AN INVESTIGATION INTO THE HIGH LEVEL PRODUCTION OF PROTEINS IN TOBACCO USING TRANSGENIC PLANTS OR VIRAL VECTORS.

by

MICHELLE M. GEHRINGER

A thesis submitted to the Faculty of Science (Department of Microbiology) of the University of Cape Town in fulfilment of the requirements for the degree of **Master of Science.**

February 1996

ACKNOWLEDGEMENTS

This project and thesis would not have been at all possible without the contributions of several people:

A big thank you to **Prof. Ed Rybicki** for input, suggestions and encouraging me through the lows.

To **AECI** and **Dr. Karl Kunert** for funding the project.

To **Prof. M.B. von Wechmar** for advice, support and the use of the plant rooms.

To **Kenneth**, an invaluable critic and advisor.

To **Janet** for filling in the blanks.

To my **labmates** for their advice, entertainment and the odd reagent.

To **John Fitton** for computer assistance.

To **Beth.**

SUMMARY

The aim of this project was to construct a high level plant expression vector from the RNA 3 of cucumber mosaic virus strain Y (CMV Y). The 5'- and 3'-untranslated regions (UTRs) of this genome segment were reverse transcribed, cloned and sequenced. The chloramphenicol acetyl transferase gene (CAT) was inserted between the two UTRs. This artificial viral cDNA (5'cat3') was cloned immediately downstream of the cauliflower mosaic virus 35 S promoter at the transcription initiation site to make a DNA vector. An RNA vector construct was made by placing the 5'cat3' segment under the control of a T7 RNA promoter sequence. *In vitro* transcripts, as well as linearised DNA vector constructs were inoculated onto CMV infected plants. Inoculated plants were monitored for CAT expression. No CAT could be detected in total protein extracts of inoculated plants. No CAT mRNA could be detected in northern blots of total RNA extracted from inoculated plants. The vector constructed from the 5'- and 3'-UTR of the RNA 3 of CMV Y did not appear to contain all the necessary attributes for a viral expression vector.

To study the expression of a foreign antigen in tobacco, the L1 capsid protein of human papillomavirus type 16 was cloned into *Agrobacterium tumefaciens* and used to make transgenic *Nicotiana tabacum*. Kanamycin resistant tobacco plants were shown to carry the L1 capsid gene using PCR screening, but western blots on total protein extracts of the transformed plants were indeterminate. Further studies are needed to determine whether the antigen is produced and if it is correctly processed.

LIST OF ABBREVIATIONS

(+)	messenger sense
(-)	complementary sense
ACEI	angiotensin I converting enzyme inhibitor
ACMV	African cassava mosaic virus
AIMV	alfalfa mosaic virus
AMV	avian myeloblastosis virus
ApMV	apple mosaic virus
ATP	adenosine-5'-triphosphate
BBMV	broad bean mottle virus
BMV	brome mosaic virus
bp	base pair
°C	degrees Celsius
CaMV	cauliflower mosaic virus
CaMV 35S	CaMV 35S promoter region
CAT	chloramphenicol acetyl transferase
CIP	calf intestinal phosphotransferase
CMV	cucumber mosaic virus
cDNA	complementary DNA
CP	coat protein
CRPV	cottontail rabbit papillomavirus
cv	cultivar
Da	Dalton
DHFR	dihydrofolate reductase
DIG	digoxigenin
DNA	deoxyribonucleic acid
DNase	deoxyribonuclease
dNTP	2'-deoxynucleoside-5'-diphosphate
ds	double-stranded

EDTA	ethylenediaminetetra-acetic acid
et al.	et alia
EtBr	ethidium bromide
Fig.	figure
FMDV	foot-and-mouth disease virus
g	gram
GCG	Genetics Computer Group
Gus	β-glucuronidase
HBsAg	hepatitis B surface antigen
HIV	human immunodeficiency virus
HPV	human papillomavirus
HRV	human rhinovirus
HSA	human serum albumin
IGR	intergenic region
IFN-γ	γ-interferon
kb	kilobase pairs
kDa	kilodalton
l	litre
λ	lambda
M	molar
MCS	multiple cloning site
m^7Gppp	7-methyl guanosine triphosphate
min	minutes
ml	millilitre
mM	millimolar
mol	moles
MSV	maize streak virus
MuMLV	murine Moloney leukemia virus
ng	nanogram
NPTII	neomycin phosphotransferase
OD	optical density
ORF	open reading frame

ORSV	odontoglossum ringspot virus
PCR	polymerase chain reaction
PDV	prune dwarf virus
p.i.	post infection
PNK	polynucleotide kinase
PSV	peanut stunt virus
RI	replicative intermediate
RNA	ribonucleic acid
RNase	ribonuclease
rpm	revolutions per minute
RT	reverse transcriptase
SDS	sodium dodecyl sulphate
ss	single-stranded
SSC	salt-sodium citrate buffer
STE	sodium-Tris-EDTA buffer
TAE	Tris-acetate electrophoresis buffer
TBE	Tris-borate electrophoresis buffer
TEV	tobacco etch virus
TMV	tobacco mosaic virus
Tris-HCl	Tris(hydroxymethyl)-aminomethane
TSV	tobacco streak virus
U	unit
µg	microgram
µl	microlitre
UTR	untranslated region
UV	ultraviolet
VLP	virus-like particle
v/v	volume per volume
w/v	weight per volume

CONTENTS

CHAPTER 4:

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

LITERATURE SURVEY

"Africa is not poor, but it is, absolutely, impoverished" says Thomas Odhiambo, president of the African Academy of Sciences. The *per capita* gross national product (GNP) was a low U.S. $340 in 1992 and the average life expectancy a mere 52 years. This gloomy picture can be altered by a two pronged attack: adequate primary health care, especially immunisation against common diseases such as malaria, hepatitis (and hopefully one day HIV) and secondly, by providing an industry capable of providing important industrial chemicals at low cost without depriving the communities of natural resources. Plant biotechnology could have a major impact on these aspects: both oral vaccines and industrially important chemicals could be manufactured in crop plants at low cost. The advantage of the production of these biomaterials is that they are renewable and their production would not stress the local environment (Somerville, 1993). It is in this vein that the project of investigating means of synthesising foreign proteins in plants was undertaken.

TRANSGENIC PLANTS AS MANUFACTURERS

INTRODUCTION.

Recent developments in molecular biology have made it possible to produce large amounts of useful chemicals in plants. In comparison to bacterial or yeast fermentations, crop plants are capable of producing these proteins at relatively high expression levels at low cost. Proteins are generally stable and fully functional (Herbers *et al.*, 1995).

Agracetus Inc., a biotechnology company based in Wisconsin, USA, was awarded the patent in 1992 for its particle-mediated gene delivery technology to create

plants that can produce high levels of biopharmaceuticals and other industrial enzymes in plant tissues. They maintain that their plant bioreactor production system (PBP) has several advantages over bacterial fermentation or cell culture systems. Microbial fermentations require complex machinery and input costs estimated to between $10 000 and $1 million per kilogram of crude extract. If the protein of interest is produced in seeds or tissue at 0,1% - 1% of the total protein, the production costs are reduced to about $1 000 per kilogram of crude extract. PBP systems eliminate the potential of contamination of microbial or cell culture systems with harmful pathogens. Transgenic plants can produce biologically active proteins and be accumulated in food crops that already have a long history of safe use. Microbial and cell culture systems are difficult to scale up, the major problems being with aeration, cell density and product recovery. Proteins produced in plants either by gene insertion or utilising a viral vector can provide easy and consistent scale-up simply by increasing the acreage of the crop concerned (Donson *et al.*, 1991)

EXAMPLES OF PLANT-MADE PRODUCTS.

It has been demonstrated that polyhydroxy butyrate (PHB), a biodegradable plastic polymer, can be produced in transgenic *Arabidopsis thaliana*. These plants produced approximately 100 μg PHB per gram of total fresh weight (Poirier *et al.*, 1995). The same authors are now attempting to produce polyhydroxyalkanoates, naturally occurring polyesters that accumulate as inclusions in several bacteria, in plants. *A.thaliana* was chosen as a model system as it is closely related to the oil-producing crop oilseed rape, a target crop for the production of PHA on an agricultural scale.

Transgenic plants can also be designed to synthesise enzymes for commercial and technical purposes. Plant cell-wall degrading enzymes are required by several industries to either purify plant derivatives or convert waste products of the food and paper-pulp industries to animal feed. To this end a thermostable xylanase from *Clostridium thermocellum* was expressed to high levels in

transgenic tobacco. The amount of protein obtained in plant extracts comprised approximately 4% of the total protein and it retained its enzymatic activity. Targeting the protein to the apoplastic space by the proteinase inhibitor II signal peptide made for easier purification of the enzyme from intercellular fluid extracts (Herbers *et al.*, 1995).

Biologically active peptides have also been produced successfully in *Arabidopsis* and oilseed rape. A portion of an *Arabidopsis thaliana* seed storage protein gene, the 2S albumin gene, is not highly conserved and was replaced with the sequence encoding the neuropeptide Leu-enkephalin. This pentapeptide, a protein which displays opiate activity, was chosen for its small size and availability of a commercially produced version for comparative studies. Tryptic cleavage sites were included at either side of the inserted leu-encephalin protein, thereby allowing for the recovery of the peptide after digestion with trypsin from 2S albumin extracts (Vandekerckhove *et al.*, 1989). Up to 200 nmol of peptide were recovered per gram of seed.

Sijmons *et al.* (1990) set out to explore the potential of plants to produce commercially important proteins by expressing the human serum albumin protein (HSA) in potato plants. In man, HSA is produced as prepro-albumin by the liver and processed twice to yield the mature secretory polypeptide. These authors found that the prepro-signal sequence of HSA could target the protein across the membrane of potato cells, but that the recombinant protein was incompletely processed. They were able to construct a hybrid plant/human precursor which resulted in secretion of correctly processed HSA that was indistinguishable from the original human protein. The starch potato demonstrates the feasibility of harvesting a crop for two yields: the process of wet milling potatoes allows both starch from the tubers and a commercially viable protein from the liquid fraction to be harvested. The liquid fraction is usually a waste product (Sijmons *et al.*, 1990).

Gene transfer to plants also offers the opportunity for genetic modification of crops. The potential exists to improve the flavour of edible plant products by

genetic manipulation. The gene coding for monellin, a protein which elicits a sweet flavour, was placed under the control of fruit-ripening specific promoters and transferred to lettuce and tomato plants (Penarrubia *et al.,* 1992). Expression of the gene in the transgenic plants resulted in the accumulation of monellin protein in tomatoes and lettuce leaves. This illustrates that gene transfer of foreign sequences offers an alternative means of enhancing the flavour and quality of fruits and vegetables.

PLANT VIRUSES AS VECTORS

The production of transgenic plants is a costly and lengthy process. An alternative to this approach is the engineering of simple plant viruses so that they will express foreign genes at high levels in infected plants. The major advantage of this technique is that it is far quicker to engineer a virus than it is to engineer a plant. A variety of foreign genes may be inserted into a viral vector with the distinct advantage that the virus product is expressed immediately, rather than after the weeks it takes to regenerate plants from transformed tissue (Boyer and Haenni, 1994).

Nicotiana tabacum is a suitable crop for the production of viable proteins as it is fast-growing and can be infected with several types of plant viruses that are not too severe on the host plant. Tobacco crops provide a large yield per plant of leaf mass and the infrastructure is all ready in place for processing large amounts of leaf tissue. The concept of using plants as factories has captured the imagination of the popular press. Tom Turpen of Biosource Technologies is quoted in the New Scientist as saying that their system of using tobacco mosaic virus (TMV) to produce an experimental vaccine against malaria, could produce as much as 250 kilograms of vaccine per month from a hectare of tobacco plants infected with the modified TMV (Coghlan 1995).

DNA VIRAL VECTORS.

Since the characterisation of plant DNA viruses there has been a growing interest in their potential as vectors for delivering genes into plants. Cauliflower mosaic virus (CaMV), a double stranded DNA virus which replicates by means of reverse transcription of its genomic RNA template, was the first virus to be used as a plant virus vector (Brisson *et al.*, 1984; Mason *et al.*, 1987).

