sec and 72°C for 30 sec, final extension was done at 72°C for 7 min. Samples were kept at 15°C until used.

PCR product formation was confirmed by agarose gel electrophoresis. The PCR products were separated on an agarose gel (1% w/v), in 1x TAE electrophoresis buffer (0.48% w/v Tris-base; 0.11% glacial acetic acid; 0.5 M EDTA, pH 8.0). Electrophoresis was performed at 110 V for 1 h. Gels were soaked in deionized water containing 1 µg/ml ethidium bromide for 20 min and stained DNA was visualized using an UV transilluminator. PCR products were excised from the gel and the DNA purified using a Wizard® SV Gel and PCR Clean-Up System kit (Promega) according to manufacturer's instructions. Subsequently, the PLRV CP gene with *Nco*1 and *Xho*1 cut sites was cloned into pGEM-T Easy (Promega) and sequenced as described (see chapter 3 of this thesis). Thereafter, the PLRV CP gene was digested from the recombinant pGEM-T Easy plasmid using *Nco*1 and *Xho*1 (New England Biolabs) at the restriction enzyme cut sites that were added at the 5' ends of the forward and reverse primer, respectively.

Subsequently, the PLRV CP gene was subcloned into the *Nco*1 and *Xho*1 site of pET14-b (Novagen). The ligation reaction for subcloning was performed as described above. Before transformation of the PLRV CP pET14-b construct into the respective expression strains, the construct was transformed into non-expression host *E. coli* strain JM109 (Promega) as described (see chapter 3 of this thesis) and purified as described above.

EcoR1 and Xho1 subcloning into pGEX-6P-2

Before the restriction enzymes *Eco*R1 and *Xho*1 were used for sub-cloning, a new set of primers was used with these restriction enzyme cut sites to amplify the PLRV CP by PCR from the pGEM-T Easy plasmid used for the *Nco*1 and *Nde*1 restriction enzyme digestion. The additional bases of the *Nco*1 forward primer were altered to enable subcloning with *Eco*R1. For the *Eco*R1 forward primer (5'-<u>GAATTC</u>**AG**ATGGGTACGGTCGTGGTTAAAGG-3´, *Eco*R1-cut site underlined) two additional bases were added for the PLRV CP (shown in bold) to be expressed in frame with GST. The corresponding amino acid encoded by these nucleotides (C**AG**) is glycine that has a side-chain that would not interfere with protein production. The sequence of the reverse primer (for *Xho*1 subcloning) was the same as above. The PCR reaction that was used for the different primer sets as well as insert preparation was as described above.

The PLRV CP gene with *Eco*R1 and *Xho*1 cut sites was cloned into pGEM-T Easy (Promega) and sequenced as described (see chapter 3 of this thesis). Thereafter, the PLRV CP gene was digested from the recombinant pGEM-T Easy plasmid using *Eco*R1 and *Xho*1 (New England Biolabs) at their restriction enzyme cut sites that were added at the 5' ends of the forward and reverse primer, respectively.

The PLRV CP gene was then subcloned into the *Eco*R1 and *Xho*1 site of pGEX-6P-2 (Amersham) downstream of the sequence encoding GST. The ligation reaction for subcloning was performed as described above. Before transformation of the PLRV CP pGEX-6P-2 construct into the respective expression strains, the construct was transformed into non-expression host *E. coli* strain JM109 (Promega) as described (see chapter 3 of this thesis). White colonies containing the PLRV CP pGEX-6P-2 constructs were subjected to direct PCR as described (see chapter 3 of this thesis) using pGEX promoter primers flanking the gene (forward primer: 5'-GGG CTG GCA AGC CAC GTT TGG TG-3'; reverse primer: 5'-CCG GGA GCT GCA TGT GTC AGA GG-3') and a toothpick scrape of the colonies as template DNA. The pGEX forward and reverse primers were synthesized by the DNA Synthesis Laboratory, Department of Molecular & Cell Biology, University of Cape Town, South Africa. The PLRV CP pGEX-6P-2 constructs were purified from *E. coli* strain JM109 using a Plasmix Miniprep kit (Talent) according to manufacturer's instructions.

Sequencing analysis of the PLRV CP pET14-b and PLRV CP pGEX-6P-2 constructs

The PLRV CP pET14-b construct was sequenced using T7 forward primer as described see chapter 3 of this thesis) in order to confirm that the PLRV CP sequence was correctly cloned for expression (see addendum C of this thesis). Sequence chromatograms were edited in Chromas v1.45 (Technelysium Pty., Tewantin, Australia).

The PLRV CP pGEX-6P-2 was sequenced using the pGEX forward and reverse primers as described (see chapter 3 of this thesis) in order to confirm that the PLRV CP sequence was correctly cloned for expression (see addendum D of this thesis). Sequence chromatograms were edited in Chromas v1.45 (Technelysium Pty., Tewantin, Australia).

Expression of PLRV CP in the pET14-b system in E. coli strain BL21(DE3)pLysS

The PLRV CP pET14-b construct containing the *Nco*1 and *Xho*1 restriction enzyme cut sites was transformed into *E. coli* BL21(DE3)pLysS (Novagen) cells. Transformation of the expression hosts and inoculation of colonies was performed as described (see chapter 3 of this thesis), however approximately 200 ng of the PLRV CP pET14-b construct and 50 µl of competent cell suspension were used. The LB plates were supplemented with the appropriate antibiotics for the *E. coli* strain BL21(DE3)pLysS in the required concentrations (50 µg/ml ampicillin and 34 µg/ml chloramphenicol) and multiple colonies were inoculated per LB medium supplemented with 1% glucose and the specific antibiotics for the strain (see above). Overnight

cultures of transformed BL21(DE3)pLysS strain was inoculated to a dilution of x 50 into TB medium (12 g/L tryptone, 24 g/L yeast extract, 0.0004% glycerol, in 0.01 M potassium phosphate buffer) containing 1% glucose with 50 µg/ml ampicillin and 34 µg/ml chloramphenicol. The cultures were grown for approximately 3 to 4 h to an OD_{600} of 0.6. Expression of the *E. coli* strain BL21(DE3)pLYSs with PLRV CP pET14-b construct was induced by the addition of IPTG to 0.4 mM. The induced cells were grown for a further 4 h at 37°C. Cells were harvested for protein purification by placing the induction mixture on ice for approximately 5 min followed by centrifugation at 10,000 x g for 5 min and the cell pellets resuspended in a tenth of the original cell culture volume TEN-50 buffer (20 mM Tris, pH 8; 1 mM EDTA, pH 8; 0.2 mM DTT, 0.1% Triton X100, 50 mM NaCl; 10% glycerol) and stored at -80°C. One ml aliquots were taken at different growth points during the induction, their OD_{600} values noted and stored as described above.

Protein profiles of bacterial fractions were compared by separation on sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) (Laemmli, 1970). Aliquots of the cell samples were taken according to their OD_{600} values so that the amount of cells for the different growth points was approximately the same. For the release of the proteins from the cell interior, the different time point samples that were stored at -80°C were transferred to 37°C to thaw. After thawing, the samples were again frozen at -80°C. This freeze-thaw cycle was repeated three times to break open the cells under the influence of T7 lysozyme encoded by the pLysS plasmid, which further improves lysis of the *E. coli* cells. For shearing the DNA and reducing the viscosity of the cell lysate, the samples were pushed through a G25 syringe-needle approximately three times. This process made the samples less viscous and easier to handle for SDS-PAGE analysis. An aliquot of each sample to be analyzed was mixed with an equal volume of treatment buffer (0.125 M Tris-Cl, 4% (w/v) SDS, 20% (v/v) glycerol, 10% (v/v) 2-mercapto-ethanol, pH 6.8) and 0.2 volumes of bromophenol blue (0.1% (w/v) in a 1.5 mM NaOH solution), and incubated on a 90°C hotplate for 10 min.

Electrophoresis was carried out at 20 mA in a resolving gel (dimensions: 8.4 cm x 10.4 cm x 1.5 mm) [12 or 14% (depending on requirements) T, 2.7% C_{bis}, 0.1% SDS, 0.375 M Tris-HCl, pH 8.8], using a Omeg Scientific Vertical Gel Electrophoresis Unit. Ten-well stacking gels (dimensions: 10.4 cm x 1 cm x 1.5 mm) (4.5% T, 2.7% C_{bis}, 0.125 M Tris-HCl, 1% SDS, pH 6.8) and a single tank buffer (25 mM Tris, 192 mM glycine, 0.1% SDS, pH 8.3) were used in this system. Rainbow^{TM} coloured protein molecular weight markers (Amersham) were treated in the same manner as bacterial fractions for use as protein standards (14.3-220 kDa). Protein bands were visualized after staining of the resolving gel [1 hour in 0.025% (w/v) Coomassie Blue R-250, 40% (v/v) methanol at 37°C], followed by destaining [1 hour in Destain 1, 50% (v/v) methanol, 10% (v/v) acetic acid at 37°C] and overnight in Destain 2 [7% (v/v) acetic acid, 5% (v/v) methanol] at room temperature. The relative molecular mass (M_r) of protein bands of interest was calculated by comparison of the relative mobilities (Rf-values) with those of the

protein standards that were electrophoresed on the same gel, plotted on a calibration curve (Rf-values vs log M_r). The Rf-value of a particular band was obtained by division of its mobility (mm from top of resolving gel) by the corresponding mobility of the dye front.

Expression of PLRV CP in the pET14-b system in E. coli strain Rosetta-gamiB (DE3)pLysS

The PLRV CP pET14-b construct containing the *Nco*1 and *Xho*1 restriction enzyme cut sites was transformed into *E. coli* Rosetta-gami B(DE3)pLysS (Novagen). Transformation of the expression hosts and inoculation of colonies was performed as described (see chapter 3 of this thesis), however approximately 50 ng of plasmid and 9 µl of competent cell suspension was used. The LB content of the transformation was not centrifuged before plating out. Furthermore, the LB plates were supplemented with the appropriate antibiotics for the *E. coli* Rosetta-gami B(DE3)pLysS strain in the required concentrations (50 µg/ml ampicillin, 15 µg/ml kanamycin, 12.5 µg/ml tetracycline and 34 µg/ml chloramphenicol) and multiple colonies were inoculated per LB medium supplemented with 1% glucose and the specific antibiotics for the strain (see above). Overnight cultures of the respective strains were inoculated to a dilution of x 50 into TB medium (12 g/L tryptone, 24 g/L yeast extract, 0.0004% glycerol, 0.01 M potassium phosphate buffer) containing 1% glucose with 50 µg/ml ampicillin, 15 µg/ml kanamycin, 12.5 µg/ml tetracycline and 34 µg/ml chloramphenicol. The cultures were grown for approximately 5 h to an OD_{600} of 0.6. Expression of the strain with the PLRV CP pET14-b construct was induced by the addition of IPTG to 0.4 mM. The induced cells were grown for a further 4 h at 37°C. Cells were harvested and stored as described above.

Preparation of the bacterial samples for SDS-PAGE and SDS-PAGE analysis was performed as described above.

Expression of PLRV CP in the pET14-b system in E. coli strain Rosetta-2(DE3)pLysS

The PLRV CP pET14-b construct containing the *Nco*1 and *Xho*1 restriction enzyme cut sites was transformed into *E. coli* Rosetta-2(DE3)pLysS (Novagen). Transformation of the expression hosts was performed as described above for expression of *E. coli* strain Rosetta-gami B(DE3)pLysS. The LB plates were supplemented with the appropriate antibiotics for the *E. coli* strain Rosetta-2(DE3)pLysS in the required concentrations (50 µg/ml ampicillin and 34 µg/ml chloramphenicol) and multiple colonies were inoculated per LB medium supplemented with 1% glucose and the specific antibiotics for the strain (see above). Overnight cultures of the respective strains were inoculated to a dilution of x 50 into TB medium (12 g/L tryptone, 24 g/L yeast extract, 0.0004% glycerol, 0.01 M potassium phosphate buffer) containing 1% glucose with 50 µg/ml ampicillin and 34 µg/ml chloramphenicol (BL21(DE3)pLysS. The cultures were grown for approximately 3 to 4 h to an OD_{600} of 0.6. Expression of the *E. coli* strain Rosetta-2(DE3)pLysS with the PLRV CP pET14-b construct was induced by the addition of IPTG to 0.4

mM. The induced cells were grown for a further 4 h at 37°C. Cells were harvested for protein purification and stored as described above.

Preparation of the bacterial samples for SDS-PAGE and SDS-PAGE analysis was performed as described above.

For cell lysates where the SDS-PAGE analysis indicated positive induction, the proteins were further purified. Four hours after induction of the cell culture containing the PLRV CP pET14-b construct, ammonium sulphate precipitation of the induced recombinant protein was performed as follows; the cell culture was resuspended in a tenth of the original cell culture volume of TEN-50 buffer and was taken through three cycles of freeze-thawing as described for the preparation of SDS-PAGE analysis samples. Subsequently, to 18 ml of the cell lysate, 12 ml of saturated ammonium sulphate was added to achieve 40 % ammonium sulphate saturation and precipitation. The sample was incubated for 2 h at 4°C and centrifuged for 20 min at 27,200 x g. To 29 ml of the supernatant a further 5.8 ml saturated ammonium sulphate was added to achieve 50 % ammonium sulphate precipitation. The sample was incubated for 2 h at 4°C and centrifuged for 20 min at 27,200 x g. The pellet was redissolved in 3 ml TEN-50 buffer and a further 2.5 ml of saturated ammonium sulphate was added to 2.5 ml of the redissolved pellet and incubated for 2 h at 4°C for a second 50% ammonium sulphate precipitation. The mixture was centrifuged for 20 min at 27,200 x g and the remaining pellet was redissolved in 1 ml TEN-50 buffer. Aliquots were taken from each fraction during the precipitation, (40% precipitation pellet (pellet resuspended in 1 ml TEN-50 buffer), 40% precipitation supernatant (1 ml of the supernatant was stored), first 50% precipitation pellet (500 µl of the TEN-50 resuspended pellet was stored), first 50% precipitation supernatant (1 ml of the aliquot was stored), second 50% precipitation pellet (pellet resuspended in 1 ml TEN-50 buffer), second 50% precipitation supernatant (1 ml aliquot stored) and dialyzed at 4°C overnight against two changes of TEN-50 buffer (without glycerol and Triton-X100). These samples were subsequently used for SDS-PAGE analysis.

Expression of PLRV CP in the pGEX-6P-2 system in E. coli strain BL21(DE3)pLysS and Rosetta-2(DE3)pLysS

The PLRV CP pGEX-6P-2 construct was transformed into *E. coli* BL21(DE3)pLysS (Novagen) and Rosetta-2(DE3)pLysS (Novagen) cells. Transformation of the expression hosts was performed as described above for expression of *E. coli* strain Rosetta-gami B(DE3)pLysS. The LB plates were supplemented with the appropriate antibiotics for the *E. coli* strains BL21(DE3)pLysS and Rosetta-2(DE3)pLysS in the required concentrations (100 µg/ml ampicillin and 34 µg/ml chloramphenicol) and multiple colonies were inoculated per LB medium supplemented with 1% glucose and the specific antibiotics for the strains (see above). Overnight cultures of the respective strains were inoculated to a dilution of x 50 into TB medium (12 g/L tryptone, 24 g/L yeast extract, 0.0004% glycerol, 0.01 M potassium phosphate buffer)

containing 1% glucose with 100 μg/ml ampicillin and 34 μg/ml chloramphenicol. The cultures were grown for approximately 3 to 4 h to an OD_{600} of 1.0. Expression of the strains with the PLRV CP pGEX-6P-2 construct was induced by the addition of IPTG to 0.1 mM. The induced cells were grown for a further 4 h at 37°C. Cells were harvested for protein purification and stored as described above.

Preparation of the bacterial samples for SDS-PAGE and SDS-PAGE analysis was performed as described above.

Column purification

For purification of cell cultures containing the PLRV CP pGEX-6P-2 constructs GST•Bind^{TM} Resin (Novagen) was used. The pellet of the induced cells, resuspended in TEN-50 buffer and stored at – 80°C, was treated with cycles of freezing and thawing as described above for the SDS-PAGE samples. Thereafter, the samples were also treated with a syringe as described for the SDS-PAGE samples, however, because of the volume of the sample a 1.8 x 40 mm Luerlock needle, 18 G, 21G, 23G and 25G needles were used in order of the greatest syringe size to the least to ease the syringe treatment. The syringe treated sample was centrifuged at 16,000 x g for 20 min at 4°C and the pellet discarded so that only the soluble phase was used for further purification (it was established that the recombinant fusion protein was largely expressed in the soluble fraction). The soluble phase of the cell lysate was pulse sonicated briefly (four repeats of 10 sec each) in a Beckman Sonicator Model W-225R Cell Disrupter and centrifuged again at 16,000 x g for 20 min at 4°C. The supernatant was further purified through a non-pyrogenic, low protein binding Millipore 0.45 μM filter unit and subsequently used for column purification.

Before applying the cell lysate to the column, the GST•Bind^{TM} Resin (10 ml column volume) was washed with three column volumes of sterile distilled water (column flow performed under gravity force) and equilibrated with three column volumes ice cold PBS (0.15 M, pH 7.2). Thereafter, approximately two column volumes of the cleared and filtered cell lysate was applied to the GST•Bind^{TM} Resin and the flowthrough collected. The cell lysate flowthrough was reapplied to the column and the flowthrough collected to reapply it a third time to the column. The column was washed with nine column volumes of a PBS buffer with a higher NaCl concentration and pH as well as added Triton X-100 to decrease non-specific binding to the column (300 mM NaCl, pH 8, 1% Triton X-100). GST-fusion proteins that bound to the column were eluted with ten 2 ml volumes (representing a fifth of the column volume) of reduced glutathione (50 mM Tris pH 8.0, 10 mM reduced glutathione). Fractions (2 ml) were collected in Eppendorf tubes (2 x 1 ml) and stored at -80°C. The column was washed with three column volumes of cleansing buffer 1 (0.1 M borate buffer, pH 8.5 containing 0.5 M NaCl) and three column volumes of cleansing buffer 2 (0.1 M acetate buffer, pH 4.5 containing 0.5 M NaCl). For storage the column was washed with a half column volume of 20% ethanol and stored at 4°C.

Fractions of each purification step were collected and stored at –80°C in Eppendorf tubes for SDS-PAGE analysis. Five µl of each fraction was electrophoresed with SDS-PAGE as described above for confirmation of purification.

Antibody production

The purified recombinant GST-PLRV CP fusion protein was dialyzed at 4°C against deionized water for 4 h. Thereafter, the dialysis tubes were placed in 0.1 x PBS for dialysis overnight at 4°C. The mixture was subsequently concentrated to a seventh of the original volume on a Speedvac Savant rotary evaporator and the protein concentration determined using the Bradford method (Bradford, 1976) that was modified for use in microtitre plates. A standard solution of BSA (fraction V, 2 mg/ml in 0.1 x PBS) was used to prepare a standard dilution series. Each standard was pipetted in duplicate into a microtitre plate and 250 µl of Bradford reagent [0.01% (w/v) Coomassie Brilliant Blue G-250, 4.7% (w/v) ethanol, 8.5% (w/v) phosphoric acid] was added and incubated for 2 min at room temperature. The plate was placed into a Titertek Multiscan spectrophotometer and the absorbances were read at 620 nm. BSA standards were used to set up a standard curve to calculate concentrations of samples of unknown protein concentration. Alternatively, the protein concentration of some of the samples was determined using a BCA protein assay (Pierce) according to the manufacturer's instructions after overnight dialysis of the respective samples against two changes of PBS to remove the glutathione.

Antibodies were raised to expressed GST/PLRV-CP fusion protein according to Bellstedt *et al.* (1987) with the use of naked bacteria. Briefly, for immunisation purposes the recombinant PLRV CP was firstly adsorbed to naked bacteria. A standard 2 mg/ml suspention of naked bacteria was prepared in sterile water for use in the adsorption to the antigenic recombinant PLRV CP. Protein, 120 µg, was added to 250 µl of naked bacteria solution and homogenized with a loosely fitting Teflon plunger. The mixture was subsequently dried by rotaty evaporation on a Speedvac Savant and resuspended in PBS to the desired final concentration (40 µg of recombinant GST-PLRV CP fusion protein adsorbed to 200 µg naked bacteria/0.5 ml PBS). Rabbits were immunised with this amount of recombinant PLRV CP adsorbed to naked bacteria on days 0, 3, 8, 14, 17, 21, 28, 31, 35, 84, 87 and 90. Test bleeds of 2 ml were drawn on days 0, 21 and 28 and larger bleeds (20 ml) on days 42 and 97.

Protease cleavage

In order to cleave GST from the GST-PLRV CP fusion protein, the GST-PLRV CP fusion protein was bound and washed on the GST•Bind™ Resin as described previously. The bed volume in this case was 2 ml and the bacterial lysate applied to the column was approximately twelve bed volumes. The column was washed with ten bed volumes of PreScission cleavage buffer (50 mM Tris-HCl, 150 mM NaCl, 1 mM EDTA, 1 mM DTT, pH 7.5) and equilibrated at 4°C for 30 min.

PreScission Protease enzyme (Amersham) (20 units of enzyme) was added per mg fusion protein that was bound per ml of resin and incubated on a rotor at low speed overnight at 4°C. Samples were eluted with 20 bed volumes of PreScission cleavage buffer that were collected in either 1, 2 or 4 ml fractions. Samples were concentrated to a eighth of their original volume by rotary evaporation on a Speedvac Savant. These samples were analysed on a SDS-PAGE gel as described previously and approximately 1 µg samples were loaded onto the gel, in the cases where the protein concentration of the fractions was too low, a maximum amount of 20 µl was loaded on the gel.

Western blot

Prior to western blotting, all samples were subjected to 14% SDS-PAGE, as described before. Separated protein was subsequently transferred electrophoretically (16-20 h at 120 mA) to a nitrocellulose membrane (0.45 µm, Schleicher and Schuel) in a buffer consisting of 0.05 M Tris, 0.2 M glycine and 20% (v/v) methanol (pH 8.3). The western blot analysis was performed in a multi-step procedure, all steps were carried out under constant agitation and the antibody incubating steps was carried out at 37°C. Firstly, the unoccupied area of the nitrocellulose membrane was blocked with casein buffer (154 mM NaCl, 0. 5% (w/v) casein, 10 mM Tris-Cl, 0.02% thiomersal, pH 7.6) for 20 min. This was followed by incubation with the desired dilution (in casein buffer) of the primary antibody (either anti-PLRV (Neogen), 1/200, 1/500 or produced anti-recombinant fusion protein, 1/1000) for 1 h. Excess antibody was removed by three washes of 5 min each in PBS-Tween (0.15 M PBS, pH 7.2, supplemented with 0.1% (v/v) Tween-20). The membrane was subsequently incubated with goat anti-rabbit antibodies (Sigma, diluted 1/500 in casein buffer) for 1 h. After repetition of the washing step, the membrane was incubated with rabbit peroxidase-anti-peroxidase complex (Sigma,) for 1 h and washed again. Specific protein bands were visualized as black/blue precipitates by addition of substrate solution (4-chloro-1-naphtol (18 mg) dissolved in cold methanol (6 ml) which was added to 30 ml PBS (pH 7.2) containing 9 µl of a 34% H_2O_2 solution). The substrate reaction was terminated after 30 min at room temperature by washing the membrane three times with distilled water. Due to the light-sensitivity of substrate reaction products, membranes were blotted dry and stored wrapped in aluminium foil.

In a separate western blot prepared as described above up to the blocking step, the membrane was incubated with a 1/100 dilution of goat anti-GST antibodies (Amersham) in filtered casein buffer. A PBS-Tween buffer was used for wash steps as above. The membrane was incubated in bovine anti-goat IgG-HRP (Santa Cruz Biotechnology) (1/10 000 dilution in filtered casein buffer). After SuperSignal West Pico Chemiluminescent Substrate (Pierce) addition according to the manufacturer's instructions, the wetted blot was exposed to standard autoradiographic film for 30 sec.

Results

Amplification and cloning of the PLRV CP gene into pGEM-T Easy

As described in chapter 3 of this thesis, the CP gene of PLRV contained in the plant samples was readily amplified following the sample preparation technique and using the optimized RT-PCR. Successful cloning could be confirmed using the confirmatory PCR as described (see chapter 3 of this thesis). Sequence analysis confirmed that the RT-PCR product was the PLRV CP gene (627 bp) (results not shown).

Nco1 and Nde1 subcloning into pET14-b

When the *Nco*1 and *Nde*1 digested products of the PLRV CP pGEM-T Easy construct were used for subcloning, colonies with PLRV CP pET14-b constructs were not obtained. It was found after repeated attempts of digestions with the restriction enzymes *Nco*1 and *Nde*1 that *Nde*1 may not have digested the pGEM-T Easy and pET14-b plasmids completely (Fig. 1 and Fig. 2).
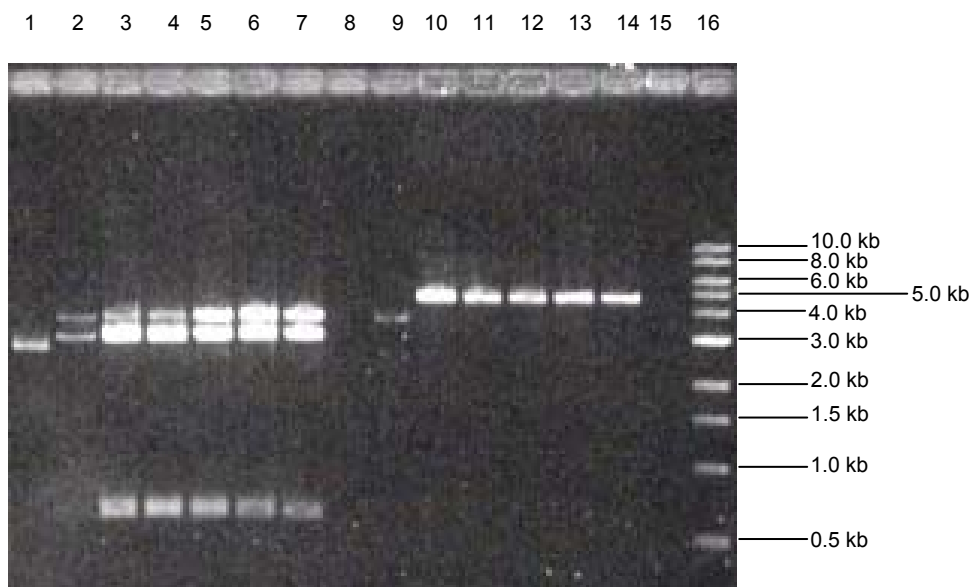
Fig. 1. *Nco*1 and *Nde*1 restriction enzyme digests of PLRV CP pGEM-T Easy construct and pET14-b plasmid. Lane 1: undigested PLRV CP pGEM-T Easy construct; lane 2-7: restriction enzyme digests of the PLRV CP pGEM-T Easy construct; lane 8: no sample; lane 9: pET14-b plasmid undigested; lane 10-14: restriction enzyme digests of the pET14-b plasmid; lane 15: no sample; lane 16: 1 kb DNA ladder (New England Biolabs).

Evidence that double restriction enzyme digestion of the PLRV CP pGEM-T Easy construct was not complete is shown in Fig. 1 (lanes 2-7). When the pET14-b plasmid was digested with *Nco*1 only, it was found that the pET14-b plasmid was fully digested to yield a single cleaved plasmid band (Fig. 2, lane 4, 5). When the pET14-b plasmid was digested with *Nde*1 only it was found that the pET14-b plasmid was not completely digested, which appears to be the result of incomplete cleavage by the *Nde*1 restriction enzyme (Fig. 2, lane 6, 7).

For this reason, an approach in which the restriction enzymes *Nco*1 and *Xho*1 were used for cleaving the PLRV CP gene from the PLRV CP pGEM-T Easy construct, was followed.
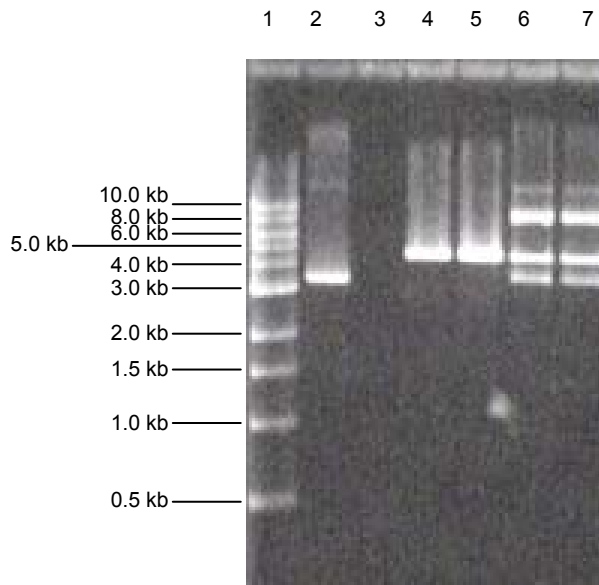
Fig. 2. *Nco*1 and *Nde*1 restriction enzyme digests of PLRV CP pGEM-T Easy construct and pET14-b plasmid. Lane 1: 1 kb DNA ladder (New England Biolabs); lane 2: pET14-b plasmid undigested; lane 3: no sample; lane 4-5: *Nco*1 restriction enzyme digest of pET14-b plasmid; lane 6-7: *Nde*1 restriction enzyme digest of pET14-b plasmid.

Nco1 and Xho1 subcloning into pET14-b

Digestion of the PLRV CP pGEM-T Easy construct and the pET14-b plasmid with *Nco*1 and *Xho*1 was successful as is shown in Fig. 3. Sequencing revealed that the PLRV CP gene was successfully cloned into the expression plasmid (pET14-b) using these restriction enzymes.
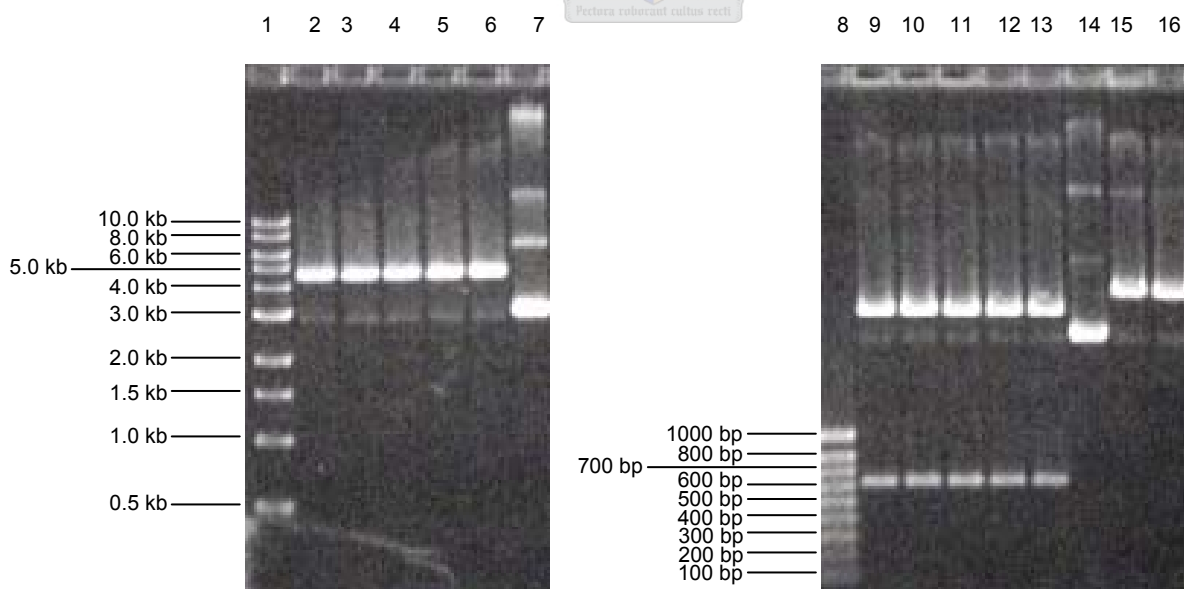
Fig. 3. *Nco*1 and *Xho*1 restriction enzyme digests of PLRV CP pGEM-T Easy construct and pET14-b plasmid. Lane 1: 1 kb DNA ladder (New England Biolabs); lane 2-6: *Nco*1 and *Xho*1 restriction enzyme digests of pET14-b plasmid; lane 7: pET14-b plasmid undigested; lane 8: HyperLadder IV (Bioline); lane 9-13: *Nco*1 and *Xho*1 restriction enzyme digests of the PLRV CP pGEM-T Easy construct; lane 14: PLRV CP pGEM-T Easy construct undigested; lane 15: *Xho*1 restriction enzyme digest of PLRV CP pGEM-T Easy construct; lane 16: *Nco*1 restriction enzyme digest of PLRV CP pGEM-T Easy construct.