The non-essential CaMV ORF II was replaced with the dihydrofolate reductase (*dhfr*) gene, a gene which confers resistance to methotrexate in *Escherichia coli* (Brisson, *et al.*, 1984). Turnip plants inoculated with the modified virus were found to develop symptoms in the same period as the control plants, suggesting that the inclusion of the *dhfr* gene was not detrimental to virus replication. Viral DNA carrying the *dhfr* gene was isolated from virus particles obtained from systemically infected leaves, proving that the insert had been retained in the viral genome. Methotrexate studies were done to indicate the functionality of the DHFR enzyme and found it to be fully active. However expression of DHFR was significantly reduced after two or three cycles of infection. This phenomenon of insert instability was later found to be independent of the insert size (De Zoeten, 1989).

De Zoeten *et al.* (1989) were able to produce human interferon αD (IFN αD) in turnip plants using a CaMV ORF II vector. Greater insert stability was achieved by linking the IFN αD gene to the absolute beginning and end of the ORF II gene, thereby preserving the original ORF arrangement. Vectors carrying short untranslated regions flanking the inserted gene were found to be unstable. This confirms that CaMV could be used to successfully deliver and express foreign genes in plants provided that the original ORF organisation is maintained. The addition of any extraneous nucleotides before the inserted ORF resulted in the insert being lost after 2 - 3 cycles of inoculation (Brisson *et al*, 1984; De Zoeten *et al.*, 1989).

Geminiviruses are apparently suitable candidates for plant vectors as collectively they infect a wide range of monocotyledonous (monocot) and dicotyledonous (dicot) plants (Stanley, 1993). They replicate via a double stranded (ds) DNA intermediate in the plant cell nucleus. The potential of maize streak virus (MSV) a geminivirus infecting monocots, as a gene vector, was investigated by Shen and Hohn (1994). A reporter gene coding for β-glucuronidase (GUS) was inserted into the non-coding region of the genome of MSV and used to infect maize seedlings by using *Agrobacterium*-mediated inoculations. This MSV-GUS construct was found to induce more blue spots in maize seedlings than in control seedlings inoculated with only a GUS construct. The increase in GUS expression was even more pronounced 1 week after inoculation (Shen & Hohn, 1994). The MSV-GUS construct was however unable to move to adjacent tissues or induce symptoms in seedlings. A possible reason for the lack of movement of the 5.9 kb MSV-GUS construct could have been that it was too large to be packaged when compared to the 2.7 bp naturally occurring MSV genome.

Another problem *en route* to developing a geminivirus vector is the means of delivery of the vector. Few geminiviruses can be mechanically inoculated onto plants. Viral sequences can however, be introduced to the plant by using *Agrobacterium* containing the viral vector sequences (Stanley 1993), but this method of introducing the foreign sequence to the host plant would not be feasible for large scale inoculations.

African cassava mosaic virus (ACMV), another member of the Geminiviridae, has also been investigated for its potential as a plant expression vector. The chloramphenicol acetyl transferase (*cat*) gene was inserted in place of the ACMV coat protein gene and successfully expressed in tobacco plants. However, vector instability during plant infection poses a major problem to the development of a geminiviral vector. Vectors appear only to remain stable during systemic infection when they are comparable in size to the wild type component. This problem of instability due to size constraint could be overcome by making use of a binary vector system whereby the viral replication associated genes are supplied in *trans*

from integrated copies in the plant genome. These constitutively produced proteins could support the efficient replication of replication protein deficient ACMV replacement vectors, thereby supporting the expression of the protein of interest (Stanley, 1993).

RNA VIRAL VECTORS.

Viral expression vectors based on RNA viruses have several distinct advantages over integration of foreign proteins into plant genomes: RNA viruses have wide host ranges; they are able to move from cell to cell mediated by virally encoded movement proteins; they exhibit rapid systemic spread in plants; their infections are maintained for the lifetime of the plant; their RNA is replicated to high levels as autonomous sequences in the host which results in rapid and productive cytoplasmic gene expression (Ahlquist *et al.*, 1987; Donson *et al.*, 1991). Tobacco mosaic virus (TMV) and brome mosaic virus (BMV) are two examples of RNA viruses that have been used successfully as plant expression vectors. The ability to effectively manipulate RNA viruses through a cDNA copy allows for simpler cloning strategies and increases their appeal as viral vectors. The cDNA copy can then be transcribed *in vitro* to produce infectious RNA molecules (Donson *et al.*, 1991; Hamamato *et al.*, 1993; Kumagai *et al.*, 1993).

Although perhaps the most successful RNA virus engineering recently has been that of cowpea mosaic virus, this vector system relies on the expression of peptide sequences inserted into the capsid and is not capable of expressing full-length foreign polypeptides. It will therefore not be discussed in this review, but will be dealt with below in the section on vaccines (Usha *et al.*, 1993).

The TMV genome consists of a single 6395 nucleotide molecule of messenger sense, single stranded RNA encoding four proteins. The two replicase proteins are translated directly from the genomic RNA, while the movement protein and the coat protein are translated from two 3'-coterminal subgenomic RNA's produced during replication. Its helical structure means that there are no packaging

constraints that are found in nonhelical viruses (Donson *et al.*, 1991). TMV easily infects plants and a large amount of virus is produced after infection with the virus (Hamamoto *et al.*, 1993). Early attempts to express foreign proteins in TMV by gene replacement resulted in either failure of the modified virus to infect plants (Joshi *et al.* 1990) or loss of long distance movement (French *et al.* 1986). Addition of an extra viral subgenomic promoter to drive expression of the foreign gene resulted in the gene sequence being deleted. This was thought to be due to recombination between the two repeated subgenomic promoter sequences within the viral construct (Dawson *et al.* 1989). Donson *et al.* (1991) avoided the problem of deletion of foreign inserts due to homologous flanking sequences by using two heterologous promoters in their vector. They used the TMV-U1 promoter to drive expression of the subgenomic RNAs for the foreign gene and the functionally similar odontoglossum ringspot virus (ORSV) promoter to synthesis ORSV coat protein. ORSV showed very low sequence similarity (45%) to TMV-U1 in that region. Two reporter genes were inserted independently into the vector, TB2; namely DHFR and the neomycin phosphotransferase (NPTII) genes. The constructs both moved systemically in *Nicotiana benthamiama* although the symptoms were milder than for wild type TMV U-1. Hybrid RNAs containing both the TMV U-1 and the inserted sequences were encapsidated by the ORSV coat protein, thereby facilitating their transmission on passage to subsequent plants. DHFR was retained without any deletions in plants after 10 passages. The NPTII gene showed sequence deletions of different portions in each independent passage series of the vector. Such deletions would not appear to be a defect of the TB2 vector, based on the stability of the DHFR sequence, but rather an inherent feature of the inserted sequence itself (Donson *et al.*, 1991).

Vector TB2 was subsequently used to express biologically active α-trichosanthin to high levels in tobacco plants (Kumagai *et al.*, 1993). The α-trichosanthin gene was placed under the control of the TMV U-1 subgenomic promoter in the TB2 vector and inoculated onto tobacco plants. Two weeks after infection, α-trichosanthin had accumulated to levels of at least 2% of the total soluble protein.

This demonstrated that vector TB2 provides an effective tool for expression of foreign sequences in plants.

Hamamoto *et al.* (1993) made use of a six base sequence that follows the stop codon for the TMV 130K protein gene and which allows readthrough of the stop codon to create a novel vector. They designed an expression vector in which the six base sequence is inserted immediately after the stop codon for the coat protein (CP) gene and immediately upstream of the inserted foreign gene sequence; in this case the angiotensin-I-converting enzyme inhibitor peptide (ACEI). The vector produced both intact CP as well as a fused protein consisting of CP and ACEI (CP-ACEI). The amount of CP-ACEI produced was 100μg/g of fresh tissue in both the tobacco and tomato leaves and 10μg/g of fresh tissue in tomato fruit. Both CP and CP-ACEI were found to successfully encapsidate the TMV vector RNA. This unexpected result may be due to the fact that the C-terminal portion of the coat protein projects out from the TMV particle. They are in the process of determining whether the vector can be used for the production of longer peptides or proteins.

Brome mosaic virus (BMV) is a small isometric virus of grasses. Its genome is divided into three separate RNA components which are packaged in separate virions as capped, messenger sense, single stranded RNAs. RNA's 1 & 2 encode the replicase proteins. RNA 3 encodes both the movement protein as well as the highly expressed coat protein from the subgenomic RNA 4, produced from the 3' portion of RNA 3 (Ahlquist *et al.,* 1987). The ability of this subgenomic RNA to direct high-level protein synthesis makes it an ideal candidate for an expression vector. French *et al.* (1986) developed an *in vitro* transcription system to express infectious BMV RNAs from cloned complementary DNA (cDNA) copies of the genomic components. The cloned cDNA copies allowed for recombinant DNA techniques to be used to create hybrid viral RNAs. Initial vector studies were done using variants of RNA 3 where the CP gene was replaced by the chloramphenicol acetyltransferase (*cat*) gene. Infectious transcripts of RNAs 1 & 2 and the hybrid RNA 3 were used to infect barley protoplasts and monitored for

CAT activity. CAT was expressed from those RNA 3 constructs carrying the CAT insert in the same orientation as the original viral genes. Its expression was improved when the inserted gene was placed in frame with the upstream CP initiation codon (French *et al.*, 1986). It was also shown that the RNA3/CAT hybrids accumulated to a lesser degree than wild type RNA3 in infected protoplasts. Ahlquist *et al.* (1987) demonstrated that capped BMV transcripts were infectious to whole plants whereas noncapped transcripts were not.

Mori *et al.* (1993) were able to successfully produce human gamma interferon (IFN-γ) in tobacco protoplasts using a BMV-derived vector system. The BMV strain ATCC66 used in their study encodes two types of coat protein: full length CP1 and a truncated version, CP2 which is translated from a second initiation codon downstream of CP1 (Mise *et al.*, 1992). Chimeric RNA clones were constructed by replacing each of the viral genes with IFN-γ using biologically active cDNA clones of BMV. They found that the hybrid RNA transcripts bearing IFN-γ in place of the CP2 were replicated most efficiently and produced the most foreign protein (approximately 10% of total extractable protein). IFN-γ produced in tobacco protoplasts retained its biological activity but at a reduced level when compared to that of IFN-γ produced in animal cells.

VACCINE PRODUCTION IN PLANTS

Plants hold great promise as low-cost vaccine production systems. Vaccines derived from plants should be simpler to harvest than vaccines made in the conventional ways of fermentation and cell culture, and could potentially be stored at room temperature, thereby eliminating the need for the cold chain (Coghlan, 1995). Plant vaccines also eliminate the fear of contamination with potentially infective animal viruses which always is a problem with vaccines manufactured in animal tissue culture (Thanavala *et al.*, 1995).

Mason *et al.* (1992) and Thanavala *et al.* (1995) were able to express the hepatitis B surface antigen (HBsAg) in transgenic tobacco plants. Using an enhanced CaMV promoter linked to the tobacco etch virus (TEV) 5' nontranslated leader sequence they were able obtain HBsAg yields of up to 66 ng/mg of soluble protein. The plant produced antigen was quantified in an immunoassay system using monoclonal antibodies directed against serum HBsAg. The HBsAg produced in tobacco aggregated to form pseudocapsid particles virtually indistinguishable fron serum derived HBsAg with respect to size, density sedimentation and antibody binding.

The authors went further to study the immunogenicity of the plant-derived HBsAg. Both tobacco-derived HBsAg and yeast-derived HBsAg were used to immunise BALB/c mice, and the immune responses to both antigens were compared. They found that the tobacco-derived HBsAg could elicit HBsAg-specific antibodies and prime T-cells *in vivo* that could be stimulated *in vitro* by the yeast-derived HBsAg, currently used in vaccines (Thanavala *et al.*, 1995).

An intriguing prospect is finding a way of using edible plants to deliver vaccines orally. This concept has generated a lot more interest since the discovery of a common mucosal immune system where specific antigen-activated lymphocytes from the gut can disseminate immunity in the intestinal, genital and respiratory tracts (Holmgren *et al.*, 1994). The binding subunit of *Escherichia coli* heat-labile enterotoxin (LT-B) is a highly active oral immunogen. Haq *et al.* (1995) made both transgenic tobacco and potato plants expressing LT-B. The peptide was produced in the correct conformation and reached levels of 5μg/g of total soluble leaf protein and 30 μg/g of soluble protein in potato microtubers. Protein levels were higher in constructs bearing the microsomal retention signal that retains the peptide in the endoplasmic reticulum with LT-B reaching leves of 14 μg/g of totoal soluble leaf protein and 110 μg/g of soluble potato microtuber protein. To study the oral immunogenicity of the plant derived LT-B, Balb/c mice were given crude soluble extract from transgenic tobacco leaves while control mice were dosed with an eqivalent amount of LT-B derived from *E.coli*. Amounts of serum and mucosal

antibodies were similar in both groups indicating that the plant-derived antigen retained the immunogenic qualities of the bacterial toxin. Neutralisation studies demonstrated that both groups of mice were protected to the same extent (Haq *et al.*, 1995). The most encouraging result in this study was that feeding mice potatoes expressing LT-B generated a sound oral immune response to the toxin. This clearly demonstrates the feasibility of using transgenic plants in the production and delivery of subunit vaccines.

Chemically synthesised peptides are often only immunogenic when linked to a larger carrier molecule. Plant viruses could possibly serve as cheap, effective antigen presentation systems. Usha *et al.* (1993) cloned an oligonucleotide encoding an epitope derived from VP1 of foot-and-mouth disease virus (FMDV) into the region of the cowpea mosaic virus (CPMV) encoding the small coat protein. The chimaeric construct bearing the epitope as an insert was able to direct capsid formation and infect whole cowpea plants. The modified virus was however unable to infect cowpeas systemically. FMDV-specific serum reacted with peptide extracted from infected leaves, demonstrating that it was correctly synthesised and folded. Subsequent analysis of progeny RNA produced in these plants revealed that the inserted sequence was rapidly lost during serial passaging, presumably due to homologous recombination between two direct repeats, created during the cloning process, on either side of the inserted sequence (Porta *et al.*, 1994).

The vector was redesigned to carry the insert in the centre of the βB-βC loop of the coat protein (Porta *et al.*, 1994). This site is exposed on the surface of the virus and is potentially the optimum site for any insertion. Two epitopes; one derived fron human rhinovirus 14 (HRV-14) and another from human immunodefieciency virus type 1 (HIV 1) were cloned into this site in such a way so as not to generate repeated sequences. Both chimaeric viruses induced normal CPMV symptoms on tobacco. Western blot analysis on virions purified from infected leaf tissue showed that both the HRV 14 and HIV 1 epitope possessed their correct antigenic properties. Virus specific antibodies to each of the

chimaeric viruses were induced in the sera of rabbits injected with purified chimaeric virions (Porta *et al.*,1994).

TMV has been engineered to carry and display an immunogenic malarial epitope on its surface. The B-cell epitope was inserted into the surface loop of the coat protein or fused to the C terminus using a leaky stop codon. Tobacco plants systemically infected with either construct gave high titres of genetically stable recombinant virus which bound very strongly to monoclonal antibodies specific for malaria (Turpen *et al.*, 1995).

CUCUMBER MOSAIC VIRUS

INTRODUCTION.

Brome mosaic virus (BMV), a tripartite virus belonging to the family Bromoviridae (Ahlquist *et al.*, 1987), was used to produce large quantities of human gamma interferon in tobacco protoplasts. This was the first example of a foreign protein being produced by an icosahedral RNA virus in plant cells (Mori *et al.* ,1993).

We have chosen to investigate the potential of another member of the Bromoviridae, cucumber mosaic virus (CMV), as a plant expression vector. CMV has a much broader host range than BMV, infecting both monocotyledonous (monocot) and dicotyledonous (dicots) plants. It has the largest known host range of any virus, infecting crops as diverse as bananas in Morocco, melons in Spain and sweet potato in Israel (Palukaitis *et al.*, 1992). It is precisely for its detrimental impact on agriculture that CMV has been so extensively studied. Several strains of this tripartite RNA virus have been fully sequenced (Davies and Symons, 1988; Rizzo and Palukaitis, 1988, 1989; Owen *et al.*, 1990; Ding *et al.*, 1994) and infectious cDNA copies of the viral components (Ahlquist *et al.*, 1987; Palukaitis *et al.*, 1992; Boyer and Haenni, 1994; Pogany *et al.*, 1994; Ding *et al.*, 1995) have greatly aided studies on the replication and life cycle of this pathogen. Mild strains of CMV exist that are efficient at systemically infecting the

host plant while not reducing the yield of plant material to a great extent (Palukaitis *et al.*, 1992). The CMV genome consists of 3 segments of single stranded (ss) RNA, each of which possesses the necessary signals for replication by the virus replication machinery. It is an ideal virus to use in constructing a viral vector as one need engineer only one segment of the viral genome as opposed to the entire genome of, for example; TMV. We envisaged the use of an artificial satellite RNA-like construct derived from CMV RNA 3 for the infective expression of foreign genes in a wide variety of host plants susceptible to this virus. Herewith follows a brief review of CMV and its potential as a plant expression vector.

TAXONOMY OF CMV.

Cucumber mosaic virus (CMV) belongs to the taxonomic family Bromoviridae (Fig.1.1) and genus Cucumovirus. CMV is the type member of the genus. Strains of CMV can be divided into two subgroups based on their antigenicity and severity of the symptoms. Subgroup I CMVs produce severe symptoms on their hosts, the type member being CMV Y which induces severe yellowing and mosaic symptoms on *Nicotiana glutinosa*. Subgroup II includes the milder strains of CMV, the type member being CMV M (Owen & Palukaitis, 1988; Rybicki, 1995). All viruses show strong similarities in particle morphology and genome organisation.

PARTICLE STRUCTURE.

Three native viral particles are required for infection by CMV, each particle consisting of the same protein shell, yet containing different RNA species (Palukaitis *et al.*, 1992). RNAs 1 & 2 are found in their own protein shells, but both RNA 3 & 4 are encapsidated together in a single particle. The separate CMV particles are not readily discernible from one other, each having the same diameter of about 28 nm and a sedimentation coefficient of approximately 98.6 Svedbergs (Rybicki, 1995).

CAPSID MORPHOLOGY

The capsid consists of 180 identical protein subunits arranged in pentamer and hexamer clusters with T = 3 symmetry. The coat protein gene of CMV is encoded on RNA 3, but expressed only from the subgenomic RNA 4. RNA 4 produces a single protein product of M_r 24,500.

CMV particles are icosahedral with a diameter of about 28 nm. The genomic RNA lines the protein shell on the inside, penetrating the protein layer to a depth of 2nm, thereby adding to the stability of the virion. Lysine residues on the coat protein appear to be critical in the stabilization of the capsid as they are thought to interact with the negatively charged phosphate groups of the viral RNA. These electrostatic interactions are easily disrupted in high salt concentrations or in the presence of the detergent SDS, allowing the particles to dissociate into their protein and RNA components (Palukaitis *et al.*, 1992). Fully functional CMV particles can be reassembled by lowering the salt concentration or removing the SDS. Heterologous reassembly has been demonstrated *in vitro* between the proteins and RNA of CMV and TAV (Chen and Francki, 1990) indicating that this association is non-specific..

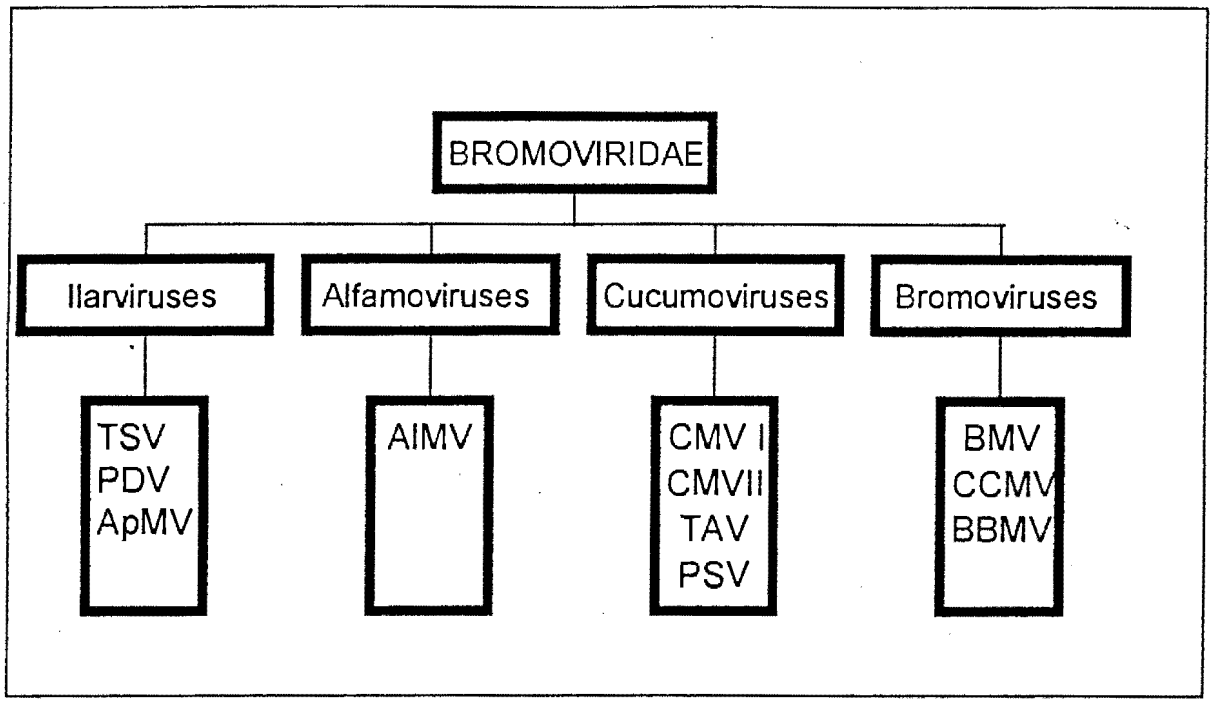

Figure 1.1. Flow chart showing the composition of the family Bromoviridae.

Abbreviations are as follows: TSV, tobacco streak virus; PDV, prune dwarf virus; ApMV, apple mosaic virus; AlMV, alfalfa mosaic virus; CMV, cucumber mosaic virus subgroup I or II; TAV, tomato aspermy virus; PSV, peanut stunt virus; BMV, brome mosaic virus; BBMV, broad bean mottle virus and CCMV, cowpea chlorotic mottle virus.

GENOME STRUCTURE.

Single stranded (ss) RNA makes up about 18% of each particle's weight. CMV has three species of messenger sense RNA; RNAs I, 2 & 3 (Table 1.1). RNA 4 is a subgenomic RNA co-terminal with the 3' end of RNA 3 and is encapsidated with RNA 3 in a single particle. All four RNAs contain a conserved untranslated 3'-terminal region and their 5' termini have 7-methylguanosine caps (Rybicki, 1995). RNA 4A has recently been described by Ding *et al.* (1994) and is derived from RNA 2 (Fig 1.2).

RNA species	Nucleotides	ORF	M_r of protein (kDa)
RNA 1	3357	1a	111
RNA 2	3050	2a	97
RNA 3	2216	3a	30
RNA 4	1031	CP	24
RNA 4A	682	2b	11.3 *

Table 1.1

MOLECULAR CHARACTERISTICS OF Fny-CMV, A SUBGROUP I MEMBER.

* Characterised for CMV Q, subgroup II. (Ding *et al.*, 1994)

RNA 1.