EcoR1 and Xho1 subcloning into pGEX-6P-2

The CP containing insert could be cleaved from pGEM-T Easy with the restriction enzymes *Eco*R1 and *Xho*1 as shown in Fig. 4. Sequencing revealed that the PLRV CP gene was successfully cloned into the expression plasmid (pGEX-6P-2) using these restriction enzymes.
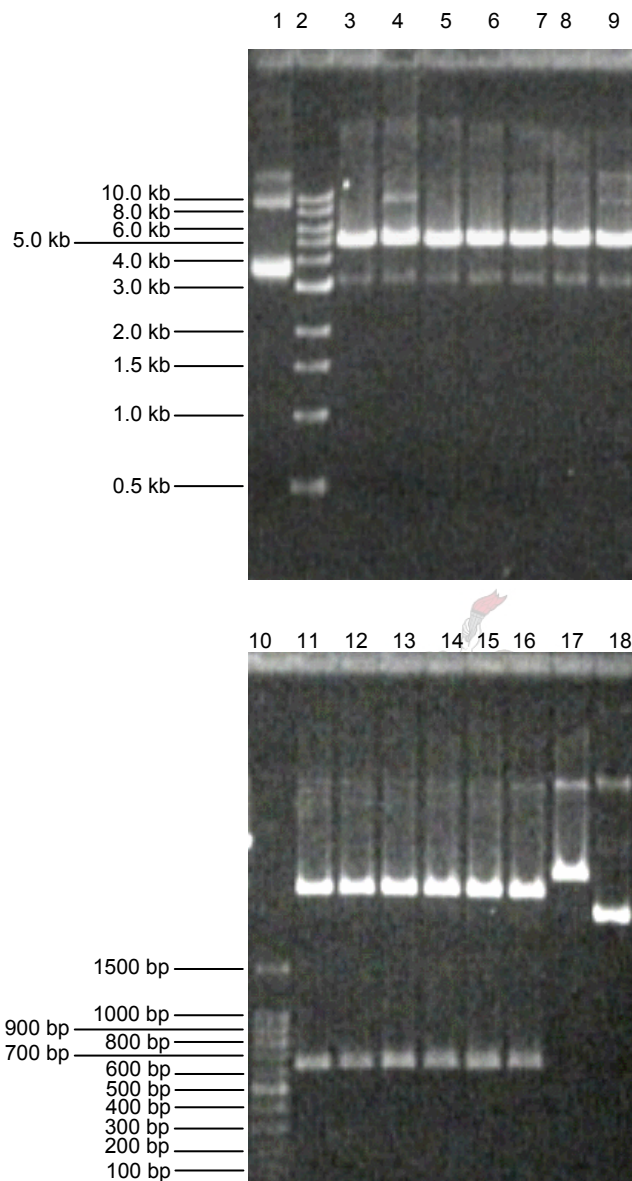
Fig. 4. *Eco*R1 and *Xho*1 restriction enzyme digests of PLRV CP pGEM-T Easy construct and pGEX-6P-2 plasmid. Lane 1: pGEX-6P-2 plasmid undigested; lane 2: 1 kb DNA ladder (New England Biolabs); lane 3-7: *Eco*R1 and *Xho*1 restriction enzyme digests of pGEX-6P-2 plasmid; lane 8: *Eco*R1 restriction enzyme digest of pGEX-6P-2 plasmid; lane 9: *Xho*1 restriction enzyme digest of pGEX-6P-2 plasmid; lane 10: 100 bp DNA ladder (Promega); lane 11-15: *Eco*R1 and *Xho*1 restriction enzyme digests of the PLRV CP pGEM-T Easy construct; lane 16: *Eco*R1 restriction enzyme digest of PLRV CP pGEM-T Easy construct; lane 17: *Xho*1 restriction enzyme digest of PLRV CP pGEM-T Easy construct; lane 18: PLRV CP pGEM-T Easy construct undigested.

Sequencing analysis of the PLRV CP pET14-b and PLRV CP pGEX-6P-2 constructs

Sequence analysis of both of these PLRV CP constructs was performed to confirm correct cloning with a view to ensuring that the PLRV CP could be expressed successfully. Each of the sequences of these PLRV CP constructs revealed the unaltered ribosome binding site,

restriction enzyme cut sites used for their cloning, both primers, the ATG start site and the PLRV CP gene in correct orientation and the stop site (see addendum C and D of this thesis).

Expression of PLRV CP in the pET14-b system in E. coli strain BL21(DE3)pLysS

The PLRV CP pET14-b construct was successfully transformed into the *E. coli* strain BL21(DE3)pLysS as shown by antibiotic selection. The absence of protein bands at 23 kDa when the transformants were induced (lanes 1-4) in comparison to negative controls (lane 6) gave evidence that no recombinant proteins were produced (Fig. 5). The weak presence of a band at 23 kDa in lanes 2 and 3 was not interpreted as indicative of protein expression but was attributed to the loading of higher relative amounts of total proteins in these lanes.
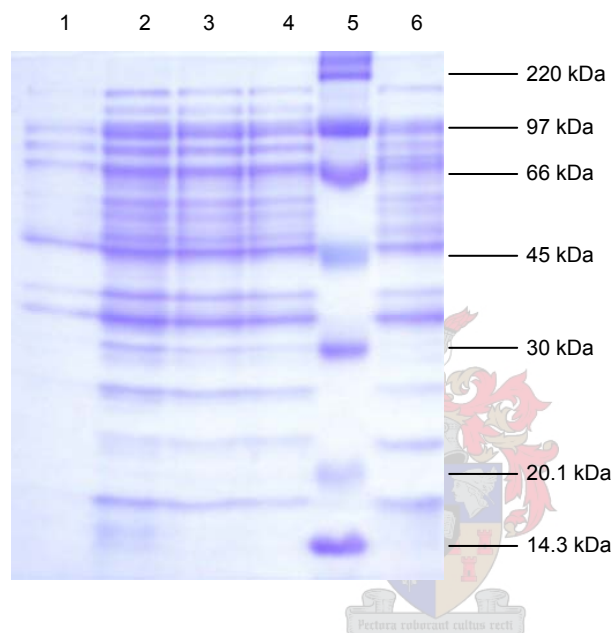
Fig. 5. Profiles of protein expression of PLRV CP in the pET14-b system in *E. coli* strain BL21(DE3)pLysS using reducing SDS-PAGE (14%). Lane 1: BL21 before induction (0 h); lane 2: BL21 4 h after induction; lane 3: BL21 10 h after induction; lane 4: BL21 12 h after induction; lane 5: Rainbow molecular protein marker (Amersham); lane 6: *E. coli* strain Rosetta-gami B(DE3)pLysS with PLRV CP pET14-b construct, 12 h cell growth without induction (negative control). (BL21 = *E. coli* strain BL21(DE3)pLysS with PLRV CP pET14-b construct)

Expression of PLRV CP in the pET14-b system in E. coli strain Rosetta-gami B(DE3)pLysS

Expression of the PLRV CP pET14-b construct in the *E. coli* strain Rosetta-gami B(DE3)pLysS resulted in the production of two proteins (Fig. 6). Neither of these proteins had the expected size of PLRV CP (23 kDa). The molecular weights of the two unknown expressed proteins as shown in Fig. 6, lane 6 (indicated by arrows), were determined with a standard curve. The size of the one protein was 15.6 kDa (indicated by lower arrow) and that of the other protein was 20.9 kDa (indicated by upper arrow). In an attempt to increase expression of the these proteins the conditions for expression were altered (the concentration for IPTG induction was increased from 0.4 mM to 1.0 mM, more than one colony was inoculated for the overnight culture instead of one and 1% glucose was added to the LB plates and overnight culture growth medium).
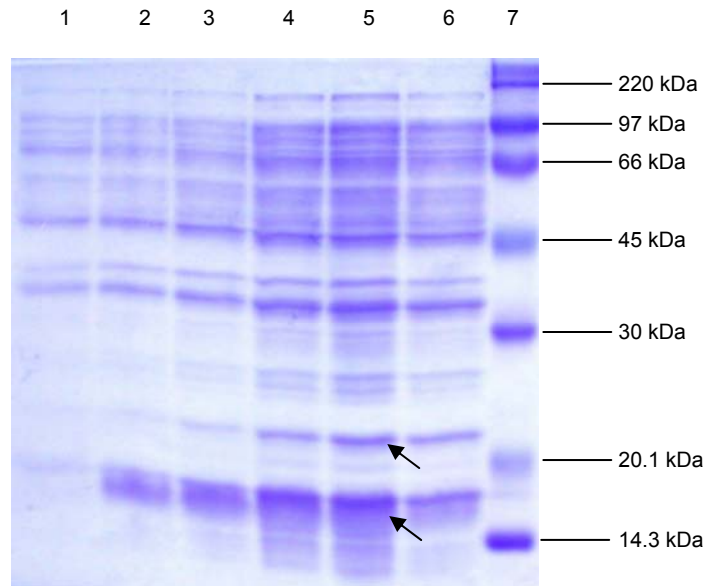
Fig. 6. Profiles of protein expression of PLRV CP in the pET14-b system in *E. coli* strain Rosetta-gami B(DE3)pLysS using reducing SDS-PAGE (14%). Lane 1: Rosetta-gami before induction (0 h); lane 2: Rosetta-gami 4 h after induction; lane 3: Rosetta-gami 6 h after induction; lane 4: Rosetta-gami 8 h after induction; lane 5: Rosetta-gami 10 h after induction; lane 6: Rosetta-gami 12 h after induction; lane 7: Rainbow molecular protein marker (Amersham). Arrows indicate the two recombinant proteins that might have been formed as a result of induction. (Rosetta-gami = *E. coli* strain Rosetta-gami B(DE3)pLysS with PLRV CP pET14-b construct)

Despite of these changes made in the conditions of expression, the amounts of the two induced proteins was not increased (Fig. 7). However, when the conditions of expression were altered, apparently only the recombinant protein of approximately 15.6 kDa was produced (Fig. 7, lanes 3, 6, 7, 8).
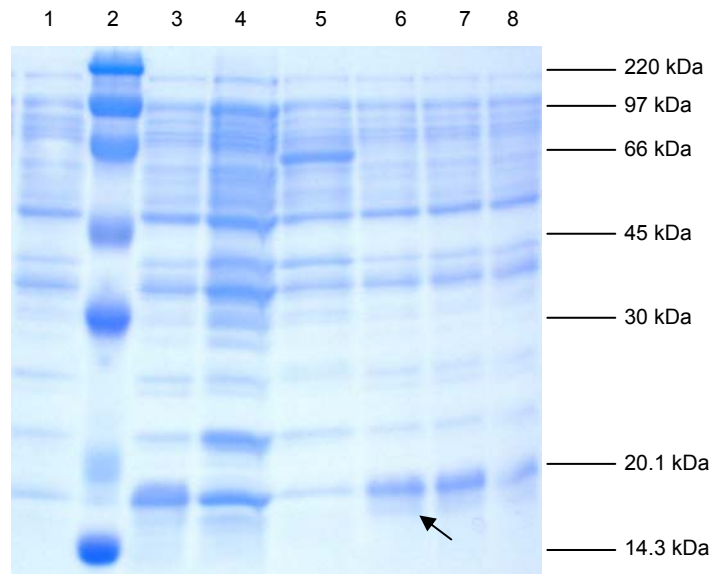
Fig. 7. Profiles of protein expression of PLRV CP in the pET14-b system in *E. coli* strain Rosetta-gami B(DE3)pLysS with altered expression conditions using reducing SDS-PAGE (14%). Lane 1: Rosetta-gami before induction (0 h); lane 2: Rainbow molecular protein marker (Amersham); lane 3: Rosetta-gami 8 h after induction; lane 4: Rosetta-gami 12 h cell growth without induction (negative control); lane 5: positive control 4 h after induction; lane 6: Rosetta-gami 6 h after induction; lane 7: Rosetta-gami 4 h after induction; lane 8: Rosetta-gami 2 h after induction. Arrow indicates the recombinant protein that might have been formed as a result of induction by IPTG. (Rosetta-gami = *E. coli* strain Rosetta-gami B(DE3)pLysS with PLRV CP pET14-b construct; positive control = *E. coli* strain Rosetta-gami B(DE3)pLysS with PLRV CP pET14-b construct containing a 65 kDa expressible protein)

Expression of PLRV CP in the pET14-b system in E. coli strain Rosetta-2(DE3)pLysS

The results of the expression of the PLRV CP pET14-b construct in Rosetta-2(DE3)pLysS are shown in Fig. 8. Induction was shown to be time dependent with an increase in recombinant protein up to 4 h (lanes 1, 3, 7 and 9) in comparison to uninduced negative controls taken at the same time points (lanes 2, 4, 6 and 8). Expression of the PLRV CP pET14-b construct in the *E. coli* strain Rosetta-2(DE3)pLysS resulted in the production of one protein (Fig. 8, lane 3, 7 and 9) of approximately 15.6 kDa. The second protein (20.9 kDa) that was expressed when the PLRV CP pET14-b construct was expressed in Rosetta-gami B(DE3)pLysS (in Fig. 6 but not in Fig. 7), was not expressed in this bacterial strain.

To ensure that the expressed protein shown in Fig. 8, lanes 3, 7 and 9 was not an expression artifact, the expression was repeated (Fig. 9, lanes 1, 4 and 8) with an additional negative control: the *E. coli* strain Rosetta-2(DE3)pLysS that was not transformed with either the pET14-b plasmid alone or the PLRV CP pET14-b construct (negative control A) (Fig. 9, lanes 2, 5 and 9). Negative control B, an *E. coli* strain Rosetta-2(DE3)pLysS that was transformed with the PLRV CP pET14-b construct (with PLRV CP gene) but not induced was also included (Fig. 9, lanes 3, 6 and 10).
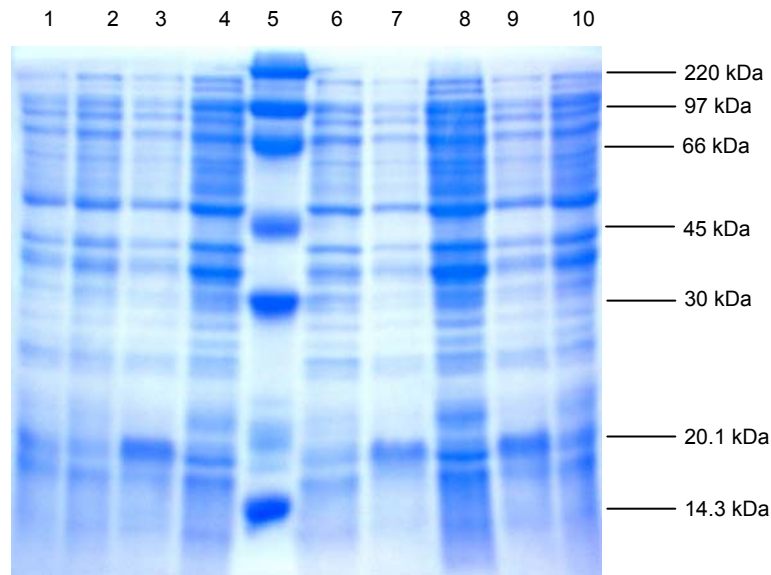
Fig. 8. Profiles of protein expression of PLRV CP in the pET-14b system in *E. coli* strain Rosetta-2(DE3)pLysS using reducing SDS-PAGE (14%). Lane 1: Rosetta-2, 1 h after induction, lane 2: Negative control 1 h cell growth ; lane 3: Rosetta-2, 3 h after induction; lane 4: Negative control 3 h cell growth; lane 5: Rainbow molecular protein marker (Amersham); lane 6: Rosetta-2 before induction (0 h); lane 7: Rosetta-2, 4 h after induction; lane 8: Negative control 4 h cell growth; lane 9: Rosetta-2, 2 h after induction; lane 10: Negative control 2 h cell growth. (Rosetta-2 = *E. coli* strain Rosetta-2(DE3)pLysS with PLRV CP pET14-b construct; Negative control = *E. coli* strain Rosetta-2(DE3)pLysS with PLRV CP pET14-b construct, cell growth without induction).
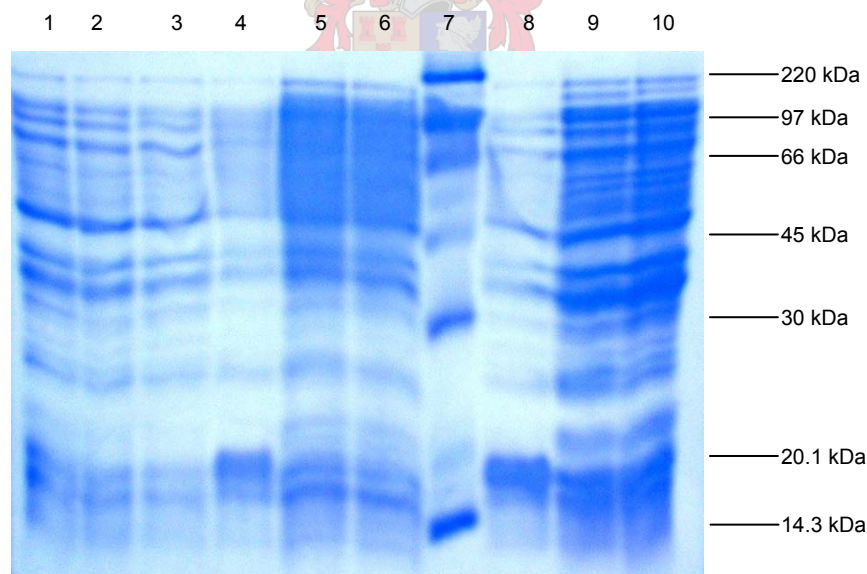
Fig. 9. Profiles of protein expression of PLRV CP in the pET-14b system in *E. coli* strain Rosetta-2(DE3)pLysS with additional negative controls using reducing SDS-PAGE (14%). Lane 1: Rosetta-2 before induction (0 h); lane 2: Negative control A before induction (0 h); lane 3: Negative control B, 0 h cell growth; lane 4: Rosetta-2, 2 h after induction; lane 5: Negative control A, 2 h after induction; lane 6: Negative control B, 2 h cell growth; lane 7: Rainbow molecular protein marker (Amersham); lane 8: Rosetta-2, 4h after induction; lane 9: Negative control A, 4 h after induction; lane 10: Negative control B, 4 h cell growth. (Rosetta-2 = *E. coli* strain Rosetta-2(DE3)pLysS with PLRV CP pET14-b construct; Negative control A = *E. coli* strain Rosetta-2(DE3)pLysS without PLRV CP pET14-b construct; Negative control B = *E. coli* strain Rosetta-2(DE3)pLysS with PLRV CP pET14-b construct, cell growth without induction).

A protein was present in the induced cell cultures with the PLRV CP pET14-b construct (Fig. 9). This protein was not found in the induced but untransformed *E. coli* strain Rosetta-2(DE3)pLysS negative control nor in the uninduced transformed control. In this induction only one potential recombinant protein was produced and not two as was the case with the expression shown in Figs. 7 and 8. The protein that was formed was in the size order of 15.6 kDa and is smaller than the predicted size of PLRV CP (23 kDa). Therefore, the protein that is induced is indeed a product of transformation. Subsequently, attempts were made to purify the induced protein by means of ammonium sulphate precipitation.