RNA 1 encodes an 111kDa polypeptide, 1a from a single open reading frame (Fig.1.2). Protein 1a, together with the 2a protein and a 54kDa host derived protein, has been demonstrated to form part of the active viral replicase complex (Hayes and Buck, 1990 a&b). The carboxy terminal 141 amino acids of the 1a protein are predominantly basic, a common feature among nucleic acid binding proteins (Palukaitis *et al.*, 1992). This region also contains the six conserved motifs found in both DNA and RNA helicases, suggesting a functional role for this protein in the replicase complex, either for unwinding a double stranded replicative intermediate or unwinding of extensive secondary structure in the ss genomic RNA molecule (Kamer and Argos, 1984; Habili and Symons, 1989; Palukaitis *et al.* 1992).

RNA 2.

RNA 2 encodes a single polypeptide of M_r 94kDa (Table 1.1), which has also been implicated in viral replication (Hayes and Buck, 1990 a&b). The sequence contains several motifs common amongst other Bromoviruses as well as in the 183K protein of TMV, including the highly conserved amino acid motif glycine-aspartic acid-aspartic acid (GDD), associated with viral replicase proteins (Kamer

and Argos, 1984; Habili and Symons, 1989; Palukaitis *et al.*, 1992). The viral replicase is able to accept messenger sense RNA and synthesise double-stranded replicative intermediary RNA and genomic RNA for encapsidation. Despite the homology with other replicase associated proteins at the amino acid level, the replicase is highly template specific, accepting only CMV as a template (Palukaitis *et al.*, 1992).

The production of a new CMV protein (2b) has recently been described by Ding *et al.* (1994). They determined the complete nucleotide sequence of CMV-Q RNA 4A and found it to be a novel subgenomic RNA derived from RNA 2. The 2b open reading frame of 100 codons overlaps the C-terminal portion of the 2a gene. RNA 4A is 682 nucleotides long and is 99.3% homologous to the terminal 682 bases of RNA 2 of CMV Q, a subgroup II strain. Sequence comparison to other published sequences revealed the presence of this extra ORF in CMV strains Y and Fny, both subgroup I strains, as well as in the cucumoviruses TAV and PSV J. The sequence was however found to be absent in bromoviruses and alfalfa mosaic virus, indicating that this is a unique feature of cucumoviruses (Ding *et al.*, 1994).

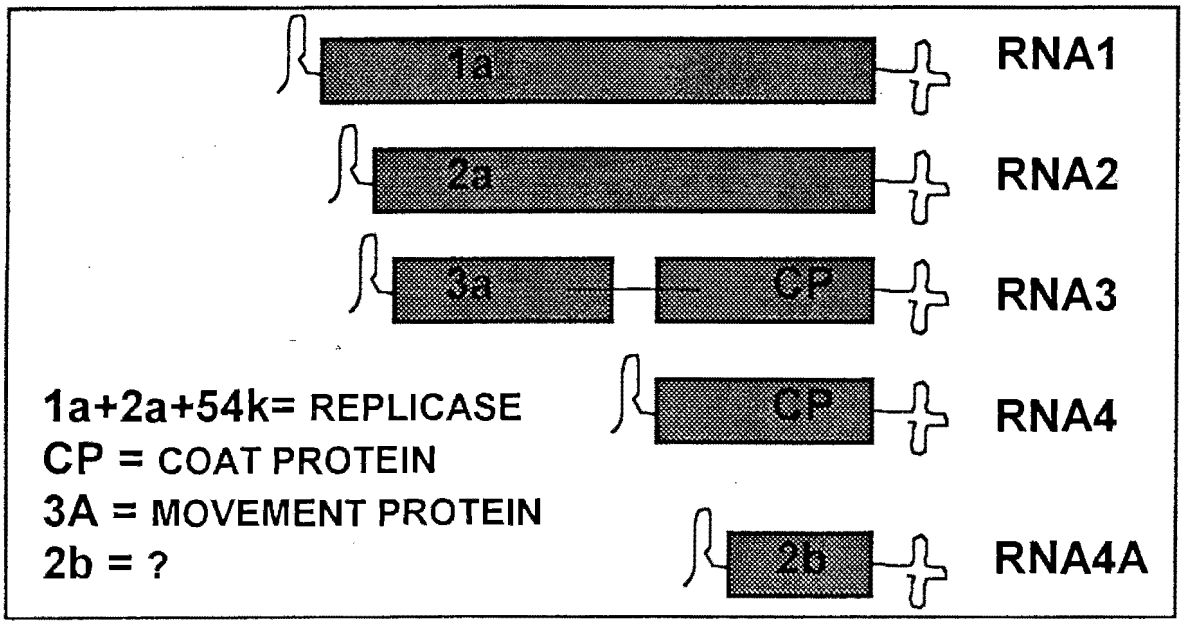

Figure 1.2. DIAGRAM OF THE GENOME STRUCTURE OF CMV.

RNA 3 AND 4.

RNA 3 of CMV contains two open reading frames (Fig 1.2). The 5'-proximal ORF encodes a protein that has been shown to be involved in movement of the virus in the plant host, the 3a protein (Davies and Symons, 1988; Palukaitis *et al.*, 1992). The second ORF encodes the 24kDa coat protein. It is translated from the subgenomic RNA 4 which is colinear with the 3' terminal end of RNA 3 (Fig 1.2). The nucleotide sequence between CMV subgroup I RNA 3s is very highly conserved in isolates from different countries and passage histories, indicating a possible common viral function for RNA 3 amongst these viruses. This conservation of sequence is not extended to subgroup II CMV strains which display considerable divergence to CMV subgroup I types (Owen *et al.*, 1990). The divergence of nucleotide sequence among other members of the Bromoviridae is not reflected at the level of secondary structures where the conservation suggests a functional role for the 5' and 3' UTRs as well as the 3a protein (O'Reilly *et al.*, 1991).

Suzuki *et al.* (1991) constructed cDNA mutants from full-length CMV Y RNA 3 to have deletions in either the 3a or coat protein sequences. These mutants were able to replicate in tobacco protoplasts but were unable to induce wild type systemic symptoms on tobacco when inoculated together with infectious transcripts of RNA 1 & 2. Their results suggested that both the 3a protein and the coat protein are involved in virus transport in the host plant tissues. Subsequent studies on CMV Kin strain have proved that the CMV 3a protein is involved in cell-to-cell trafficking by increasing the size of the plasmodesmata (Vacquero *et al.*, 1994; Ding *et al.*, 1995).

RNA 3 has recently been implicated in inducing a hypersensitive response in *Arabidopsis thaliana*. CMV Y causes necrotic lesions on *A. thaliana* ecotype C24 whereas CMV O spreads systemically in this host. In plants infected with CMV Y, the virus was confined to the inoculated leaves and levels of pathogenesis-related-1 protein (prp-1) were found to be raised in the regions surrounding CMV

induced necrotic lesions. CMV O spread systemically in *A. thaliana* plants, but did not show raised levels of prp-1. Comparison of the host responses to pseudorecombinants between CMV Y and CMV O showed that the prp-1 inducing element was confined to RNA 3 (Takahishi *et al.*, 1994).

The coat protein has been identified as the chlorosis-inducing domain of CMV. CMV Fny causes a green mosaic pattern on tobacco while CMV M induces a bright yellow/white chlorosis. Recombination between cDNA clones of the RNA 3 from these two strains localised the chlorosis-inducing region to a section of the coat protein gene containing two nucleotide differences(Shintaku *et al.*, 1992). These authors went on to use site directed mutagenesis to demonstrate that the secondary structure at these amino acids was the determining factor in chlorosis induction and not the nucleotide sequence as such.

5' TERMINAL STRUCTURE.

Structural features at, or near, the 5' ends of many viral RNAs are thought to play an important role in viral replication. The first 150 base pairs of each CMV RNA could form a strong stem loop structure, similar to that found in BBMV (Pogany *et al.*, 1994) and which appears to play an important role in the replication of beet necrotic yellow vein virus, another ssRNA virus (Gilmer *et al.*, 1993).

The presence of a 5' 7-methyl guanosine cap structure enhances the stability and infectivity of BMV transcripts (Ahlquist *et al.*, 1987; Hayes & Buck, 1990b). The yield of progeny viral RNA from inoculated cDNA decreases seriously with the addition of extra, non-viral bases at the 5' end (Janda *et al.*, 1987; Hayes and Buck, 1990b). The mechanism of this decrease in infectivity is not known. It is suspected that additional 5' nucleotides result in extra bases at the 3' terminus of the (-) strand and that this interferes with the initiation of (+) strand synthesis by the viral replicase (Ahlquist *et al.*, 1992).

THE INTERGENIC REGION.

CMV carries a third untranslated region on its RNA3, the intergenic region (IGR) as illustrated in figure 3.4. This region has been found to play an important role in the replication of BMV RNA 3. (Marsh *et al.*, 1991). Deletion of sections of the internal control region of BMV decreased RNA 3 replication in protoplasts to less than 15% of the wild type levels (Pogue *et al.*, 1992).

Boccard and Baulcombe (1993) were able to demonstrate that mutants in the CMV RNA 3 of the Kin strain that resulted in truncated coat protein being produced, were still able to replicate to levels of 20% of the wild type virus. Mutants containing similar deletions in the coat protein **and** in the intergenic region were however unable to replicate at all. The orientation of the inserted intergenic region was also important as insertion of the IGR in the incorrect orientation did not restore the replication ablilty to the CP-IGR deleted RNA 3. By constructing various deletion mutants of the IGR, they were able to show that a 250-nucleotide region in the intergenic region was required for CMV RNA 3 accumulation. This region contains a conserved UA tract also found in other subgroup II CMVs (Boccard and Baulcombe, 1993). A second conserved motif in the IGR has a similarity to the ICR2 motif found in the RNA polymerase III promoter of eukaryotes. This motif is also found in the 5'-end of BMV RNAs 1 & 2 (Marsh and Hall, 1989). The IGR appears to affect (+)-strand accumulation in both CMV and BMV (Pogue *et al.*, 1992; Boccard and Baulcombe, 1993) and that the ICR2-like motif is necessary for efficient synthesis of RNA 3. The presence of these ICR2-like regions suggests that a host RNA polymerase III subunit and/or one of its cofactors could participate in viral RNA replication. However, none of these components has a molecular mass close to that of one of the host factors associated with the polymerase of CMV or BMV (Zaccomer *et al.*, 1995).

3' -TERMINAL STRUCTURE.

All three CMV RNAs have a conserved, untranslated 3' terminal region of about 200 nucleotides. This common sequence takes on one of two structures depending on the presence or absence of Mg^{2+}. If Mg^{2+} is absent, the 3' terminus takes on a tRNA-like structure which in turn can be charged with tyrosine by plant derived amino acyl tRNA synthetases (Palukaitis *et al.*, 1992; Rybicki, 1995). The presence of Mg^{2+} destabilises the hairpins to form a pseudoknot (Palukaitis *et al.*, 1992).

The addition of 6 to 33 extra non-viral nucleotides at the end of the 3' terminus of BMV does not appear to affect infectivity of transcripts (Ahlquist *et al.*, 1987; Suzuki *et al.*, 1991). These transcripts would not have been able to be aminoacylated and yet were equally infectious in protoplasts as transcripts lacking the 3' extensions. This would suggest that aminoacylation is not required in the initial stages of infection (Ahlquist *et al.*, 1987).

VIRUS REPLICATION.

CMV replication appears to be very similar to other RNA viruses. A general life cycle of the virus can be drawn up by relying on data common to several plant and animal RNA viruses.

Entry.

CMV is transmitted by more that 60 species of aphids in a non-persistent manner (Rybicki, 1995). Recent studies on both potyviruses and cucumoviruses have shown that the efficiency of transmission is primarily a function of the coat protein of the virus (Perry *et al.*, 1994). Chen and Francki (1990) have demonstrated by means of *in vitro* encapsidation studies that the transmissibility phenotype of these viruses is conferred by the coat protein. Once inside the host cell the CMV

must disassemble to release the viral RNA. The exact mechanism of uncoating remains unclear.

Replication.

RNAs 1 & 2 are immediately translated by the host cytoplasmic ribosomes to make proteins 1a and 2a (Palukaitis *et al.*, 1992). These two proteins, together with the 54kDa host protein, form the viral replicase (Hayes and Buck, 1990 a&b), which becomes associated with the genomic viral RNA. It is thought that the replicase recognises the secondary structure at the 3' end of the RNA as this feature is so highly conserved amongst the RNA viruses. This has been demonstrated for BMV where it was found that the 3' terminal structure, common to all the Bromoviridae, contains all the signals required for the initiation of replication (Miller *et al.*, 1986). A complementary RNA of negative sense is

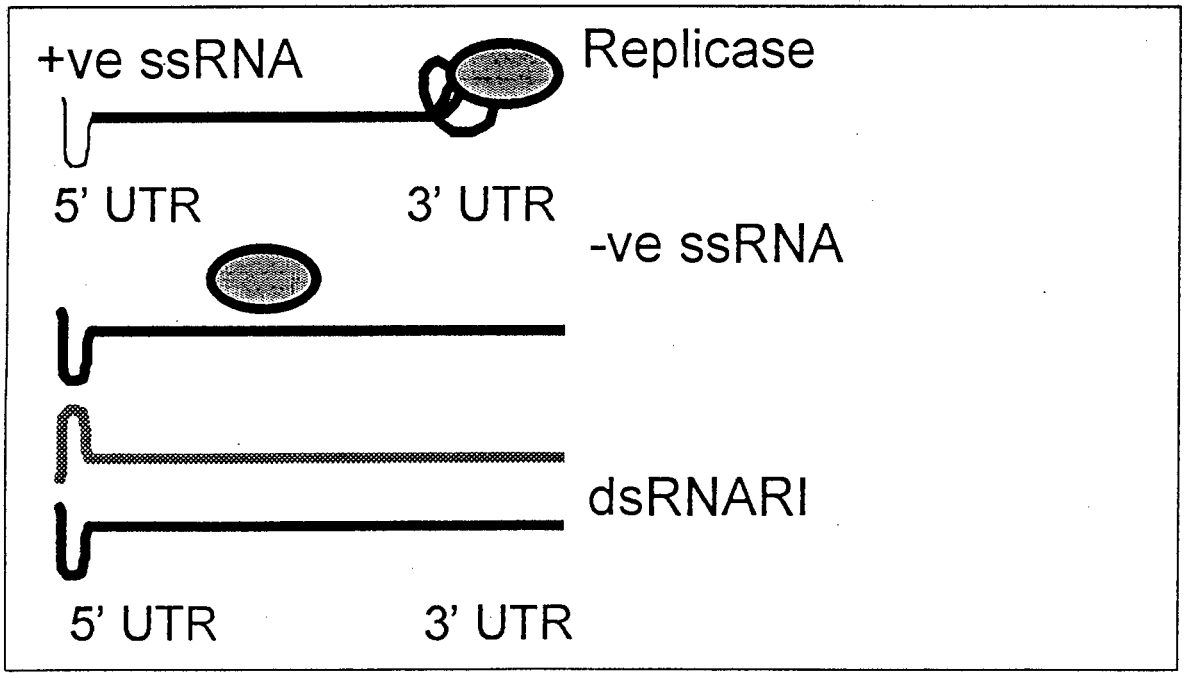

Figure 1.3. Proposed replication strategy of CMV.

The replicase complex recognises the 3' terminus of the (+) genomic RNA. It synthesises (-) sense RNA from the (+) template to make a double stranded (ds) replicative intermediate (RI) RNA.

synthesised off the messenger sense genomic template RNA to form a double stranded (ds) replicative intermediate (RI) (Fig. 1.3).

At some point the a switch occurs and the replicase commences synthesising messenger sense RNA from the negative sense templates. It is thought that the 5' end "breathes" by forming a stem-loop structure in the ds RNA RI complex. The replicase would appear to recognise this loop on the negative strand and synthesise messenger sense genomic RNA from either the ds RI or ss negative sense RNA. This recognition appears to occur more efficiently than at the 3' end structure of the genomic RNA, as the amount of positive sense RNA is usually two orders of magnitude higher than negative sense RNA. The intergenic region of RNA 3 is also capable of forming a stemloop structure (Fig. 1.4). This feature may be even more efficient at capturing the replicase complex as more of the subgenomic RNA 4 is made than any other RNA in the infection process.

Mutations of the 2a protein of BMV have altered the ratio of (+) to (-) strand RNA synthesis (Kroner *et al.*, 1989). However studies on BMV replication in protoplasts suggest that the coat protein may be the regulator of (+) strand synthesis as the levels of (+) and (-) RNAs 1 & 2 are nearly identical in the absence of RNA 3. Virus formation could remove (+) RNA from the replicating pool thereby favouring its formation to maintain an equilibrium. Marsh *et al.* (1991) implicated sections of the intergenic region as determinants of asymmetric replication in BMV.

More research is needed before a clear picture of the replication of CMV can be described. What is clear from the present literature is that there is no single mechanism of regulation of synthesis of the various (+) and (-) viral RNAs.

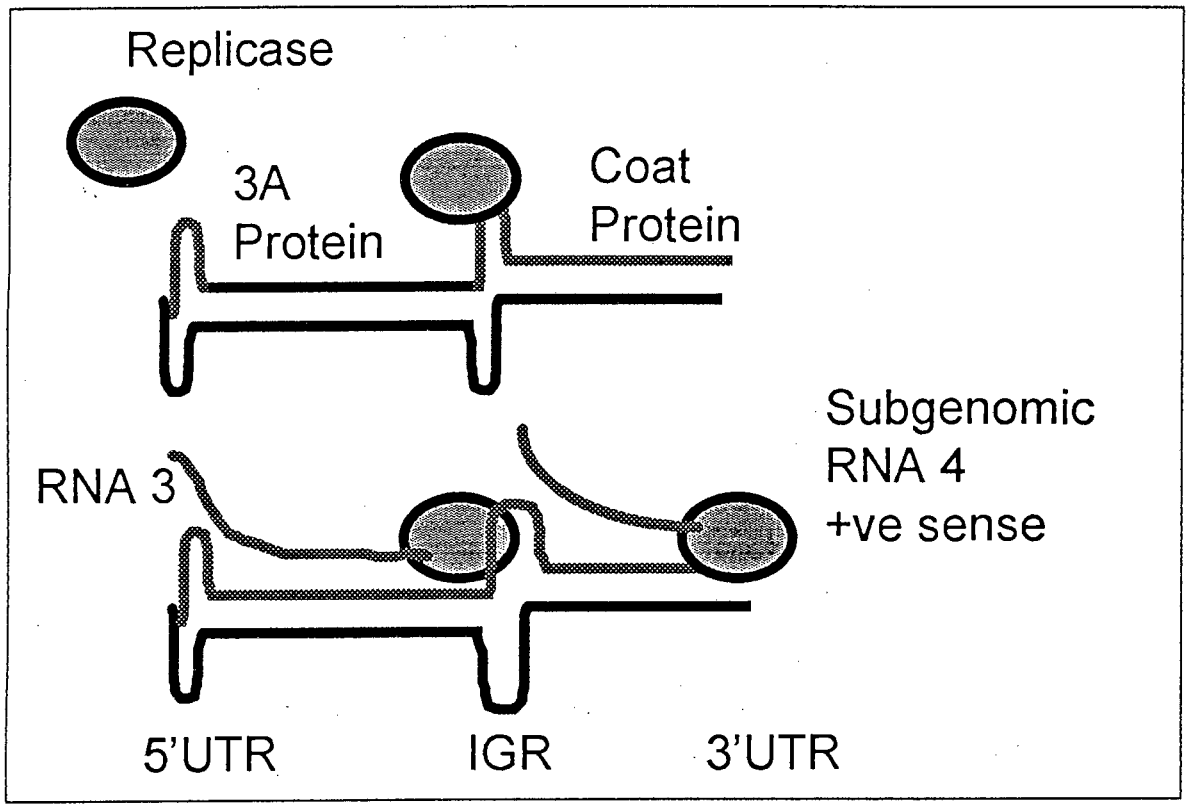

Figure 1.4. Proposed replication strategy of CMV (cntd.)

The 5' end and the intergenic region of the RI is proposed to form a stem-loop which is recognised by the replicase complex. (+) sense RNA is synthesised from the (-) strand of the RI.

DISCUSSION

The production of foreign proteins in plants is fast becoming routine for a number of crop species. The process of generating transgenic plants expressing the gene of interest typically depends on the transformation of plant leaf discs, protoplasts or callus tissue with cloned DNA, selection for transformation by expression of a marker gene and regeneration of the material into whole plants. Transformed regenerants must be allowed to set seed before enough plants can be obtained to allow any reliable estimates of the yield of foreign protein. The levels of expression of the transgene are generally unpredictable, depending on transgene copy number and site of integration into the plant genome (Finnegan & McElroy, 1994). This can largely be compensated for by regenerating many independent transformants and screening for the desired trait. These procedures typically take weeks and even months in some crop species. Although there are several cases of stable expression, it would appear that a large number of transgenes are inactivated by the host; a phenomenon known as gene silencing (Finnegan & McElroy, 1994). The integration of foreign DNA into the plant genome is often unstable following propagation, leading to the loss of the newly aquired gene (Finnegan & McElroy, 1994).

An alternative approach to the expression of foreign genes in plants has been developed over the last couple of years which has none of the drawbacks associated with the plant transformation system. It is now possible to engineer simple plant viruses so that they will express foreign genes at high levels in infected plants (Kumagai *et al.*, 1993; Hamamoto *et al.*, 1993; Thanavala *et al.*, 1995), at a time and plant developmental stage of the experimenter's choosing in as wide a range of crops as are infected by the virus. The major advantage of this system is that it is far easier to manipulate a small viral genome than it is to manipulate a plant genome. A segmented RNA virus such as CMV offers a more simplified approach as its largest genome segment is 3.3kb compared to TMV, another RNA virus with a genome of 6.3kb. A distinct advantage of virus-mediated gene expression is that the engineered product may be tested

immediately it is made by simply infecting the plant. This eliminates the weeks and months required to generate transgenic plants to express the same protein. The system is easily scaled up to increase the amount of protein produced by merely increasing the number of plants inoculated with the modified virus (Donson *et al.*, 1991). This system also eliminates the use of selectable markers such as antibiotic resistance genes, as if the virus infects, the protein will be expressed as a matter of course. A further advantage is that it is of no consequence if the protein is lethal to the plant: as long as the plant tissue is harvested before it dies, the product is still extractable.

Although transgenic plants *per se* are not as flexible a system, they could be used in conjunction with viral vectors to limit viral escape. The viral replication functions could be supplied *in trans* in the host plant genome (Stanley, 1993) and the gene to be expressed would be inserted in place of the excised gene. The deletion of such an essential gene from the viral genome would cripple the virus and hopefully prevent its escape into the environment . This process could also be applied to DNA viruses (caulimoviruses and geminiviruses) to overcome the size constraint on inserts found in these viruses (De Zoeten,1989; Stanley, 1993). As mentioned previously, DNA vectors based on geminiviruses are difficult to introduce into their host plants (Stanley 1993; Shen and Hohn, 1994). DNA viruses would appear to be ideal for expressing foreign genes in plants as the viral genome could be directly manipulated without having to make a complementary DNA copy of the viral genome first, as is the case with RNA viruses.

Viruses can also be used not only to synthesise foreign proteins, but as epitope presentation systems. CpMV, another segmented RNA virus, has been sucessfully used to express two epitopes; one from HRV-14 and another from HIV, on the surface of its capsid (Porta *et al.*, 1994). This yet again demonstrates the flexibily of an RNA viral expression system as opposed to expressing the gene of interest in a transgenic plant.