Induction of the *E. coli* strain Rosetta-2(DE3)pLysS with the PLRV CP gene in pET14-b was repeated and the protein was purified from the 4 h after induction cell culture. The recombinant protein could not be purified to complete homogeneity, but in comparison with the unprecipitated fraction, the precipitated fraction shown in Fig. 10, lane 7 and 9, contains a greater concentration of the recombinant protein. Furthermore, the second 50% precipitation shown in lane 9, Fig. 10 resulted in an additional enrichment of the recombinant protein in comparison with the precipitated fraction shown in lane 7, Fig. 10.
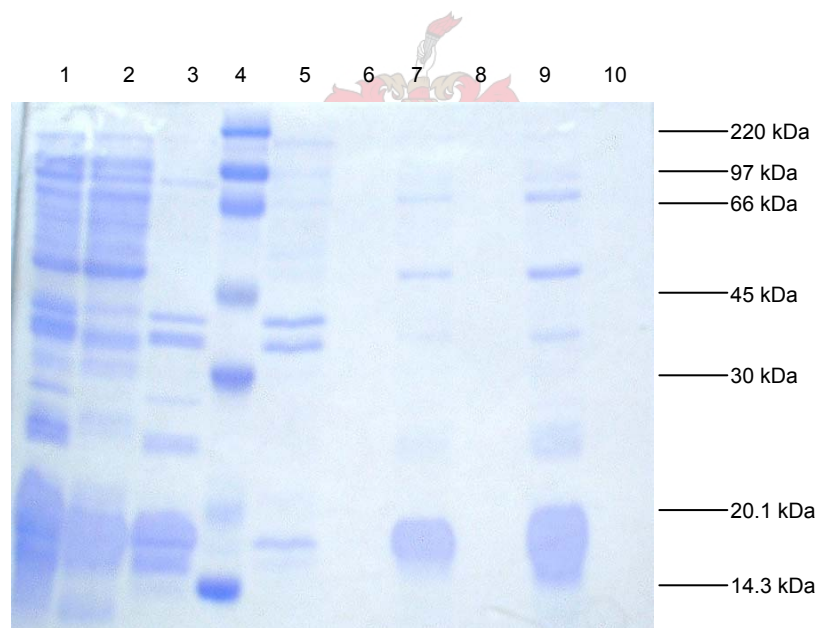
Fig. 10. Protein profiles of ammonium sulphate precipitation of the recombinant protein formed by *E. coli* strain Rosetta-2(DE3)pLysS with the PLRV CP gene in pET14-b using reducing SDS-PAGE (14%). Lane 1: Rosetta-2, 4 h after induction; lane 2: Rosetta-2, 4 h after induction, soluble phase of cell lysate; lane 3: Rosetta-2, 4 h after induction, insoluble phase of cell lysate; lane 4: Rainbow molecular protein marker (Amersham); lane 5: Ammonium sulphate precipitation (40%), pellet; lane 6: Ammonium sulphate precipitation (40%), supernatant; lane 7: Ammonium sulphate precipitation (50%), pellet; lane 8: Ammonium sulphate precipitation (50%), supernatant; lane 9: Ammonium sulphate second 50% precipitation, pellet; lane 10: Ammonium sulphate second 50% precipitation, supernatant. (Rosetta-2 = *E. coli* strain Rosetta-2(DE3)pLysS with PLRV CP pET14-b construct). The fraction of lane 1 was used for ammonium sulphate precipitation.

In order to test whether the induced, recombinant 15.6 kDa protein (Fig. 10, lane 7, 9) was PLRV CP, investigations into the identity of this protein were conducted with a western blot

using commercial anti-PLRV antibodies, which is shown in Fig. 11. All of the bands detected in the western blot were of low intensity and can in all likelihood be ascribed to non-specific binding. Commercial antibodies against PLRV do not bind with the protein of molecular weight 15.6 kDa indicating that the induced protein is not likely to be PLRV CP and is thus possibly an artifact. However, the western blot indicated that the antibodies to PLRV CP bound weakly to a protein contained in the sample shown in lane 4 of a molecular weight of approximately 23 kDa which would be in agreement with the expected size of PLRV CP. If the faint band in the size order of 23 kDa (Fig 11, lane 3) does indicate a low level of PLRV CP production, this level of expression would be too low for efficient protein isolation, purification and immunization.
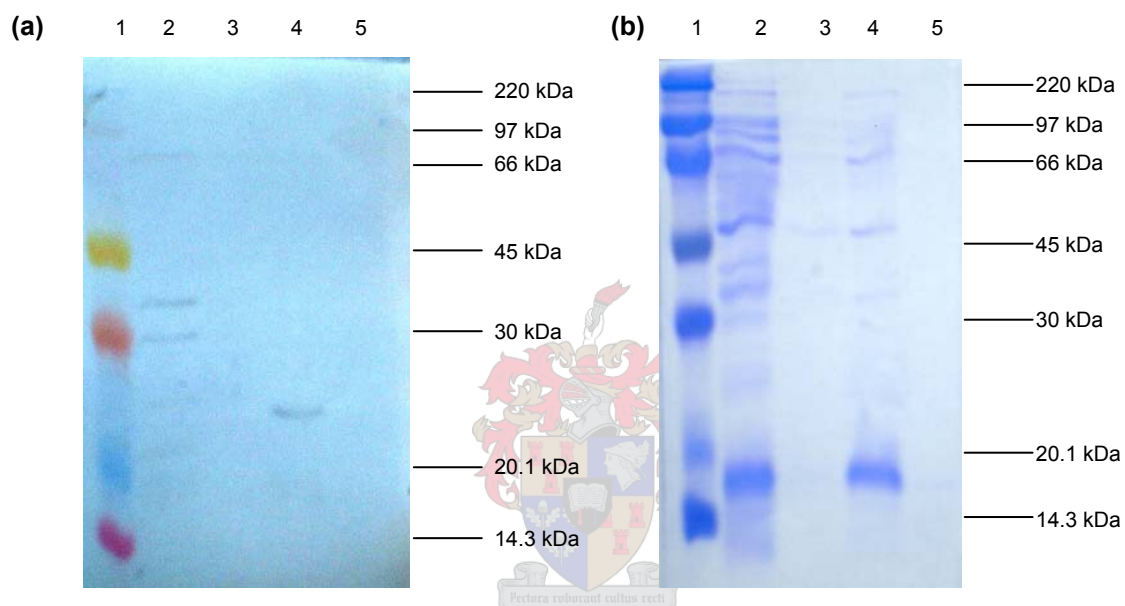
Fig. 11. (a) Western-blot and (b) protein profiles using reducing SDS-PAGE (14%) of recombinant protein purified by ammonium sulphate precipitation. Commercial anti-PLRV antibodies were used for the detection of proteins. For **(a)** and **(b)**: lane 1: Rainbow molecular protein marker (Amersham); lane 2: Rosetta-2, 4 h after induction; lane 3: Rosetta-2, 4 h after induction (sample x10 diluted); lane 4: Ammonium sulphate second 50% precipitation, pellet; lane 5: Ammonium sulphate second 50% precipitation, pellet (sample x10 diluted). (Rosetta-2 = *E. coli* strain Rosetta-2(DE3)pLysS with PLRV CP pET14-b construct). The weak bands seen in (a) could be ascribed to non-specific binding due to the size differences of the detected proteins.

From these results it was concluded that expression of PLRV CP using the pET14-b system with the present PLRV CP constructs was not successful, and even if the expression was performed in strains that supplement the AGA/AGG rare codons underrepresented in *E. coli* strains, it was still unsuccessful. For this reason, the expression of PLRV CP as a fusion protein in the pGEX-6P-2 system was subsequently attempted.

Expression of PLRV CP in the pGEX-6P-2 system in E. coli strain BL21(DE3)pLysS and Rosetta-2(DE3)pLysS

For comparative purposes, but also to circumvent potential shortages in rare codons, the PLRV CP pGEX-6P-2 construct was transformed into *E. coli* strains BL21(DE3)pLysS and Rosetta2-

(DE3) pLysS to determine the efficiency of expression. Transformation of the PLRV CP pGEX-6P-2 construct into *E. coli* strains Rosetta-2(DE3)pLysS and BL21(DE3)pLysS was successful as shown by antibiotic selection.
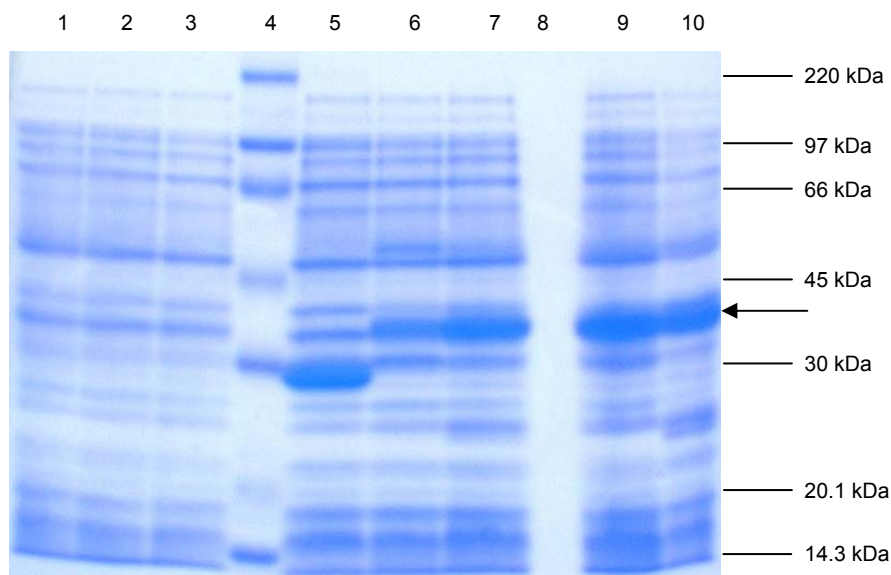
Fig. 12. Profiles of protein expression of PLRV CP in the pET-14b system in *E. coli* strain BL21(DE3)pLysS and Rosetta-2(DE3)pLysS with the PLRV CP gene in pGEX-6P-2 using reducing SDS-PAGE (12%). Lane 1: BL21-GEX before induction (0 h); lane 2: BL21-GEXins before induction (0 h); lane 3: Ros2-GEXins before induction (0 h); lane 4: Rainbow molecular protein marker (Amersham); lane 5: Positive control BL21-GEX 4 h after induction; lane 6: BL21-GEXins 4 h after induction; lane 7: Ros2-GEXins 4 h after induction; lane 8: no sample; lane 9: Ros2-GEXins 4 h after induction, soluble fraction; lane 10: Ros2-GEXins 4 h after induction, insoluble fraction. Arrow indicates the produced recombinant PLRV CP-GST fusion protein. (BL21-GEX = *E. coli* strain BL21(DE3)pLysS with pGEX-6P-2 plasmid without insert; BL21-GEXins = *E. coli* strain BL21(DE3)pLysS with PLRV CP pGEX-6P-2 construct; Ros2-GEXins = *E. coli* strain Rosetta-2(DE3)pLysS with PLRV CP pGEX-6P-2 construct)

Expression of the recombinant PLRV CP fusion protein was successful. The control clone that expresses GST (Fig 12, lane 5) shows that a protein of 26 kDa is expressed which conforms to the size of GST. The size of the PLRV CP is 23 kDa, therefore the calculated size of the fusion protein is 49 kDa. However, the molecular weight of the PLRV CP fusion protein was slightly smaller (approximately 40 kDa, Fig 12, lane 6, 7, 9, 10). Expression of the protein encoded by the PLRV CP pGEX-6P-2 construct in *E. coli* strain Rosetta-2(DE3)pLysS was found to be stronger than expression of the protein encoded by the same PLRV CP pGEX-6P-2 construct in *E. coli* strain BL21(DE3)pLysS. Furthermore, the soluble fraction of induction in Rosetta-2(DE3)pLysS containing the PLRV CP pGEX-6P-2 construct resulted in a higher yield of recombinant fusion protein than the yield of the complete fraction of induction of the same PLRV CP pGEX-6P-2 construct in *E. coli* strain BL21(DE3)pLysS (Fig. 12, lane 6, 9).

To establish which growth point after induction produces the most protein, the induction was repeated and samples were taken at different growth points after induction (Fig. 13).
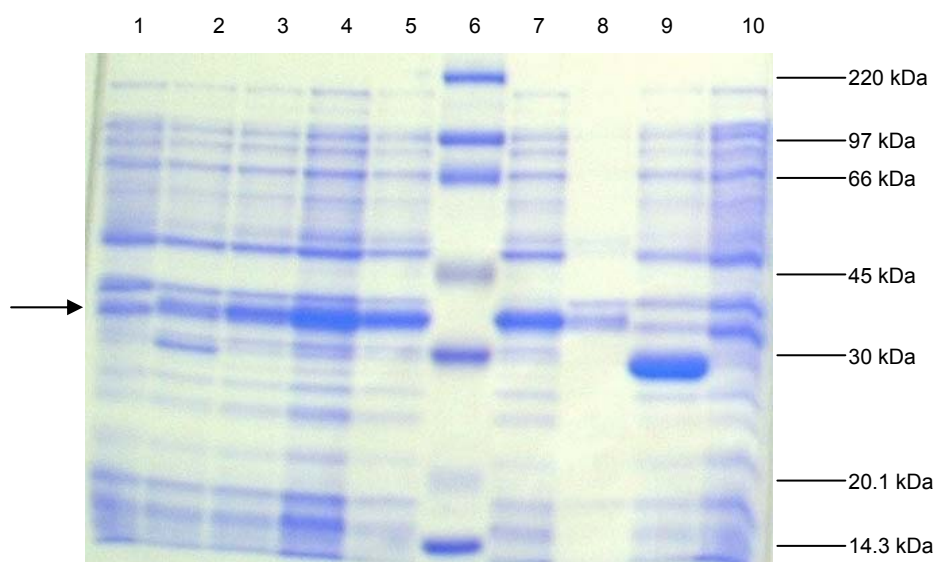
Fig. 13. Protein profiles of time course expression of the PLRV CP gene in pGEX-6P-2 by *E. coli* strain Rosetta-2(DE3)pLysS using reducing SDS-PAGE (12%). Lane 1: Ros2-GEXins before induction (0 h); lane 2: Ros-GEXins 1 h after induction; lane 3: Ros2-GEXins 2 h after induction; lane 4: Ros2-GEXins 3 h after induction; lane 5: Ros2-GEXins 4 h after induction; lane 6: Rainbow molecular protein marker (Amersham); lane 7: Ros-GEXins 4 h after induction, soluble phase; lane 8: Ros-GEXins 4 h after induction, insoluble phase; lane 9: Ros2-GEX 4 h after induction (postive control); lane 10: Ros2-GEXins 4 h cell growth without induction (negative control). Arrow indicates the produced recombinant PLRV CP-GST fusion protein. (Ros2-GEXins = *E. coli* strain Rosetta-2(DE3)pLysS with PLRV CP pGEX-6P-2 construct; Ros2-GEX = *E. coli* strain Rosetta-2(DE3)pLysS with pGEX-6P-2 plasmid without insert).

The greatest amount of recombinant protein was formed at 3 h after induction (Fig. 13, lane 4), but this does not necessarily mean that the greatest amount of soluble phase proteins were formed at 3 h. Furthermore, the sample in lane 4 might have been overloaded as the cellular proteins are also in a higher concentration than the cellular proteins in the other fractions. Fig. 13, lane 7 shows that the greatest amount of protein that was formed is found in the soluble phase. The 4 h after induction culture was used for purification of PLRV CP-GST fusion protein.

Column purification

Results from the first trial of purification of the PLRV CP-GST fusion protein using the GST bind resin column, revealed that purification was not optimized because there were several impurities in the eluates (results not shown). In the repeat experiment the eluates were re-purified on the column using a PBS-wash buffer with a 2 x higher NaCl concentration and pH as well as added Triton X-100 to decrease non-specific binding to the column. After analysis of the samples on a SDS-PAGE gel, a Bradford protein determination was performed on the samples and the protein content was less than that expected from the visual amount on the gel. According to the Bradford results of three samples the protein content was 27 µg/ml, 20 µg/ml and not enough to be detected, respectively. Therefore, the percentage basic amino acid composition of BSA used for the standard curve was compared with that of the recombinant fusion protein and it was found to be the same. Reasons for the low Bradford detection of the protein samples in comparison with the clear view on the SDS-PAGE gel could be that the basic amino acids of the

recombinant PLRV CP in the fusion protein were buried in the GST fusion protein. With another protein determination method (the BCA protein assay) the protein content was determined (after the glutathione was removed with dialysis). The method confirmed that the amount of protein in the samples was low. The reason for this low yield of protein could be two-fold: Firstly, the starting material for this purification consisted of the previous eluates and not the cell lysate or, secondly, it could have been as a result of the high viscosity of the cell lysate resulting in non-specific binding to the column. The proteins were subsequently used for only the first round of immunization. For the next repeat experiment the induced cell mixture was again used for purification and the cell lysate was briefly sonicated and filtered. This resulted in the purification of a greater amount of homogenous recombinant fusion protein (Fig. 14 (b), lane 1 to 8). The fusion protein eluted from fraction four to ten (Fig. 14 (b), lane 1-3, 5-8). However, some of the fractions were not pure enough for immunization and had to be further purified (Fig. 14, lane 1-3, 5). A Bradford protein determination was performed on the samples shown in Fig. 14 (b), lane 6, 7, and 8, in these samples the protein fractions were pure enough for immunization. The total yield of recombinant protein was 2044 µg.

Antibody production

In an attempt to produce antibodies to native PLRV, the purified recombinant fusion protein was used for immunization. Antibody production was successful as determined by a western blot of day 0, 28 and 97 antiserum (results not shown). The titre of the day 97 antiserum was very high and a dilution of 1/1000 was used to detect the recombinant antigen. The recombinant fusion protein and GST, produced by Rosetta-2(DE3)pLysS with a pGEX-6P-2 plasmid, could be detected by these antibodies (results not shown).

Protease cleavage

In an attempt to produce antibodies to native PLRV, the GST was cleaved from the recombinant fusion protein using PreScission protease for purification of recombinant PLRV CP (Fig. 15).

Cleavage of the recombinant fusion protein was successful, but the cleaved PLRV CP remained on the column and did not elute in the expected fractions as shown in Fig. 15. The cleaved PLRV CP only eluted with glutathione when the GST also eluted and was therefore contaminated with GST (Fig. 16, lane 8). A standard curve revealed that the size of the bigger band in Fig. 16, lane 8 (top arrow) is 26.1 kDa, which agrees with that of GST (26 kDa). The smaller band below the cleaved GST (Fig. 16, lane 8, lower arrow), which could be PLRV CP (23 kDa), has a molecular mass of 25.2 kDa.
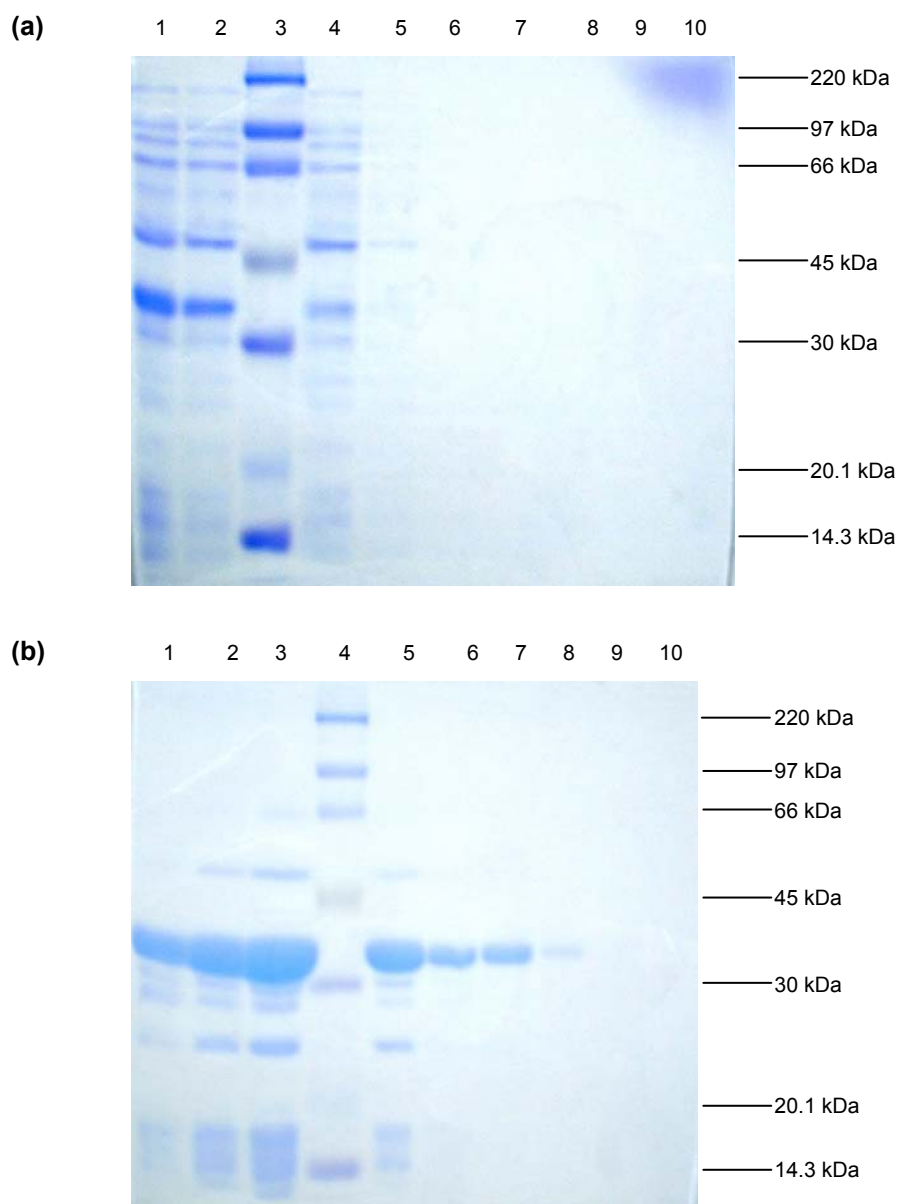
Fig. 14. Protein profiles of GST column purification of recombinant PLRV CP-GST fusion protein affinity purified on GST column using reducing SDS-PAGE (12%). (a) Lane 1: Ros2-GEXins 4 h after induction, soluble phase cell lysate; lane 2: Ros2-GEXins 4 h after induction, soluble phase cell lysate after 0.45 μM filtering; lane 3: Rainbow molecular protein marker (Amersham); lane 4: Ros2-GEXins 4 h after induction, soluble phase cell lysate after 0.45 μM filtering and once passed through the column; lane 5: first wash step eluate; lane 6: second wash step eluate; lane 7: third wash step eluate; lane 8: first 2 ml glutathione elution; lane 9: second 2 ml glutathione elution; lane 10: third 2 ml glutathione elution. **(b)** Lane 1: fourth 2 ml glutathione elution; lane 2: fifth 2 ml glutathione elution; lane 3: sixth 2 ml glutathione elution; lane 4: Rainbow molecular protein marker (Amersham); lane 5: seventh 2 ml glutathione elution; lane 6: eight 2 ml glutathione elution; lane 7: ninth 2 ml glutathione elution; lane 8: tenth 2 ml glutathione elution; lane 9: eluate of column wash buffer 1; lane 10: eluate of column wash buffer 2. (Ros2-GEXins = *E. coli* strain Rosetta-2(DE3)pLysS with PLRV CP pGEX-6P-2 construct; wash = column wash with PBS buffer).
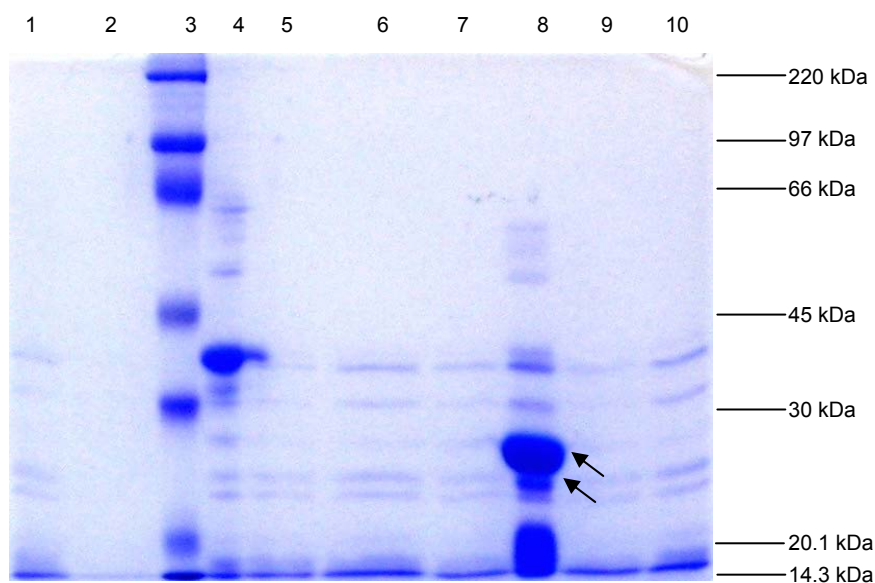
Fig. 15. Protein profiles of the cleavage of recombinant PLRV CP-GST fusion with PreScission protease using reducing SDS-PAGE (14%). The eluate of the column was collected in one or two ml fractions. Lane 3: Rainbow molecular protein marker (Amersham); lane 4: purified recombinant fusion protein (uncleaved); Eluation without glutathione: lane 10; Fraction 3 (1 ml); lane 6: Fraction 4 (1 ml); lane 7: Fraction 5 (1 ml); lane 9: Fraction 6 (1 ml); lane 1: Fraction 7 (1 ml); lane 5: Fraction 8 (1 ml); lane 2: Fraction 26 (2 ml sample); after glutathione elution: lane 8: Fraction 3 (2 ml). Approximately 1 μg of each sample was loaded onto the gel, except for the glutathione eluted protein in lane 8 where 10 μl of the fraction was loaded onto the gel. Arrows indicate the two proteins formed as a result of protease cleavage of PLRV CP-GST fusion protein.

Western blot

The purified PLRV CP fusion protein was used for immunization. This was done to test whether antibody formation against fusion protein would induce high levels of antibodies that would recognize the CP, which in turn could be used to detect native PLRV. Additionally, an attempt was made to cleave the GST from the PLRV CP fusion protein in order to produce antibodies to the cleaved PLRV CP which in turn would recognize native PLRV. Western blots were performed with antibodies to PLRV, recombinant fusion protein and GST to establish whether cleavage of the recombinant GST-PLRV CP fusion protein was successful and whether the recombinant PLRV CP was similar in conformation to the native PLRV CP (Fig. 16-18). The results of the western blot analysis performed using the anti-PLRV antibodies are shown in Fig. 16. This blot was performed with the same antibodies prepared against native PLRV, which could not detect the protein that was produced using the pET14-b system (Fig. 11).

This western blot showed that the commercial anti-PLRV antibodies detected a protein with a size of 24.8 kDa (Fig. 16, lane 1, which is approximately the size of PLRV CP of 23 kDa) and a protein with a size of 36.5 kDa (Fig. 16, lane 3, which is approximately the size of the fusion protein of 49 kDa) confirming that these antibodies detected the cleaved PLRV CP and the GST-PLRV CP fusion protein. From these results it can be deduced that commercial antibodies against PLRV bound weakly to both the GST-PLRV CP fusion protein and the cleaved PLRV CP portion of the fusion protein.
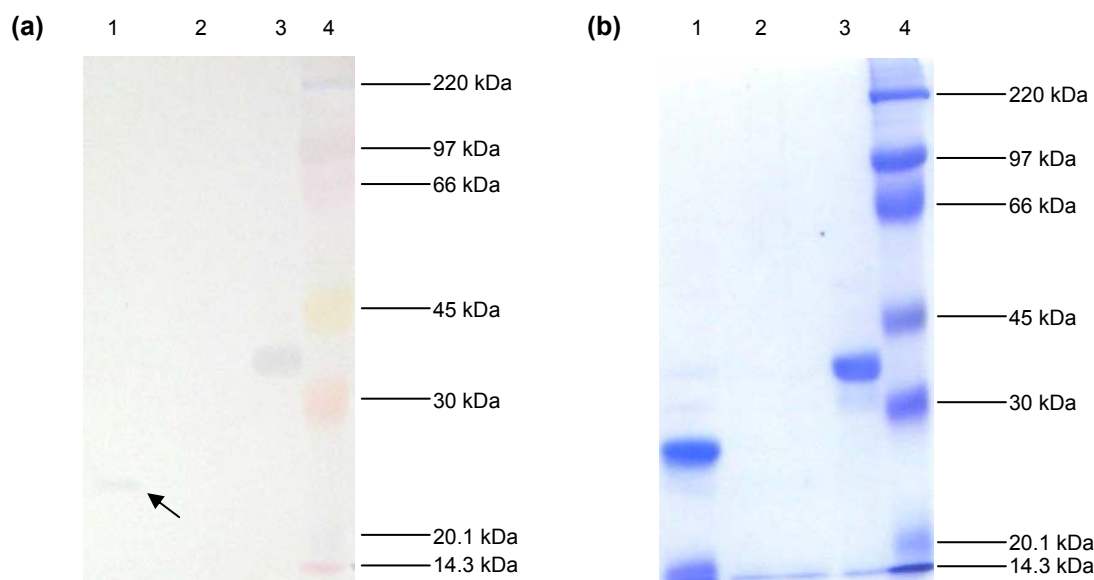
Fig. 16. (a) Western-blot and (b) protein profiles using reducing SDS-PAGE (14%) of protease cleavage of recombinant fusion protein by anti-PLRV antibodies. Anti-PLRV antibodies were used for the detection of proteins (the same antibodies as those used for the western blot presented in Fig. 12). For **(a)** and **(b)**: lane 1: glutathione elution fraction; lane 2: cleaved recombinant fusion protein eluate fraction from GST column (approximately 1 μg); lane 3: purified recombinant fusion protein (uncleaved, approximately 1 μg); lane 4: Rainbow molecular protein marker (Amersham). Samples used are the same as those used for electrophoresis presented in Fig. 15. Arrow indicates the band detected by anti-PLRV antibodies of an approximate size of 24.8 kDa (approximately the size of PLRV CP, 23 kDa).

The results of the western blot analysis performed using the anti-recombinant fusion protein antibodies produced in the present study are shown in Fig. 17.

This western blot showed that the anti-GST-PLRV CP antibodies detected two proteins (Fig. 17, lane 2), one with a size of 23.0 kDa (which is the size of PLRV CP) and another with a size of 26.6 kDa (which is approximately the size of GST of 26 kDa). The anti-GST-PLRV CP antibodies also detected a protein with a size of 40 kDa (Fig. 17, lane 4, which is approximately the size of GST fusion protein of 49 kDa). Taken together, these results confirm that the antibodies detected the cleaved GST and the GST-PLRV CP fusion protein strongly. PLRV CP was apparently also detected weakly (Fig. 17, lane 2) although other bands were also detected indicating some non-specific binding. From these results it can be deduced that antibodies produced against the GST-PLRV CP fusion protein bound strongly to the GST protein and poorly to the PLRV CP portion of the fusion protein.

The results of the western blot analysis performed using commercial anti-GST antibodies are shown in Fig. 18. This western blot showed that the commercial anti-GST antibodies detected a protein with a size of 27 kDa (which is approximately the size of GST of 26 kDa) and a protein with a size of 37.6 kDa (which is approximately the size of GST fusion protein of 49 kDa) confirming that these antibodies detected the cleaved GST (Fig. 18, lane 2) and the GST-PLRV CP fusion protein (Fig. 18, lane 4). Other bands were also detected indicating some non-specific binding. From these results it can be deduced that antibodies produced against the

GST protein strongly react with the recombinant fusion protein and the cleaved GST, but not with the cleaved PLRV CP.
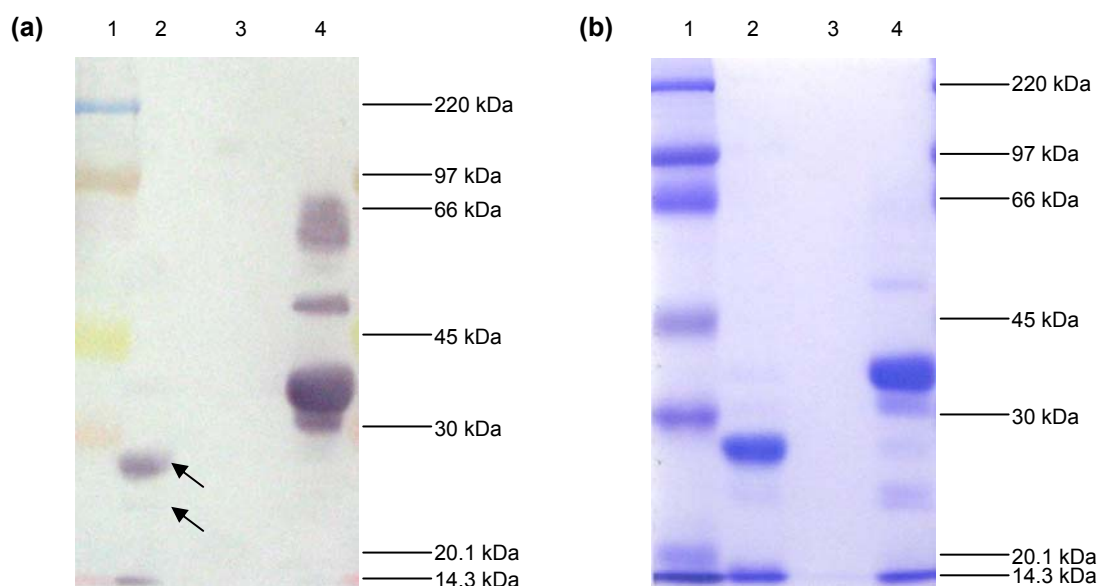
Fig. 17. (a) Western-blot and (b) protein profiles using reducing SDS-PAGE (14%) of protease cleavage of recombinant fusion protein by anti-recombinant fusion protein antibodies. For **(a)** and **(b)**: lane 1: Rainbow molecular protein marker (Amersham); lane 2: glutathione elution fraction; lane 3: cleaved recombinant fusion protein eluate fraction from GST column (approximately 1 µg); lane 4: purified recombinant fusion protein (uncleaved, approximately 1 µg). Arrows indicate the bands detected by anti-recombinant fusion protein antibodies approximately the size of 23.0 kDa (bottom arrow) and 26.6 kDa (top arrow).
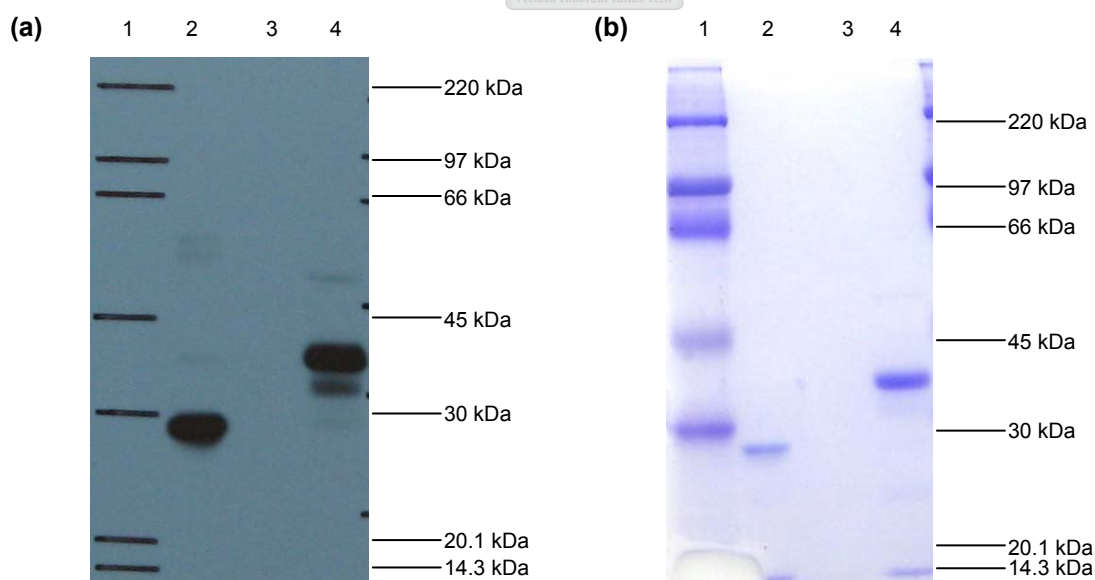
Fig. 18. (a) Western-blot and (b) protein profiles using reducing SDS-PAGE (14%) of protease cleavage of recombinant fusion protein by anti-GST antibodies. For **(a)** and **(b)**: lane 1: Rainbow molecular protein marker (Amersham); lane 2: glutathione elution fraction; lane 3: cleaved recombinant fusion protein eluate fraction from GST column (approximately 0.1 µg); lane 4: purified recombinant GST-PLRV CP fusion protein (uncleaved, approximately 0.1 µg).