OBJECTIVES OF THE PROJECT

As can be seen from the foregoing, there are numerous ways of expressing foreign proteins in plants; from creating a transgenic plant to express the desired protein, to having a modified virus synthesise it. The objective of this project was to express a protein constitutively in a tobacco plant at a low level of expression and then show increased expression utilising a plant expression vector. The transgenic plant would act as a suitable positive control for subsequent analysis of the vector product at the protein and RNA levels.

The L1 capsid protein of HPV was chosen as it has not been previously expressed in a plant system and a DNA clone of it was readily available. The L1 capsid protein of HPV is notoriously difficult to express in *E. coli*, although it has been expressed in the correct conformation in human *cos* cells and also in insect cells (Hines *et al.*, 1994). The potential also exists to use this protein as a human vaccine to HPV as animals vaccinated with L1 of related papillomaviruses have been protected against subsequent challenge by the wild type virus (Jarret *et al.*, 1991; Campo *et al.*, 1993). The success of the tobacco expressed LT-B in generating a suitable antibody response in mice after oral vaccination (Haq *et al.*, 1995), bodes well for oral immunisation of individuals utilising the L1 of HPV-16.

CMV was chosen for modification into a viral vector as a near relative, BMV has been shown to be effective in expressing a foreign protein in tobacco protoplasts (Mori *et al.*, 1993). The RNA 3 molecule was chosen as the potential exists to insert two genes and still maintain the size of the original genome fragment. It also exhibits a greater host range than BMV, thereby increasing its appeal as a broad host range expression vector. Once the efficacy of the vector had been demonstrated using reporter genes, the levels of expression of the L1 protein could be compared to the constitutive levels in the transgenic plant.

CHAPTER 2

EXPRESSION OF HUMAN PAPILLOMAVIRUS MAJOR CAPSID PROTEIN IN TRANSGENIC TOBACCO

INTRODUCTION.

The past decade has brought about large advances in plant tissue transformation techniques. It is now relatively simple to insert and express foreign genes in plant cells, which can subsequently be regenerated to mature plants capable of expressing the gene and passing it on to the next generation in a Mendelian fashion (Von Bodman *et al.*, 1995).

One of the most studied recent examples of the expression of a foreign antigen in transgenic plants has been the insertion of the gene encoding the hepatitis B surface antigen (HBsAg) into tobacco plants (Mason *et al.*, 1992). Partially purified HBsAg was used to immunise mice intraperitoneally (Thanavala *et al.*, 1995). Immunisation of mice with plant extract containing the antigen induced an immune response similar to the one obtained by immunising mice with yeast derived HBsAg (Thanavala *et al.*, 1995). T-cells isolated from mice immunised with the tobacco-derived HBsAg could be stimulated *in vitro* with the yeast-derived antigen. These authors were able to conclusively prove that both the B- and T-cell epitopes of the HBsAg were preserved in their transgenic plants.

An example of the success of oral immunisation is found in a paper by Haq *et al.* (1995) where crude soluble extracts from transgenic tobacco and potato plants expressing the *Escherichia coli* heat-labile enterotoxin (LT-B) were used to immunise mice by gavage. Gavage entails force-feeding the mouse by means of a tube inserted down the oesophagus to the stomach. The mice developed both

serum and gut mucosal anti-LT-B immunoglobulins that neutralised the enterotoxin in cell proliferation assays (Haq *et al.*, 1995).

Papillomaviruses are double stranded DNA viruses that cause cutaneous warts and mucosal condylomata in a wide variety of animals. Human papillomaviruses (HPV) display specific tissue tropisms (Zur Hausen and De Villiers, 1994). Benign condylomata in the cervical mucosa often contain low-risk HPV types such as HPV-6 and HPV-11, whereas cervical carcinomas most frequently contain high-risk HPV types such as HPV-16 and HPV-18 (Cason *et al.*, 1993; Hines *et al.*, 1994).

Antibodies generated in rabbits against the major capsid protein (L1) of a virus related to HPV, cottontail rabbit papillomavirus (CRPV), provided the rabbits with protective immunity against subsequent infections with the wild type virus (Lin *et al.*, 1992), despite low titres of neutralising antibodies. Similar results were obtained by a group working on bovine papillomavirus (BPV) L1 protein inoculated into calves. These animals were able to generate serum neutralising antibodies to BPV when reinoculated with the BPV L1 antigen (Jarrett *et al.*, 1991; Campo *et al.*, 1993). Breitburd *et al.* (1995) tested the ability of vaccination with virus-like particles (VLPs) ressembling CRPV to protect rabbits against papillomas induced by the wild type virus. None of the rabbits vaccinated with CRPV VLPs and subsequently challenged with CRPV particles developed cancer within 1 year of infection. No protection was however detected in rabbits vaccinated with native or denatured BPV L1 or with denatured CRPV L1.

Both these studies indicate that the L1 capsid protein of HPV could be important in the development of a vaccine for human papillomavirus. L1 protein has been previously expressed in human *cos* cells under the control of the SV40 promoter. The recombinant L1 protein was of the appropriate size (55kDa) and formed the correct conformational epitopes necessary for the formation of a neutralising antibody response (Ghim, *et al.*, 1992). Kirnbauer *et al.* (1992) were able to express L1 capsid protein of both BPV and HPV-16 in insect cells via a

baculovirus expression vector. The L1 proteins were expressed to high levels and assembled into capsid-like particles. The self-assembled BPV L1 was able to generate a higher titre of neutralising antibodies in rabbits than the L1 extracted from recombinant bacteria. Hines *et al.* (1994) expanded on the strains tested by expressing the L1 proteins of HPV-1, 6, 11 & 16 in both *cos* cells, under the SV40 promoter and in a baculovirus expression system. The recombinant L1 proteins also expressed conformational epitopes in both the *cos* and insect cells that were type specific and displayed neutralising epitopes (Hines *et al.*, 1994).

The success of plant-derived hepatitis B surface antigen and the heat labile enterotoxin of *E. coli* in providing a protective immune response after oral vaccination with the protein extracts containing the antigens is encouraging for the development of oral vaccines. We decided to try and express another antigen that holds promise as an oral immunogen: the L1 capsid protein of HPV-16 (Cason *et al.*, 1993) in transgenic tobacco plants. Expression of L1 protein in transgenic tobacco plants would give an indication of the basal level of expression in the plant. Once the viral vector had been successfully constructed, the gene encoding the L1 protein could be inserted into the vector and the two levels of expression be compared.

MATERIALS AND METHODS.

DNA cloning procedures.

pSK-L1 was kindly donated by W. Burgers and Dr.A. Williamson (Medical Microbiology, University of Cape Town). The L1 open reading frame was excised as a ClaI/XbaI fragment and ligated into ClaI/XbaI cut pART7 as shown in figure 2.5 (Gleave, 1992). This placed the L1 ORF downstream of the CaMV 35S promoter in the pART7 primary cloning vector. The orientation of the insert was confirmed by restriction enzyme analysis. The CaMV 35S promoter, L1 ORF and *ocs*3' region (transcription termination region of the octopine synthase gene) cartridge was excised as a NotI fragment and ligated into NotI digested binary vector, pART27 (Gleave 1992). pART27 carries the RK2 minimal replicon for maintainance in *Agrobacterium*, the ColE1 origin of replication for

high-copy maintainance in *E. coli* and the Tn7 spectinomycin/streptomycin resistance gene as selectable marker. The T-DNA of pART27 carries the kanamycin resistance gene which is expressed in both *E. coli* and *Agrobacterium.* The presence of the insert in pART27 was confirmed by restriction enzyme analysis.

Agrobacterium transformation.

Agrobacterium tumefaciens C58C1(pMP90) was transformed with the L1-pART27 clone by the freeze-thaw method of Hooykaas (1988). *Agrobacterium* was grown to an OD of 0.4 in 5ml Luria broth containing gentamycin (40µg/ml) and rifampicin (100 µg/ml). All the following steps were done on ice or at 4°C. The cells were centifuged at 3000g for 5 minutes and resuspended in 1ml ice-cold 20mM $CaCl_2$. Aliquots of 0.1 ml were dispensed into precooled eppendorf tubes and 1µg of plasmid DNA was added to the cells. The tubes were frozen in liquid nitrogen and then heat shocked at 37°C for 5 minutes. 1ml Luria broth was added to the transformed cells and they were then left to shake at 30°C for 4 hours. The cells were then plated onto Luria agar plates containing spectinomycin (100 µg/ml) and rifampicin (100 µg/ml). *Agrobacterium* transformants were obtained after 2 days growth at 30°C. The clones were checked for inserts by polymerase chain reaction utilising primers L1F (5'-CCCGGGATGT CTCTTTGGCT GCCTAG-3') and L1R (5'-GCCTCGAGT TACAGCTTAC GTTTTTTGC-3') on samples picked from the plates with a plastic pipette tip. PCR reaction conditions were the same as in Appendix I except that the cycles were: 1 minute @ 93°C; 1 minute @ 55°C and 1 minute @70°C for a total of 25 cycles.

Plant transformations.

Cut pieces of sterile *N. tabacum* cv. Petit Havana SR1 were transformed with the modified *Agrobacterium* containing the L1-protein gene as described by Horsch *et al.* (1985). Transformed calli were selected on regeneration medium supplemented with kanamycin at a concentration of 100µg/ml. Resistant calli were induced to regenerate shoots and roots, transferred to soil and maintained in a greenhouse (Horsch *et al.*, 1985).

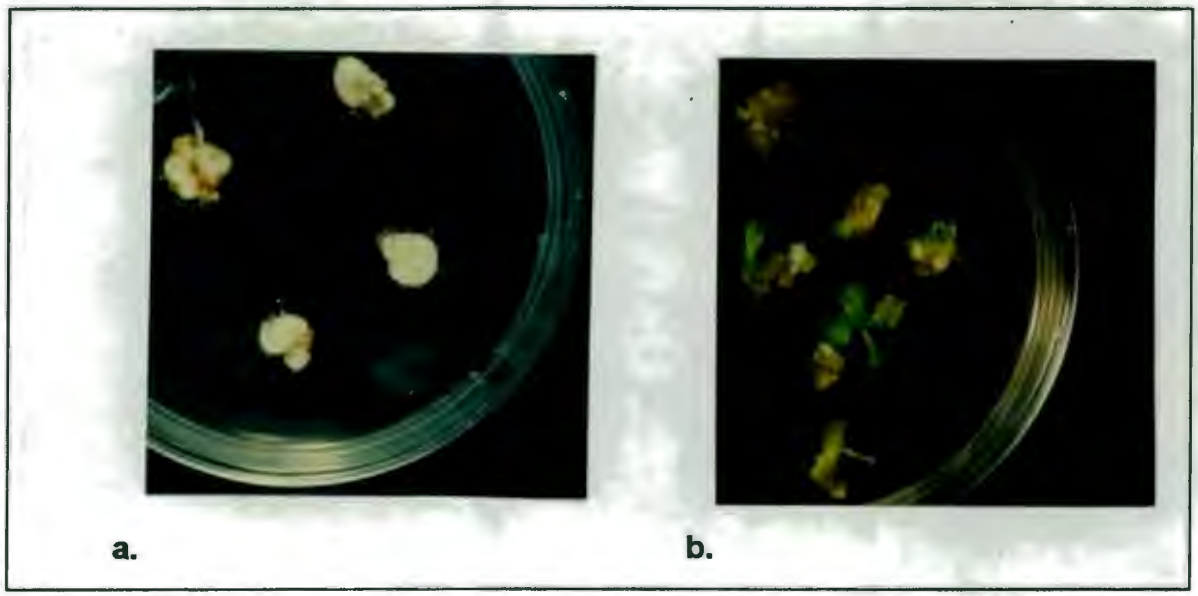

Figure 2.1. Leaf transformations utilising *Agrobacterium* transformed with L1-pART27. (a) Untransformed leaf discs that are not kanamycin resistant. Leaf discs become pale and die off. (b) Transformed leaf discs showing callus formation. The untransformed sections of the leaf discs die off but the kanamycin resistent calli continue to proliferate.