Conclusions

The PLRV CP gene was successfully amplified with primers that were designed to recognize the 5' and 3' ends of the gene. These primers were extended to include suitable restriction sites for subcloning into expression vectors. As the *Nde*1 restriction enzyme resulted in poor cleavage. Only the restriction enzyme combination of *Nco*1 and *Xho*1 could be used for subcloning into the pET14-b plasmid. Additionally, the PLRV CP was subcloned into the pGEX plasmid using *Eco*R1 and *Xho*1. Successful subcloning of full length PLRV CP into both plasmids could be confirmed by sequencing. The cloning and expression of recombinant PLRV CP has also been successfully performed by other authors (López *et al.*, 1994; Treder and Lewosz, 2001).

Expression of recombinant PLRV CP encoded by the PLRV CP pET14-b construct in *E. coli* strain BL21(DE3)pLysS was not successful. Other authors also found the expression of recombinant PLRV CP to be problematic (López *et al.*, 1994) and they deduced that the reason for this poor expression was the presence of arginine codons in the N-terminal region of PLRV CP which are rarely used in *E. coli*. Sequence analysis (see addendum C of this thesis) revealed that a total of 11 arginine residues encoded by codons that are rarely used in *E. coli* are found within the first 60 codons on the N-terminal side of PLRV CP. For this reason, expression of recombinant PLRV CP was attempted in the *E. coli* strain Rosetta-gami B(DE3)pLysS strain as it supplies tRNAs for the codons AGA (Arg), AGG (Arg), AUA (Ile), CUA (Leu), CCC (Pro) and GGA (Gly) on a compatible chloroamphenicol-resistant plasmid, pRARE. However, even though two recombinant proteins were produced, neither was of the correct size. Another strain, the *E. coli* strain Rosetta-2(DE3)pLysS, that supplies an additional tRNA for the rare codon for arginine (*E. coli* strain Rosetta-2(DE3)pLysS contains a compatible chloroamphenicol-resistant plasmid, pRARE2, coding for the codons AGA (Arg), AGG (Arg), CGG (Arg), AUA (Ile), CUA (Leu), CCC (Pro) and GGA (Gly)), and is easier to cultivate, was also used, but in this strain only the smaller protein was produced. It was established that the protein was indeed a result of induction of the PLRV CP pET14-b construct and not an endogenous protein. The induced protein could be purified with ammonium sulphate precipitation to near homogeneity. However, it was concluded that this protein was not full length PLRV CP as antibodies against native PLRV could not detect it. Thus, in spite of this choice of strains we were unable to produce full length recombinant PLRV CP. It could be that the recombinant protein that was formed is a truncated form of PLRV CP that never assumes its correct conformation. Even though the N-terminal region of the protein would have been formed, and this region is antigenically the most important (Prof M.H.V. van Regenmortel, Strasbourg, personal communication; chapter 3 of this thesis), it could be that the truncated protein folds very differently in comparison to the full length PLRV CP and antibodies against native PLRV CP therefore did not recognize it.

To overcome this problem, expression was attempted in the pGEX expression system. For this purpose, the PLRV CP sequence was cloned 3' to the GST gene in the pGEX plasmid. When expressed, this would result in a fusion protein in which GST was on the N-terminal side relative to the PLRV CP. López *et al.* (1994) cloned the PLRV CP gene to the 3' end of the β-galactosidase gene to give a fusion construct. Expression of this construct resulted in the production of a β-galactosidase PLRV CP fusion protein. The reason given for the success of this expression was that the fusion solves the codon usage problem because it moves the arginine rich region closer to the carboxy terminus of the fusion protein, as a result of which translation could be less hindered at the sensitive initiation site by the rare codons (Chen and Inouye, 1990). By increasing the distance between the initiation codon and the arginine codons the production of recombinant protein can be increased almost linearly up to eight fold (Chen and Inouye, 1990). Therefore, it is not only the presence of the rare codons but also the position of the codons in an mRNA that determines the effectiveness of expression of recombinant proteins with rare codons in *E. coli*. Possibly for this reason, our approach for the production of the recombinant GST-PLRV CP fusion protein in *E. coli* strains BL21(DE3)pLysS and Rosetta-2(DE3)pLysS was therefore also successful. However, expression in Rosetta-2(DE3)pLysS was stronger, which shows that not only the addition of an N-terminal fusion protein but also the addition of tRNAs for rare codons in the expression strain improves expression. An apparent anomaly was that the recombinant GST-PLRV CP fusion protein had an apparent molecular weight of 40 kDa whilst the theoretical molecular weight was 49 kDa. López *et al.* (1994), found that the recombinant β-galactosidase PLRV CP fusion protein had an apparent molecular weight of 116 kDa as opposed to a theoretical molecular weight of 139 kDa. A possible explanation for this phenomenon could be that SDS-PAGE gels do not show the proteins according to their estimated size (Rubinson *et al.*, 1997). In this case, the recombinant fusion protein migrated faster than expected, whereas Rubinson *et al.* (1997) found that their recombinant GVA CP migrated slower than the size deduced from the nucleotide sequences. Martin *et al.* (1990) noted that a purified viral protein migrated to an observed size smaller than would be expected, which was attributed to anomalous running of the polypeptide in the gel due to its conformation (there was a proline-rich tract in the protein) or to posttranslational modification or degradation during virus purification. Therefore, native viral proteins and recombinant viral proteins can migrate to positions that are not expected. For the recombinant PLRV CP or recombinant GST-PLRV CP fusion protein, the proteins might migrate faster as a result of the charge contribution of the numerous arginines. Mayo *et al.* (1989) reported that the net charge of PLRV CP was +24 calculated from the difference between the total number of lysine and argininine residues and the total number of aspartic and glutamic acid residues. This feature is shared with other luteovirus CPs as well as those of other unrelated viruses (Mayo *et al.*, 1989).

Improved release of the recombinant GST-PLRV CP fusion protein from the cell lysate into the soluble phase was achieved by sonification, followed by centrifugation and filtration. For the purification of the recombinant fusion protein a standard GST-binding resin was used and this was only successful when the composition of the PBS wash buffer was altered. An attempt was made to cleave the GST from the recombinant fusion protein to purify PLRV CP. Although cleavage of the fusion protein was successful, separation of the GST protein from the PLRV CP was not successful as the cleaved PLRV CP eluted together with the GST after release with glutathione.

In western blots in which antibodies were used against PLRV, the recombinant GST-PLRV CP fusion protein and GST respectively, it was found that antibodies made against the fusion protein were found to detect the PLRV CP poorly. This is in contrast to the production of antibodies to a fusion protein of GST and TSWV CP (Vaira *et al.*, 1996) which detected the native TSWV CP successfully, and could then be used in ELISA. Meng *et al.* (2003) produced the CP of RSPaV fused to the N-terminus of MBP and there antibodies could detect native RSPaV accurately in all tissues tested. However, use of these antibodies in a double antibody sandwich (DAS)-ELISA (coating with antibody) was not successful and only an indirect ELISA was effective (coating with antigen). Čeřovská *et al.* (2002) produced antibodies to PVA without a fusion and these were effective for the detection of PVA CP. These antibodies were, however, not suitable for use in a DAS-ELISA, but were effective in the indirect-ELISA. The CP of CTV fused to the N-terminal fragment of MBP to produce antibodies that were effective for the detection of CTV infected tissue in a western blot (Nikolaeva *et al.* 1995). Antibodies produced to 96% of the CP of GLRaV fused at the C-terminal portion of β-galactosidase were effective for the production of antibodies that could detect the native virus but gave a stronger reaction with the recombinant viral CP-β-galactosidase fusion protein (Ling *et al.*, 2000). However, even though this work shows that PLRV CP can be successfully produced as a fusion protein with GST, we found that our antibodies to recombinant GST-PLRV CP fusion protein were not effective in detecting the native virus in a DAS ELISA (results not shown). GST may be more immunogenic than PLRV CP, which may result in poorer antibody responses to the PLRV CP portion of the recombinant GST-PLRV CP fusion protein. For this reason, antibody production should only be considered against purified PLRV CP. Thus, in future work, purification of the cleaved PLRV CP prior to the production of antibodies against PLRV CP, will be essential.

References

Abou-Jawdah, Y., Sobh, H., Cordahi, N., Kawtharani, H., Nemer, G., Maxwell, D.P. and Nakhla, M.K. (2004). Immunodiagnosis of prune dwarf virus using antiserum produced to its recombinant coat protein. *J. Virol. Methods*, **121**: 31-38.

Anon. (1994). The pET system: your choice for expression. *inNovagen*, **1**(1): 1-3.

Anon. (2002). GST gene fusion system handbook. *Amersham Biosciences*, Edition AA: 18-1157-58.

Anon. (2006a). Novagen competent cells: what a difference a strain makes. *EMB Biosciences*, pp. 1-12. Available: www.emdbiosciences.com/novagen (Accessed on 17 January 2007).

Anon. (2006b). pET hosts strain descriptions. Available: http://www.merckbiosciences.co.uk/SharedImages/Novagen/Hoststrains.html (Accessed on 12 October 2006).

Bellstedt, D.U., Human, P.A., Rowland, G.F. and Van der Merwe, K.J. (1987). Acid-treated, naked bacteria as immune carriers for protein antigens. *J. Immunol. Methods*, **98**: 249-255.

Bradford, M.M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein using the principle of protein-dye binding. *Anal. Biochem.*, **72**: 248-254.

Čeřovská, N., Moravec, T. and Velemínský, J. (2002). Polyclonal antibodies to a recombinant coat protein of potato virus A. *Acta Virol.*, **46**: 147-151.

Chen, G.T. and Inouye, M. (1990). Suppression of the negative effect of minor arginine codons on gene expression; preferential usage of minor codons within the first 25 codons of the *Esherichia coli* genes. *Nucleic Acids Res.*, **18**(6): 1465-1473.

Hourani, H. and Abou-Jawdah, Y. (2003). Immunodiagnosis of cucurbit yellow stunting disorder virus using polyclonal antibodies developed against recombinant coat protein. *J. Plant Pathol.*, **85**(3): 197-204.

Jelkmann, W. and Keim-Konrad, R. (1997). Immuno-capture polymerase chain reaction and plate-trapped ELISA for the detection of apple stem pitting virus. *J. Phytopathol.*, **145**: 499-503.

Kaden-Kreuziger, D., Lamprecht, S., Martin, R.R. and Jelkman, W. (1995). Immunocapture polymerase chain reaction assay and ELISA for the detection of strawberry mild yellow edge associated potexvirus. *Acta Hortic.*, **385**: 33-38.

Kelly, K.C., Huestis, K.J., Austen, D.A., Sanderson, C.T., Donoghue, M.A., Stickel, S.K., Kawasaki, E.S. and Osburne, M.S. (1995). Regulation of *sCD4-183* gene expression from phage-T7-based vectors in *Esherichia coli. Gene*, **156**: 33-36.

Kumari, S.G., Makkouk, K.M., Katul, L. and Vetten, H.J. (2001). Polyclonal antibodies to the baterially expressed coat protein of faba bean nectrotic yellows virus. *J. Phytopathol.*, **149**: 543-550.

Laemmli, U.K. (1970). Cleavage of structural proteins during the assemly of the head of bacteriophage T4. *Nature*, **227**: 680-685.

Ling, K., Zhu, H., Jaing, Z. and Gonsalves, D. (2000). Effective application of DAS-ELISA for detection of grapevine leafroll associated closterovirus-3 using a polyclonal antiserum developed from recombinant coat protein. *Eur. J. Plant Pathol.*, **106**: 301-309.

López, L., Muller, R., Balmori, E., De La Riva, G., Ramírez, N., Doreste, V., López, M., Peréz, S., Oramas, P. and Selman-Housein, G. (1994). Molecular cloning and nucleotide sequence of the coat protein gene of a Cuban isolate of potato leafroll virus and its expression in *Escherichia coli. Virus Genes*, **9**(1): 77-83.

Martin, R.R., Keese, P.K., Young, M.J., Waterhouse, P.M. and Gerlach, W.L. (1990). Evolution and molecular biology of luteoviruses. *Annu. Rev. Phytopathol.*, **28**: 341-363.

Mayo, M.A., Robinson, D.J., Jolly, C.A. and Hyman, L. (1989). Nucleotide sequence of potato leafroll luteovirus RNA. *J. Gen. Virol.*, **70**: 1037-1051.

Meng, B., Credi, R., Petrovic, N., Tomazic, I. and Gonsalves, D. (2003). Antiserum to recombinant virus coat protein detects rupestris stem pitting associated virus in grapevines. *Plant Dis.*, **87**(5): 515-522.

Minafra, A., Casati, P., Elicio, V., Rowhani, A., Saldarelli, P., Savino, V. and Martelli, G.P. (2000). Serological detection of grapevine rupestris stem pitting-associated virus (GRSPaV) by a polyclonal antiserum to recombinant virus coat protein. *Vitis*, **39**(3): 115-118.

Nikolaeva, O.V., Karasev, A.V., Gumph, D.J., Lee, R.F. and Garnsey, S.M. (1995). Production of polyclonal antisera to the coat protein of citrus tristeza virus expressed in *Escherichia coli*: application for immunodiagnosis. *Phytopathology*, **85**(6): 691-694.

Rohwani, A. and Stace-Smith, R. (1979). Purification and characterization of potato leafroll virus. *Virology*, **98**: 45-54.

Rubinson, E., Galiakparov, N., Radian, S., Sela, I., Tanne, E. and Gafny, R. (1997). Serological detection of grapevine virus A using antiserum to a nonstructural protein, the putative movement protein. *Phytopathology*, **87**(10): 1041-1045.

Studier, W., Rosenberg, A.H., Dunn, J.J. and Dubendorff, J.W. (1990). Use of T7 RNA polymerase to direct expression of cloned genes. *Method. Enzymol.*, **185**: 60-89.

Tamada, T. and Harrison, B.D. (1980). Factors affecting the detection of potato leafroll virus in potato foliage by enzyme-linked immunosorbent assay. *Ann. Appl. Biol.*, **95**: 209-219.

Treder, K. and Lewosz, J. (2001). Production of antibodies against potato leafroll virus coat protein expressed in *Escherichia coli* as an antigen. *Proceedings of 'The 11th EAPR Virology Section Meeting', Triest, Czech Republic*, pp. 62-64.

Vaira, A.M., Vecchiati, M., Masenga, V. and Accotto, G.P. (1996). A polyclonal antiserum against a recombinant viral protein combines specificity with versatility. *J. Virol. Methods*, **56**: 209-219.
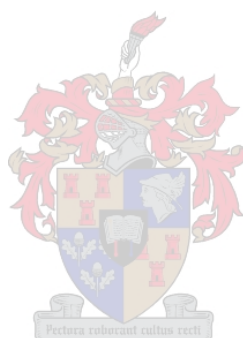
Chapter 5: Future prospectives

In this study diversified and different PLRV isolates in South Africa were identified, which indicates that there is an ongoing threat that PLRV isolates with altered and also more severe pathogenicity may evolve. Thus, the sequencing of CP genes of South African PLRV isolates on an ongoing basis is important to monitor the evolution of PLRV in this country. Furthermore, the introduction of new PLRV isolates into South Africa via newly imported potato material should be limited as new PLRV introductions could form new and more pathogenic recombinants with the already unique isolates present in South Africa. If this were to happen, the severity of PLRV infection in potatoes in South Africa might increase.

For the production of recombinant PLRV CP we have shown that this is only possible when the PLRV CP gene is expressed as a fusion protein with GST. Furthermore, the production of recombinant GST-PLRV CP fusion protein is greater in *E. coli* strains that supplement rare codons. However, for the production of antibodies to PLRV CP, the recombinant GST-PLRV CP was not effective. The results obtained indicate that the cleaved recombinant PLRV CP might be effective for the production of antibodies that recognize native PLRV CP. As the cleavage of the recombinant GST-PLRV CP fusion protein was successful, but not the subsequent purification, the purification of the cleaved GST from PLRV CP should be further optimized. The purified cleaved PLRV CP can then be used for the production of antibodies that would detect PLRV in ELISAs.

Other than producing antibodies for the detection of PLRV to the purified cleaved PLRV CP, antibodies could also be produced to antigenic regions of PLRV CP. Two very immunogenic regions found at the N-terminus of PLRV CP were identified in this study. These regions are more important antigenically than other regions of PLRV CP since they are situated in the N-terminal immunogenically important region of viral CPs. Peptides corresponding to these regions could be synthetically produced and these used for the immunization of rabbits to produce of anti-peptide antibodies. The use of this method would circumvent many problems that are associated with the production of recombinant proteins in *E. coli* as have been found in this study as well as the labour intensive process of cloning and subcloning. These anti-peptide antibodies could then be used for an effective diagnostic assay for PLRV.

Addendum A

Alignment of the sequenced South African PLRV CP nucleotide sequences with PLRV CP isolates from GenBank. Alignment of the nucleotide sequences was performed but a few gaps had to be introduced into the outgroup sequence (SPLSV). The length of the aligned matrix is 631.

```
                 10        20        30        40        50        60        70
        ....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Rood     ATGAGTACGGTCGTGGTTAGAGGAAATGTCAATGGTGGTATACAACAACCAAAGAGGCGAAGAAGGCAAT
NN3-1    ...G.................A...................G............................
NN3-2    ...G.................A...................G............................
Sand2-1  ...G.................A...................GG...........................
Sand2-2  ...G.................A...................G............................
NN334-7  ...G.................A...................G............................
Sand2-3  ...G.................A...................G............................
SandB-2  ...G.................A...................G......G.....T...............
SandB-3  ...G.................A...................G............................
PP003-11 ----................A...................G............................
WW154-20 ...G.................A...................G............................
CC5-1    ...G.................A...................G............................
CC5-3    ...G.................A...................G............................
CC6-7    ...G.................A...................G............................
TT-12    ...G.................A...................G............................
NALR7    ....................A...................G..........GA................
Aus      ....................A...................GC.........GA................
Sand1-1  ...G.................A..................G..........GA................
SandA-1  ...G.................A..................G..........GA................
SandA-3  ...G.................A..................G..........GA................
SandF-1  ...G.................A..................G...........A................
SandF-2  ...G.................A..................G..........GA................
SandF-3  ...G.................A..................G...........A................
PP0-6    ...G.................A...................G.........GA................
Z25-14   ...G.................A...................G.........GA................
DD020-1  ...G...T............A...................G.........GA................
DD020-2  ...G.................A...................G.........GA................
DD020-3  ...G.................A...................G.........GA................
NethV    ....................A...................G.T.......GA................
Neth31   ....................A...................G.........GA................
NethV4   ....................A...................G.........GA................
Neth11   ....................A...................G........G.GA................
Neth30   ....................A...................G.........GA................
Pak      ....................A...................G.........GA................
OP       ....................A...................G.........GA................
SKorRB   ....................A...................G.........GA................
NethWa   ....................A............C...G.........GA................
Cuba     ....................A...................G.G.......GA................
India    ....................A...................G.......A..GA................
Italy    ....................A...................G.......A..GA................
14.2     ....................A...................GC...C.....GA................
FrPT     ....................A...................GC...C.....GA................
Fr1      ....................AG..................G.........GA................
Zim13    ....................AG..................G.........GA................
Egp      ....................A...................G.........GA................
Pol      ....................A...................G.........GA................
SKor777  ....................A...................G.....T...GA................
Can      ....................A...................G.........GA................
SKor     ........C...........A...................G.........GA................
Neth15   ....................A...................G.........T.GA.T.............
Sco      ....................A...................G.........GA.T.............
SandH-7  ...G.................A...................G......G....GA................
SandH-10 ...G.................A...................G.........GA................
Z26-23   ...G.................A...................G.........GA................
Z26-24   ...G.................A...................G.........GA................
Z26-25   ...G.................A...................G.....G...GA................
CIP01    ....................A...................G.........G.................
Ger      ....................A...................G.........G.................
Noir     ....................A...................G.........GA................
Chi      ..............C.....................G.........GA................
Cu87     ....................A...................GC........GA..............C
FrCU87   ....................A.A.C...........G.......C....GA..............C
SPSLV    ....................AA-------......CTC.CGT--.....---GC.CA.........AG.A
```

```
                        80        90       100       110       120       130       140
                ....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Rood            CCCTTCGAAGGCGCGCTAACAGAGTGCAGCCAGTGGTTATGGTCACGGCCCCTGGGCAACCCAGGCGCCG
NN3-1           .....................................................................
NN3-2           .....................................................................
Sand2-1         .....................................................................
Sand2-2         .....................................................................
NN334-7         .....................................................................
Sand2-3         .....................................................................
SandB-2         ...........................................................T.........
SandB-3         .....................................................................
PP003-11        .....................................................................
WW154-20        .....................................................................
CC5-1           .....................................................................
CC5-3           .....................................................................
CC6-7           .....................................................................
TT-12           .....................................................................
NALR7           ......GCC................A............................................
Aus             .......C.................A............................................
Sand1-1         .......C.................A............................................
SandA-1         .......C.................A............................................
SandA-3         .......C.................A............................................
SandF-1         .......C.................A............................................
SandF-2         .......C.................A............................................
SandF-3         .......C.................A............................................
PP0-6           .......C.................A............................................
Z25-14          .......C.................A............................................
DD020-1         .......C.................T............................................
DD020-2         .......C.................T............................................
DD020-3         .......C.................T............................................
NethV           .......C.................C............................................
Neth31          .......C.................C............................................
NethV4          .......C.................C............................................
Neth11          .......C.................T............................................
Neth30          .......C.................T............................................
Pak             .......C.................T............................................
OP              .......C.................T..........................T.................
SKorRB          .......C.................T............................................
NethWa          .......C.................T............................................
Cuba            .......C.................T............................................
India           .......C.................T............................................
Italy           .......C.................T............................................
14.2            .......C.................T............................................
FrPT            .......C.................T............................................
Fr1             .......C.....T........CT..............................................
Zim13           .......C.....T........T...............................................
Egp             .......C.................T...G........................................T.......
Pol             .......C.................T..G.........................................T.......
SKor777         .......C.................T............................................
Can             .......C.................T............................................
SKor            .......C.................T........................G...................
Neth15          .......C.................T............................................
Sco             .......C.................T............................................
SandH-7         .......C.................T............................................
SandH-10        .......C.................T............................................
Z26-23          .......C.................T..........................................T.
Z26-24          .......C.................T..........................................T.
Z26-25          .......C.................T..........................................T.
CIP01           .......T.....T........T.............................................T.
Ger             .......C.................T.........................T..................
Noir            .......C.................T............................................
Chi             .......C.A...............T.....................A......................
Cu87            .......C.................T............................................
FrCU87          .......C.................T............................................
SPSLV           A..---AG......T...CA....GT...........G......A......C..A.CA..G..G....
```
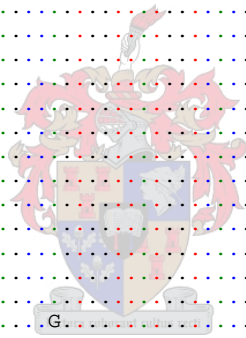
```
                 150       160       170       180       190       200       210
           ....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Rood       AAGACGCAGAAGAGGAGGCAATCGCCGCTCAAGAAGAACTGGAGTTCCCCGAGGACGAGGCTCAAGCGAG
NN3-1      ..................................................................
NN3-2      ..................................................................
Sand2-1    ..................................................................
Sand2-2    ..................................................................
NN334-7    ..................................................................
Sand2-3    ..................................................................
SandB-2    ..................................................................
SandB-3    ..................................................................
PP003-11   ..................................................................
WW154-20   ..................................................................
CC5-1      .G................................................................
CC5-3      ..................................................................
CC6-7      ............G.....................................................
TT-12      ..................................................................
NALR7      ..................................A...............................
Aus        ......T.............C.............................................
Sand1-1    ..................................................................
SandA-1    ...G..............................................................
SandA-3    ..................................................................
SandF-1    ..................................................................
SandF-2    ..................................................................
SandF-3    .....................................G............................
PP0-6      ..................................................................
Z25-14     .........G........................................................
DD020-1    ..................................................................
DD020-2    ..................................................................
DD020-3    ..................................................................
NethV      ..................................................................
Neth31     ..................................................................
NethV4     ..................................................................
Neth11     ..................................................................
Neth30     ..................................................................
Pak        ..................................................................
OP         ..................................................................
SKorRB     ..................................................................
NethWa     ..................................................................
Cuba       ..................................................................
India      ..................................................................
Italy      ..................................................................
14.2       ..................................................................
FrPT       ..................................................................
Fr1        ..................................................................
Zim13      ..................................................................
Egp        ..................................................................
Pol        ..................................................................
SKor777    ..................................................................
Can        ..................................................................
SKor       ..................................................................
Neth15     ..................................................................
Sco        ..................................................................
SandH-7    ..................................................................
SandH-10   ..................................................................
Z26-23     ..................................................................
Z26-24     ..................................................................
Z26-25     ..................................................................
CIP01      ......T...........................................................
Ger        ......T...........................................................
Noir       ......T...........................................................
Chi        ......T...........................................................
Cu87       ......T...........................................................
FrCU87     ......T....................................T.....................
SPSLV      .C....TC.....C....TGG.A.AACAA.TC..C..G...C.C.....G........A..AT.C..C.A
```

```
                    220       230       240       250       260       270       280
                ....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Rood            ACATTCGTGTTTACAAAGGACAACCTCGTGGGCAATTCCCAAGGAAGTTTCACCTTCGGGCCGAGTCTAT
NN3-1           ..................................................................
NN3-2           ..................................................................
Sand2-1         ..................................................................
Sand2-2         ..................................................................
NN334-7         ................G.................................................
Sand2-3         ..................G...............................................
SandB-2         ..................................................................
SandB-3         ..................................................................
PP003-11        ..................................................................
WW154-20        ..................................................................
CC5-1           ..................................................................
CC5-3           ...........................C......................................
CC6-7           ..................................................................
TT-12           ..................................................................
NALR7           ...........................................C......................
Aus             ...........................................C......................
Sand1-1         ...........................................C......................
SandA-1         ...........................................C......................
SandA-3         ...........................................C......................
SandF-1         ...........................................C......................
SandF-2         ...........................................C......................
SandF-3         ...........................................C......................
PP0-6           ...........................................C......................
Z25-14          ..C........................................C......................
DD020-1         ...........................................C......................
DD020-2         ...........................................C......................
DD020-3         ...........................................C......................
NethV           ...................................A.......C......................
Neth31          ...................................A.......C......................
NethV4          ...................................A.......C......................
Neth11          ...................................A.......C......................
Neth30          ...................................A.......C......................
Pak             ...................................A.......C......................
OP              ...................................A.......C......................
SKorRB          ...................................A.......C......................
NethWa          ...................................A.......C......................
Cuba            ..........................................C......................
India           ..........................................C......................
Italy           ...................................A.......C......................
14.2            ...................................A.......C......................
FrPT            ...................................A.......C......................
Fr1             ...................................A.......C......................
Zim13           ...................................A.......C......................
Egp             ...................................A.......C......................
Pol             ...................................A.......C......................
SKor777         ..........................................GC.....................
Can             ...........................................C......................
SKor            ..........................................CA.....................
Neth15          ..........................................CA.....................
Sco             ..........................................CA.....................
SandH-7         ...................................A.......C......................
SandH-10        ...................................A.......C......................
Z26-23          ...................................A.......C.............G........
Z26-24          ...................................A.......C......................
Z26-25          ...................................A.......C......................
CIP01           ...................................A.......C......................
Ger             ...................................A.......C......................
Noir            ...................................A.......C......................
Chi             ...................................A.......C......................
Cu87            ...................................A.......C......................
FrCU87          ...........................................C...............A......
SPSLV           ..T.....A..CT................AA.....G.G..AC...............TC...T.
```

```
                    290       300       310       320       330       340       350
               ....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Rood           CAGACTGTCCGGCATTCAAGGATGGAATACTCAAGGCCTACCATGAGTATAAGATCACAAGCATCTTACT
NN3-1          ...................A..........................................T.......
NN3-2          ...................A..........................................T.......
Sand2-1        ..............................................................T.......
Sand2-2        ..............................................................T.......
NN334-7        ..............................................................T.......
Sand2-3        ..............................................................T.......
SandB-2        ..............................................................T.......
SandB-3        ..............................................................T.......
PP003-11       ..............................................................T.......
WW154-20       ..............................................................T.......
CC5-1          ..............................................................T.......
CC5-3          ..............................................................T.......
CC6-7          ..............................................................T.......
TT-12          .........................................G....................T.......
NALR7          ..............................................................T.......
Aus            ..............................................................T.......
Sand1-1        ..............................................................T.......
SandA-1        ..............................................................T.......
SandA-3        ...T..........................................................T.......
SandF-1        ...........................................................G..T.......
SandF-2        ...........................................................G..T.......
SandF-3        ...........................................................G..T.......
PP0-6          ...........................................................G..T.......
Z25-14         ..................C........................................G..T.......
DD020-1        ...........................................................G..T.......
DD020-2        ...........................................................G..T.......
DD020-3        ...........................................................G..T.......
NethV          ....................................................................
Neth31         ..............................................................T.......
NethV4         ..............................................................T.......
Neth11         ....................................................................
Neth30         ....................................................................
Pak            ....................................................................
OP             ....................................................................
SKorRB         ....................................................................
NethWa         ....................................................................
Cuba           ....................................................................
India          ....................................................................
Italy          ...................................................................C
14.2           ....................................................................
FrPT           ....................................................................
Fr1            ....................................................................
Zim13          ....................................................................
Egp            ....................................................................
Pol            ....................................................................
SKor777        ....................................................................
Can            ....................................................................
SKor           ..................C.................................................
Neth15         ....................................................................
Sco            ....................................................................
SandH-7        ....................................................................
SandH-10       ....................................................................
Z26-23         ....................................................................
Z26-24         ....................................................................
Z26-25         ....................................................................
CIP01          ....................................................................
Ger            ...............T....................................................
Noir           ....................................................................
Chi            ....................................................................
Cu87           ....................................................................
FrCU87         ....................................................................
SPSLV          ....G..CGA.C.T...G.A.G....G............T...................A.........
```

```
                    360       370       380       390       400       410       420
                ....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Rood            TCAGTTCGTCAGCGAGGCCTCTTCCACCTCCTCCGGTTCCATCGCTTATGAGTTGGACCCCCATTGCAAA
NN3-1           ......................................................................
NN3-2           ......................................................................
Sand2-1         ...........................................................G..........
Sand2-2         ......................................................................
NN334-7         ......................................................................
Sand2-3         ......................................................................
SandB-2         ....................................................................T....
SandB-3         ....................................................................T....
PP003-11        ....................................................................T....
WW154-20        ....................................................................T....
CC5-1           ....................................................................T....
CC5-3           ...................T................................................T....
CC6-7           ....................................................................T....
TT-12           ....................................................................T....
NALR7           ........................G....C........................................
Aus             ........................G....C........................................
Sand1-1         ........................G....C........................................
SandA-1         ........................G....C........................................
SandA-3         ........................G....C........................................
SandF-1         .............................C...............A........................
SandF-2         .............................C...............A........................
SandF-3         .............................C...............A........................
PP0-6           .............................C...............A........................
Z25-14          .............................C...............A........................
DD020-1         ...................T.........C...............A........................
DD020-2         .............................C...............A........................
DD020-3         .............................C...............A........................
NethV           ........................T...........................................T....
Neth31          ........................T.............................................
NethV4          ........................T.............................................
Neth11          ......................................................................
Neth30          ......................................................................
Pak             ......................................................................
OP              ......................................................................
SKorRB          ......................................................................
NethWa          ......................................................................
Cuba            ...............................A......................................
India           ......................................................................
Italy           ................C.....................................................
14.2            ......................................................................
FrPT            ......................................................................
Fr1             ......................................................................
Zim13           ......................................................................
Egp             ......................................................................
Pol             ......................................................................
SKor777         ......................................................................
Can             ......................................................................
SKor            ......................................................................
Neth15          ......................................................................
Sco             ......................................................................
SandH-7         ......................................................................
SandH-10        ......................................................................
Z26-23          ......................................................................
Z26-24          ......................................................................
Z26-25          ......................................................................
CIP01           ......................................................................
Ger             .............................C........................................
Noir            .............................C........................................
Chi             .............................C........................................
Cu87            .............................C........................................
FrCU87          .............................C........................................
SPSLV           ......................A..G...AG.....T...T......AC............T...
```
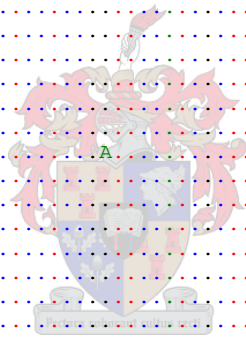
```
                        430       440       450       460       470       480       490
                 ....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Rood             ATATCATCCCTCCAGTCCTACGTCAACAAGTTCCAAATTACGAAGGGCGGCGCTAAAACCTATCAAGCGC
NN3-1            ......................................................................
NN3-2            ......................................................................
Sand2-1          ......................................................................
Sand2-2          ......................................................................
NN334-7          ......................C...............................................
Sand2-3          ...............................................G.................G....
SandB-2          ...............................A.........C..T.........................
SandB-3          ...............................A.........C..T.........................
PP003-11         ...............................A.........C..T.........................
WW154-20         ...............................A.........C..T.........................
CC5-1            ...............................A.........C..T.........................
CC5-3            ...............................A.........C..T.........................
CC6-7            ...............................A.........C..T.........................
TT-12            ...............................A.........C..T.........................
NALR7            ......................................................................
Aus              ......................................................................
Sand1-1          ......................................................................
SandA-1          ......................................................................
SandA-3          ......................................................................
SandF-1          ...C.....................................A...........G.................
SandF-2          .........................................A...........G.................
SandF-3          .........................................A...........G.................
PP0-6            .........................................A...........G.................
Z25-14           .........................................A...........G.................
DD020-1          .........................................A...........G.................
DD020-2          .........................................A...........G.................
DD020-3          .........................................A...........G.................
NethV            ......................................................C.....T.........
Neth31           ...................................C..................C.....T.........
NethV4           ......................................................C.....T.........
Neth11           G.....................................................C.....T.........
Neth30           G.....................................................C.....T.........
Pak              G.....................................................C.....T.........
OP               G......................................G..............C.....T.........
SKorRB           G.....................................................C.....T.........
NethWa           G...................C.............C.TC................C.....T.........
Cuba             G..............................A.....................C.....T.........
India            G.....................................................C.....T.........
Italy            G.....................................................C.....T.........
14.2             G.....................................................C.....TT........
FrPT             G.....................................................C..G.TT........
Fr1              G.....................................................C.....T.........
Zim13            G.....................................................C.....T.........
Egp              G.....................................................C.....T.........
Pol              G.....................................................C.....T.........
SKor777          G.....................................................C.....T.........
Can              G.....................................................C.....T.........
SKor             G.....................................................C.....T.........
Neth15           G.....................................................C.....T.........
Sco              G.....................................................C.....T.........
SandH-7          G.....................................................C.....T.........
SandH-10         G.....................................................C.....T.........
Z26-23           G.....................................................C.....T.........
Z26-24           G..............................................A......C.....T.........
Z26-25           G.....................................................C.....T.........
CIP01            G.....................................................C.....T.........
Ger              G.....................................................C.....T.........
Noir             G.....................................................C.....T.........
Chi              G.G...........T.......................................C.....T.........
Cu87             G.....................................................C.....T.........
FrCU87           G.....................................................C.....T.........
SPSLV            ...............TCC...AC...............TC.G......A.....T..G.GGT.G.TCACC...A
```

```
                   500       510       520       530       540       550       560
          ....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Rood      GGATGATAAACGGGGTAGAATGGCACGATTCGTCTGAGGATCAGTGCCGGATTCTGTGGAAAGGAAATGG
NN3-1     ..................................................................
NN3-2     ..................................................................
Sand2-1   ...........................................A......................
Sand2-2   ...........................................A......................
NN334-7   ...........................................A......................
Sand2-3   ...........................................A......................
SandB-2   ...................G.......................A......................
SandB-3   ...................G.......................A......................
PP003-11  ...................G.......................A......................
WW154-20  ...................G.......................A......................
CC5-1     ...................G.......................A......................
CC5-3     ...................G.......................A......................
CC6-7     ...................G.......................A......................
TT-12     ...................G.......................A......................
NALR7     ...........................................A......................
Aus       ...............................A...........A......................
Sand1-1   ...........................................A......................
SandA-1   ...........................................A......................
SandA-3   ...........................................A......................
SandF-1   ...........................................A............T.........
SandF-2   ...........................................A............T.........
SandF-3   ...........................................A............T.........
PP0-6     ...........................................A............T.........
Z25-14    ...........................................A............T.........
DD020-1   .......................................................T.........
DD020-2   .......................................................T.........
DD020-3   .......................................................T.........
NethV     ....................T......................A.........G............
Neth31    ....................T......................A.........G............
NethV4    ....................T...................A..A.........G............
Neth11    ....................T......................A.........G............
Neth30    ....................T......................A.........G............
Pak       ....................T......................A.........G............
OP        ...................C.T......................A.........G...........
SKorRB    ...................G.T......................A.........G............
NethWa    ....................T......................A......................
Cuba      ....................T......................A.........G............
India     ....................T......................A.........G............
Italy     ....................T......................A.........G..C.........
14.2      ....................T......................A.........G............
FrPT      ....................T......................A.........G............
Fr1       ....................T......................A.........G............
Zim13     ....................T......................A.........G............
Egp       ..........T.........T......................A......................
Pol       ..........T.........T......................A......................
SKor777   ....................C......................A.........G............
Can       ....................T......................A.........G............
SKor      A.........T.........T..............T.......A.........G............
Neth15    ....................T......................A.........G............
Sco       ....................T......................A.........G............
SandH-7   ..........T.........T......................A.........G............
SandH-10  ....................T......................A.........G............
Z26-23    ....................T......................A.........G............
Z26-24    ....................T......................A.........G............
Z26-25    ....................T......................A.........G............
CIP01     ....................T......................A.........G............
Ger       ....................T......................A.........G............
Noir      ....................T......................A.........G............
Chi       ....................T......................A.........G............
Cu87      ....................T......................A.........G............
FrCU87    ....................T......................A......................
SPSLV     ......C......C.G............T...............A......AC..G..C..C.
```
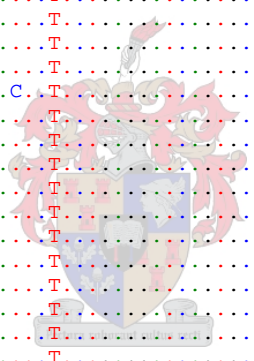
```
                  570       580       590       600       610       620       630
              ....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Rood          AAAATCTTCAGATCCCGCAGGATCCTTCAGAGTCACCATCAGGGTGGCTTTGCAGAACCCCAGATAG
NN3-1         ...................................A.............A....
NN3-2         .................................................A....
Sand2-1       ...................T........T.....A..............A....
Sand2-2       ...................T........T.....A..............A....
NN334-7       ...................T........T.....A...A..........A...A
Sand2-3       ...................T........T.....A..............A....
SandB-2       .......C...........T..............A..............A....
SandB-3       .......C...........T..............A..............A....
PP003-11      .......T.C.........T..............A..............A....
WW154-20      .......G.C.........T..............A..............A....
CC5-1         .......C...........T..............A...A..........A....
CC5-3         .......C...........T..............A..............A....
CC6-7         .......C...........T..............A..............A....
TT-12         .......C...........T..............A..............A....
NALR7         .......C........T..T..............A.....C....A....A....
Aus           ...................T........T.....A..........A...A....
Sand1-1       .......C...........T.............AA..............A....
SandA-1       .......C...........T.............AA..............A....
SandA-3       .......C...........T.............AA..............A....
SandF-1       .......C...........T..............A..............A....
SandF-2       .......C...........T..............A..............A....
SandF-3       .......C...........T..............A..............A....
PP0-6         .......C...........T..............A..............A....
Z25-14        .......C...........T..............A..............A....
DD020-1       ...................T..............A..............A....
DD020-2       .............T.....T..............A..............A....
DD020-3       ...................T..............A..............A....
NethV         .....C.....A.................................A....A....
Neth31        ...........A.................................A....A....
NethV4        ...........A.................................A....A....
Neth11        ...........A.................................A....A....
Neth30        ...........A.................................A....A....
Pak           ...........A.................................A....A....
OP            ...........A.................................A....A....
SKorRB        .....C.....A.................................A....A....
NethWa        ...........A.................................A....A....
Cuba          ...........A.................................A....A....
India         ...........A.................................A....A....
Italy         ...........G.....................A..A........A....A....
14.2          ...........A................T................A....A....
FrPT          ...........A................T..T...C.........A....A....
Fr1           ...........A.................................A....A....
Zim13         ...........A.................................A....A....
Egp           ...........A.................................A....A....
Pol           ...........A.................................A....A....
SKor777       ............................................A....A....
Can           ............................................A....A....
SKor          ...........A.................................A....A....
Neth15        .........T................A.................A....A....
Sco           .........T................A...A.............A....A....
SandH-7       ...........A.................................A....A....
SandH-10      ...........A.................................A....A....
Z26-23        ...........A.................................A....A....
Z26-24        ...........A.................................A....A....
Z26-25        ...........A.................................A....A....
CIP01         ...........A.............................A...A....A....
Ger           ...........A.............................A...A....A....
Noir          ...........A.............................A...A....A....
Chi           ...........A................T..............A.-----------
Cu87          .....C.....CA................................A....A....
FrCU87        .....C.....CA................................A....A....
SPSLV         CGGCAA.----..CAT...C.........C.........A.C......C..C.......-..A....GTAG
```