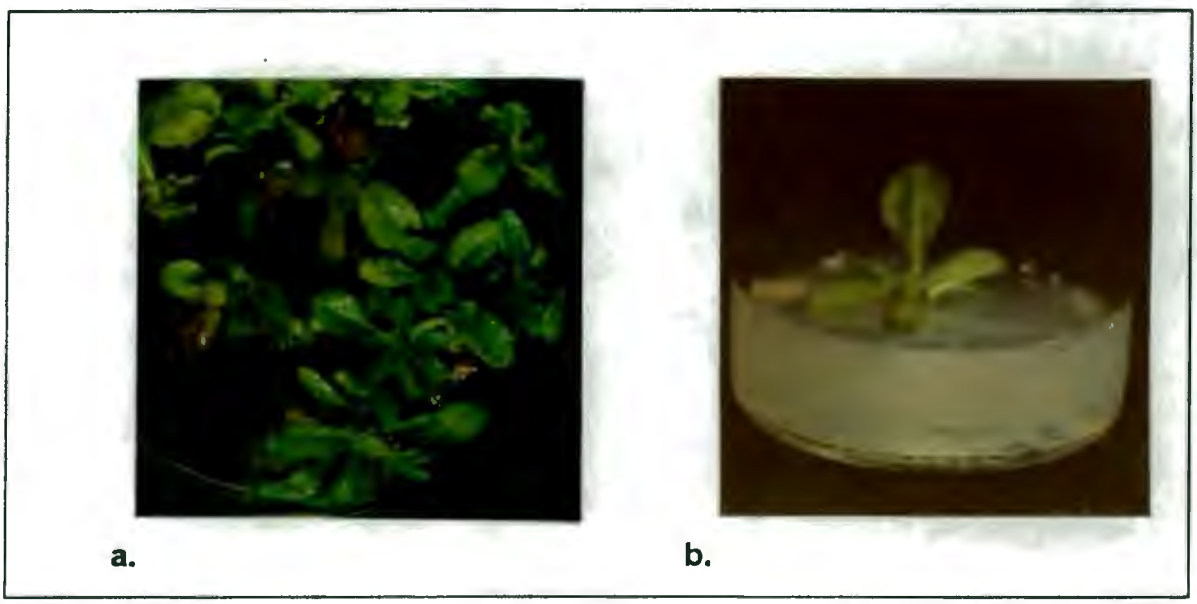

Figure 2.2. Plant regenerations. (a) Kanamycin resistant calli start to form shoots 30 days after callus formation. (b) Shoots transferred to rooting medium containing kanamycin form roots after 4-7 days.

DNA and RNA isolations from transformed plants.

Both DNA and RNA were isolated from a 300mg leaf sample utilising the method of (Nelson, 1994). Briefly, the leaf was ground up in liquid nitrogen and 500µl of lysis buffer (100mM Tris-HCl, pH 8.6, 1% sarkosyl, 4M guanidium thiocyanate, 25mM EDTA pH 8, 25mM EGTA pH 8, 100 mM β-mercaptoethoethanol) was added to the sample. 250µl phenol was added and the sample shaken vigorously. 250µl chloroform was added, the sample was mixed well and centrifuged at 10 000 rpm for 10 minutes at 4°C. The SNF was once more extracted with phenol:chloroform. 1µl of concentrated acetic acid was added to the SNF of the second extraction. 0.6 volumes of isopropanol was then added to the treated SNF and the samples were mixed and stored at -20°C for 2 hours. Samples were then allowed to thaw and were centrifuged at 4°C at 10 000 rpm for 10 minutes.. The pellet was resuspended in 300µl DEPC-treated water containing 0.1% SDS by pipetting up and down and incubating the tubes at 37°C for 10 minutes. The remaining insoluble material was removed by a brief centrifugation in the microfuge. The SNF was transferred to a new tube and 0.25 volumes of 8M lithium chloride was added to the sample. The RNA was precipitated overnight at 4°C. The RNA was pelleted at 10 000 rpm for 30 minutes at 4°C and the SNF decanted to a fresh tube for DNA recovery. The RNA pellet was resuspended in 20µl of DEPC-treated water.

The SNF from the RNA pelleting step was mixed with a $1/10^{th}$ volume of 3M NaAc pH4.5 and 2 volumes of 100% ethanol. Samples were left at -20°C for an hour before centrifugation at 10 000 rpm for 30 minutes at 4°C. The DNA pellet was washed in 70% ethanol and resuspended overnight in 20µl water.

DIG-labelling of the L1 DNA probe.

pSKL1 was digested with SmaI. The L1 fragment was excised from a 1xTAE agarose gel stained with 0.02% methylene blue (Sambrook *et al.*, 1989) and recovered utilising the Geneclean II Kit (BIO 101 Inc.). The fragment was labelled by random priming using DIG-labelled-dUTP as per Boehringer-Mannheim protocol.

PCR screening of L1 transgenic plants.

A 500ng amount of total plant DNA extracted above was used in a polymerase chain reaction to detect the L1 gene. Reactions were done as in Appendix I utilising primers L1F and L1R and the cycling conditions used above.

A volume of 10µl of each sample was loaded on a 1% 1xTBE agarose gel for electrophoresis as shown in figure 2.3 (Maniatis *et al.*, 1989). The gel was blotted, probed with the L1 DIG-labelled probe and detected as discussed in Appendix F (Fig 2.3).

Total protein isolations from L1 transgenic plants.

Leaf material weighing about 300mg was ground up in 400µl extraction buffer (0.25M Tris-HCl pH 7.8, 5mM EDTA, 1% β-mercaptoethanol) and centrifuged at 4°C for 15 minutes. 350µl supernatant fluid containing soluble proteins was drawn off and stored at -20°C. 100µl 2x disruption buffer (0.25M Tris-HCl pH 6.8, 30% v/v glycerol, 20% SDS, 29% β-mercaptoethanol, 0.01% w/v bromophenol blue) was added to the remaining SNF and pelleted debris. Samples were then boiled for 15 minutes and allowed to cool to room temperature. These samples contained both soluble and insoluble proteins.

Western blot analysis of L1 transgenic plants.

A single well on an SDS-polyacrylamide gel (5% stacking gel and 12% resolving gel) was loaded with 50µl of the soluble and insoluble protein extraction mix as per Sambrook *et al.* (1989). Recombinant L1 protein obtained from a baculovirus expression system was used as a positive control (Rose *et al.*, 1994). The proteins were transferred to a nitro-cellulose membrane using a horizontal blotting apparatus with carbon electrodes. The Boehringer-Mannheim Western Blotting Kit was used to detect L1 protein on the membranes. Rabbit anti-bovine papillomavirus antiserum (DAKO) was used as the primary antibody at a 1/100 dilution.

RESULTS

L1 TRANSFORMATION OF *AGROBACTERIUM.*

The L1 gene encoding the major capsid protein of HPV-16 was cloned into the binary vector pART27. This clone was used to transform *Agrobacterium tumefaciens* C58C1(pMP90). The resultant transformants were screened by PCR as well as by restriction enzyme analysis (results not shown).

L1 TRANSFORMATION OF TOBACCO PLANTS.

N. tabacum cv. Petit Havana SR1 leaf discs were successfully transformed by the L1-*Agrobacterium*. Kanamycin resistant calli were observed 30 days after incubation with the bacterium (Fig. 2.1). Shoots appeared after another 30 days and were transferred to rooting medium 20 days later (Fig. 2.2). Samples were taken for DNA and protein analysis after another 2 months on kanamycin containing medium (100µg/ml).

PCR SCREENING OF L1 TRANSGENIC PLANTS.

A total of 19 kanamycin resistant plants were screened of which 17 were PCR positive for the L1 gene using L1 specific primers (Fig.2.3). Transgenic positive plants gave a product of 1.6kb, the same size as the product obtained from pSK-L1 positive control. Probing with a DIG-labelled L1 DNA probe confirmed that the two negative plants contained no L1 gene and that the PCR fragments were indeed the L1 amplified gene (Fig.2.3).

WESTERN BLOT ANALYSIS OF L1 TRANSGENIC PLANTS.

A volume of 25µl total protein extract, representing about 50µg protein of each plant was loaded onto a protein gel. 250ng of recombinant L1 protein (Rose *et al.*, 1994) was loaded onto each gel as a positive control. Each gel also carried a

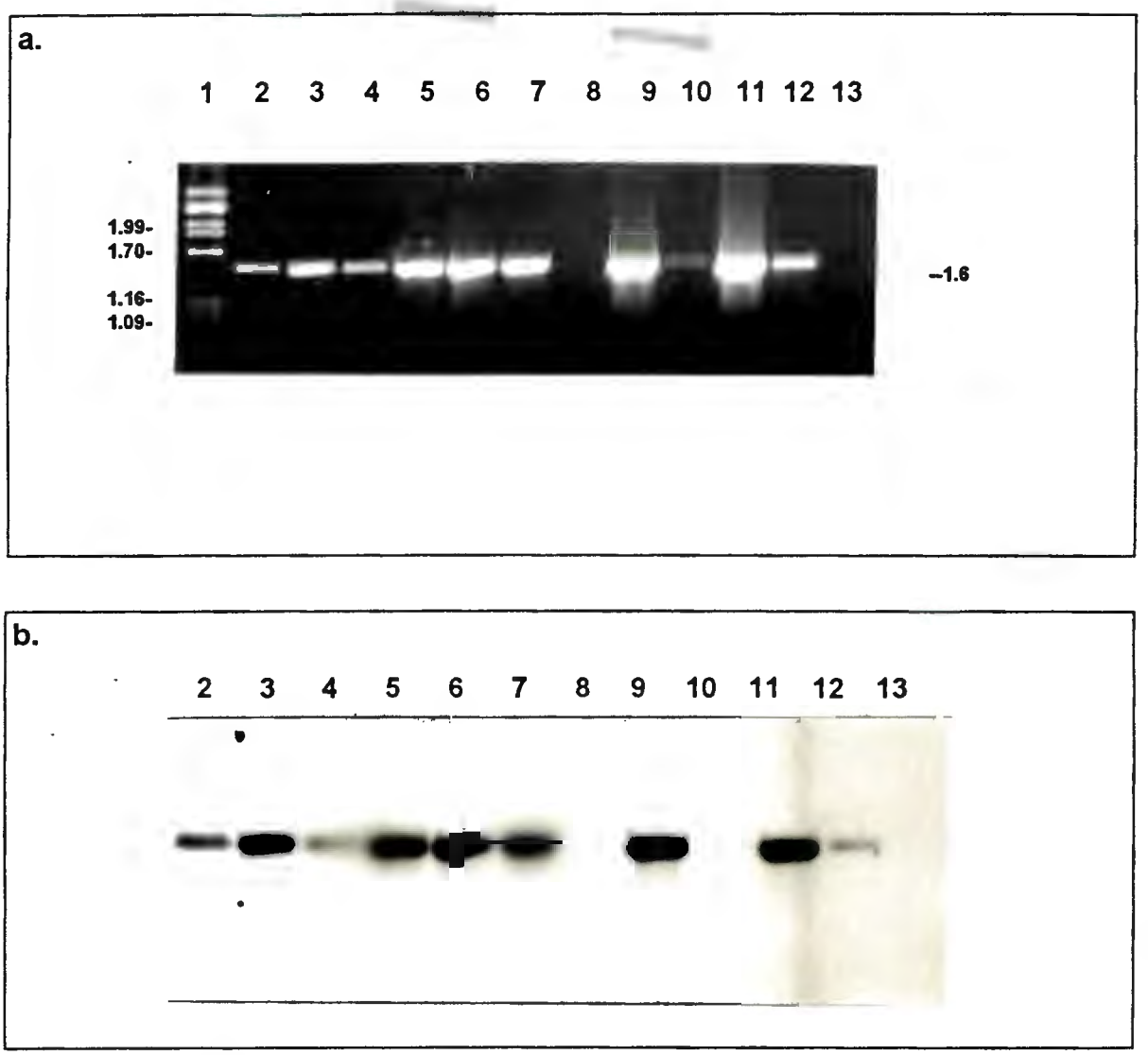

Figure 2.3. PCR analysis of L1 transgenic plants.