Addendum B

Alignment of the deduced PLRV CP amino acid sequences of one European (Neth 31) and one South African (Sand2-2) PLRV CP nucleotide sequence. These amino acid sequences show 97.1% homology. Amino acid substitutions are shown in bold underlined. Three of the amino acid substitutions are conservative, found in amino acid positions 7 (Arg/Lys), 18 (Arg/Lys) and 80 (Met/Val) and three of the amino acid substitutions are non-conservative found in amino acid positions 2 (Ser/Gly), 153 (Thr/Ile) and 192 (Thr/Pro).

```
                 10        20        30        40        50        60        70
        ....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Neth31  MSTVVVRGNVNGGVQQPRRRRRQSLRRRANRVQPVVMVTAPGQPRRRRRRRGGNRRSRRTGVPRGRGSSE
Sand2-2 MGTVVVKGNVNGGVQQPKRRRRQSLRRRANRVQPVVMVTAPGQPRRRRRRRGGNRRSRRTGVPRGRGSSE

                 80        90        100       110       120       130       140
        ....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Neth31  TFVFTKDNLMGNSQGSFTFGPSLSDCPAFKDGILKAYHEYKITSILLQFVSEASSTSSGSIAYELDPHCK
Sand2-2 TFVFTKDNLVGNSQGSFTFGPSLSDCPAFKDGILKAYHEYKITSILLQFVSEASSTSSGSIAYELDPHCK

                 150       160       170       180       190       200
        ....|....|....|....|....|....|....|....|....|....|....|....|....
Neth31  ISSLQSYVNKFQTTKGGAKTYQARMINGVEWHDSSEDQCRILWKGNGKSSDTAGSFRVTIRVALQNPK*
Sand2-2 ISSLQSYVNKFQITKGGAKTYQARMINGVEWHDSSEDQCRILWKGNGKSSDPAGSFRVTIRVALQNPK*
```
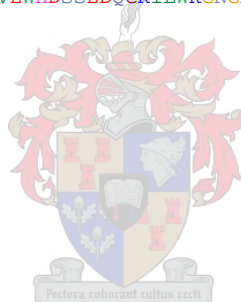
Addendum C

Sequence results (with T7 forward primer) of the pET14b vector with PLRV CP gene insert to show the ribosome binding site, *Nco*1 restriction enzyme cut site, start codon, PLRV CP gene insert, stop codon and the *Xho*1 restriction enzyme cut site. Rare arginine codons (AGA, AGG, CGG) are highlighted in yellow.

Landmarks of pET14-b vector with PLRV CP gene insert

Ribosome binding site (rbs) (underlined)	8-12
*Nco*1 restriction enzyme cut site (underlined)	18-23
ATG start codon (bold)	20-22
TAG stop codon (bold)	644-646
*Xho*1 restriction enzyme cut site (underlined)	647-652

```
        10        20        30        40        50        60        70        80        90
....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
CTTTAGAAGGAGATATACCATGGGTACGGTCGTGGTTAAAGGAAATGTCAATGGTGGTGTACAACAACCAAAGAGGCGAAGAAGGCAATC
        rbs          Nco1

        100       110       120       130       140       150       160       170       180
....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
CCTTCGAAGGCGCGCTAACAGAGTGCAGCCAGTGGTTATGGTCACGGCCCCTGGGCAACCCAGGCGCCGAAGACGCAGAAGAGGAGGCAA

        190       200       210       220       230       240       250       260       270
....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
TCGCCGCTCAAGAAGAACTGGAGTTCCCCGAGGACGAGGCTCAAGCGAGACATTCGTGTTTACAAAGGACAACCTCGTGGGCAATTCCCA

        280       290       300       310       320       330       340       350       360
....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
AGGAAGTTTCACCTTCGGGCCGAGTCTATCAGACTGTCCGGCATTCAAGGATGGAATACTCAAGGCCTACCATGAGTATAAGATCACAAG

        370       380       390       400       410       420       430       440       450
....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
TATCTTACTTCAGTTCGTCAGCGAGGCCTCTTCCACCTCCTCCGGTTCCATCGCTTATGAGTTGGACCCCCATTGTAAAATATCATCCCT

        460       470       480       490       500       510       520       530       540
....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
CCAGTCCTACGTCAACAAATTCCAAATCACTAAGGGCGGCGCTAAAACCTATCAAGCGCGGATGATAAACGGGGTGGAATGGCACGATTC

        550       560       570       580       590       600       610       620       630
....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
GTCTGAGGATCAGTGCCGGATACTGTGGAAAGGAAATGGAAAATCTTCAGACCCCGCAGGATCCTTTAGAGTCACCATCAGAGTGGCTTT

        640       650
....|....|....|....|....|....|....|....
GCAGAACCCCAAATAGCTCGAGGATCCGG
              Xho1
```
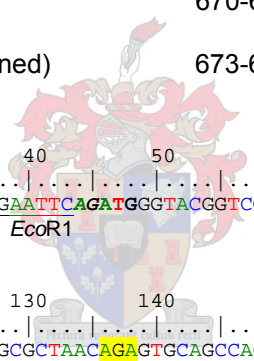
Addendum D

Sequence results (with pGEX-5' reverse primer, confirmed by the pGEX-3' forward primer) of the pGEX-6P-2 vector with PLRV CP gene insert to show the PreScission Protease cleavage site, *Eco*R1 restriction enzyme cut site, the two additional bases (AG) added for the PLRV CP gene insert to be expressed in frame with GST, start codon, PLRV CP gene insert, stop codon and the *Xho*1 restriction enzyme cut site. Rare arginine codons (AGA, AGG, CGG) are highlighted in yellow.

<u>Landmarks of pGEX-6P-2 vector with PLRV CP gene insert</u>

PreScission Protease cleavage site (underlined)   1-24

*Eco*R1 restriction enzyme cut site (underlined)   38-43

Two additional bases (AG) (bold italics)   44-45

ATG start codon (bold)   46-48

TAG stop codon (bold)   670-672

*Xho*1 restriction enzyme cut site (underlined)   673-678

```
          10        20        30        40        50        60        70        80        90
....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
CTGGAAGTTCTGTTCCAGGGGCCCCTGGGATCCCCAGGAATTCAGATGGGTACGGTCGTGGTTAAAGGAAATGTCAATGGTGGTGTACAA
      PreScission Protease                     EcoR1

          100       110       120       130       140       150       160       170       180
....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
CAACCAAAGAGGCGAAGAAGGCAATCCCTTCGAAGGCGCGCTAACAGAGTGCAGCCAGTGGTTATGGTCACGGCCCCTGGGCAACCCAGG

          190       200       210       220       230       240       250       260       270
....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
CGCCGAAGACGCAGAAGAGGAGGCAATCGCCGCTCAAGAAGAACTGGAGTTCCCCGAGGACGAGGCTCAAGCGAGACATTCGTGTTTACA

          280       290       300       310       320       330       340       350       360
....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
AAGGACAACCTCGTGGGCAATTCCCAAGGAAGTTTCACCTTCGGGCCGAGTCTATCAGACTGTCCGGCATTCAAGGATGGAATACTCAAG

          370       380       390       400       410       420       430       440       450
....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
GCCTACCATGAGTATAAGATCACAAGTATCTTACTTCAGTTCGTCAGCGAGGCCTCTTCCACCTCCTCCGGTTCCATCGCTTATGAGTTG

          460       470       480       490       500       510       520       530       540
....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
GACCCCCATTGTAAAATATCATCCCTCCAGTCCTACGTCAACAAATTCCAAATCACTAAGGGCGGCGCTAAAACCTATCAAGCGCGGATG

          550       560       570       580       590       600       610       620       630
....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
ATAAACGGGGTGGAATGGCACGATTCGTCTGAGGATCAGTGCCGGATACTGTGGAAAGGAAATGGAAATCTTCAGACCCCGCAGGATCC

          640       650       660       670       680
....|....|....|....|....|....|....|....|....|....|
TTTAGAGTCACCATCAGAGTGGCTTTGCAGAACCCCAAATAGCTCGAGCGGCCGC
                                           Xho1
```