(a) Lane 1: λ-DNA cut with PstI as molecular weight markers. The sizes of the relevant fragments are given in kilobases beside the gel. Lane 2-11: Transgenic plants numbers 2, 3, 5, 8, 9, 11, 12, 14, 15 & 16. Only plant # 12 does not carry the L1 gene on this gel. Lane 12: PCR positive control on 100ng pSK-L1. Lane 13: PCR negative control on a non-transgenic tobacco plant DNA sample. (b) Southern blot result of the PCR gel probed with the DIG-labelled L1 probe. Lanes as above with Lane 2 corresponding to transgenic plant #2.

non- transgenic plant protein extract as a negative control. Proteins were transferred to nitrocellulose membranes and incubated with antibody to disrupted BPV. The results of the Western blot are presented in figure 2.4. The L1 positive control did not light up which suggests that the detection system is at fault. No bands were observed at 55kDa which is where L1 is expected to band (Kirnbauer *et al.*, 1992). Any bands observed in this region after longer exposure times were also visible in the plant negative control protein extracts.

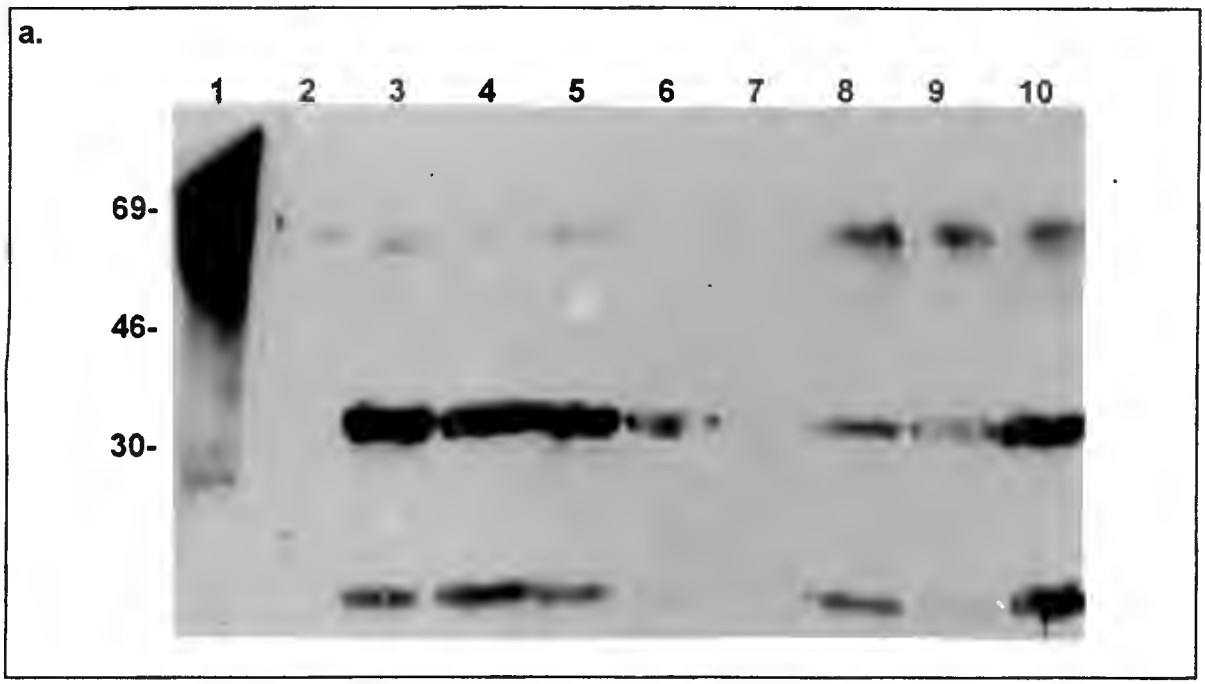

Figure 2.4. Western blot analysis of L1 transgenic plants. (a) Lane 1: Rainbow molecular weight markers (Amersham Life Sciences) Protein sizes are indicated in kDa. Lane 2: 250ng recombinant L1 protein; lane 3: transgenic negative plant protein extract; lanes 4-10: protein extracts #2, 3, 5, 8, 9, 14 & 15 from plants that were PCR positive for the L1 gene.

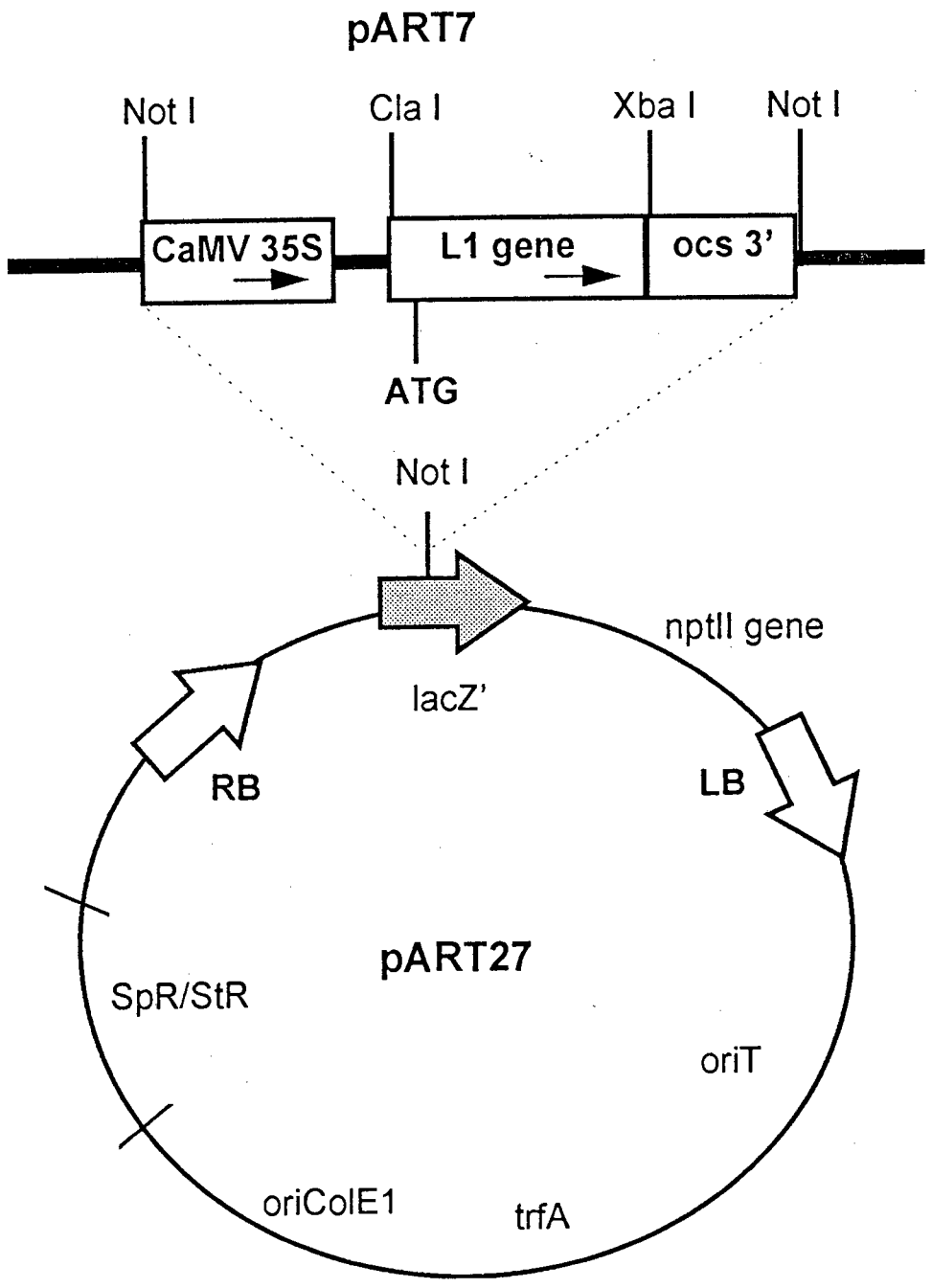

Figure 2.5. Construction of the plant transformation vector L1-pART27. The L1 gene was cloned as a ClaI/XbaI fragment into the ClaI and XbaI sites of the multiple cloning site of vector pART7, placing the gene downstream of the cauliflower mosaic virus 35S promoter (CaMV35S) and upstream of the transcription termination region of the octopine synthase gene (*ocs*3'). This region was excised as a NotI fragment from pART7 and inserted into pART27 at the NotI site, within the *lacZ* region (encoding the lac αpeptide) indicated by the coloured arrowed box. The right border (RB) and left border (LB) are indicated by clear arrowed boxes. The chimaeric *nptII* gene provides kanamycin resistance in both *E. coli* and *Agrobacterium.* SpR/StR is the bacterial selectable marker for spectinomycin/streptomycin resistance. oriT is the RK2 origen of transfer and the ColE1 allows for replication in *E. coli.*

DISCUSSION.

In order to detect L1 protein in total protein leaf extracts from leaves, a suitable antiserum is required. Antibodies to BPV have ppreviously been show to cross-react with denatured HPV-16 L1 protein (Hines *et al.*, 1994). The commercial antibodies used in this study were also made to target disrupted BPV and so should have recognised the disrupted, plant produced HPV-16 L1 antigen. The fact that the anti-BPV serum could not detect the L1 recombinant positive control protein would suggest that there is a problem with the antiserum and a new antiserum may give a better result. The colleagues (Di Marais and Dr. A. Williamson, Medical Microbiology, University of Cape Town) who provided the Rose (1994) recombinant L1 antigen have subsequently been unable to show that it reacts with L1 antiserum, indicating that the positive control antigen may be at fault. It was therefore imperative to demonstrate the presence of L1 specific RNA in plant total RNA preparations and also the integration of the L1 gene into the genome of the tobacco plant.

Northern blots were attempted to determine whether the L1 gene was being expressed, but the results were inconclusive. These blots need to be repeated once more plant material becomes available. If certain plants do show expression of L1 protein, immunohistochemistry could be done on sections of leaf material to determine whether virus-like particles are being formed in the leaf tissue.

In conclusion, 19 kanamycin resistant plants were regenerated of which 17 have been shown to carry the L1 gene in their genetic make-up (Fig. 2.3). Ideally one should do a Southern blot to determine whether the gene is integrated into the plant genome once more plant tissue is available to do a large scale DNA extraction. We were unable to show whether the L1 gene was being expressed and could not detect whether the L1 protein was present in total protein extracts or not. Further analysis of these plants may yield some interesting and insightful results as to the production of viral antigens in plants.

CHAPTER 3

MOLECULAR CLONING OF THE 5' AND 3' UNTRANSLATED REGIONS OF CUCUMBER MOSAIC VIRUS RNA 3.

INTRODUCTION

Recombinant DNA technology has made it possible to analyse viral genomes at the molecular level and study the effects of genetic modifications on their organisation and expression. Studies on RNA virus genomes have previously been hampered by the fact that these viruses do not utilise a DNA intermediate during replication, making investigations of their life cycles difficult. The possibility of obtaining infectious cDNA clones of plus-strand RNA virus genomes has since provided extensive information on replication and movement of these viruses (Boyer and Haenni, 1994). Moreover, these clones can be utilised to construct infectious satellite RNAs for expression of foreign proteins in plants (Donson *et al.*, 1991; Hamamoto *et al.*, 1993; Kumagai *et al.*, 1993 & Mori *et al.*, 1993).

We chose to investigate the potential of cucumber mosaic virus (CMV), as a plant expression vector. CMV has a much broader host range than BMV, another member of the Bromoviridae previously used in plant viral vector studies (Mori *et al.* 1991). CMV infects both monocot and dicot plants, thereby increasing its appeal as a broad host range plant expression vector. The replication functions would be provided to the satellite RNA by co-inoculation of the plants with total infectious CMV RNA extracts from wild type infected plants (Fig 3.1). This procedure would provide us with a rapid means of testing the system without having to construct full-length cDNA copies of RNAs 1 & 2.

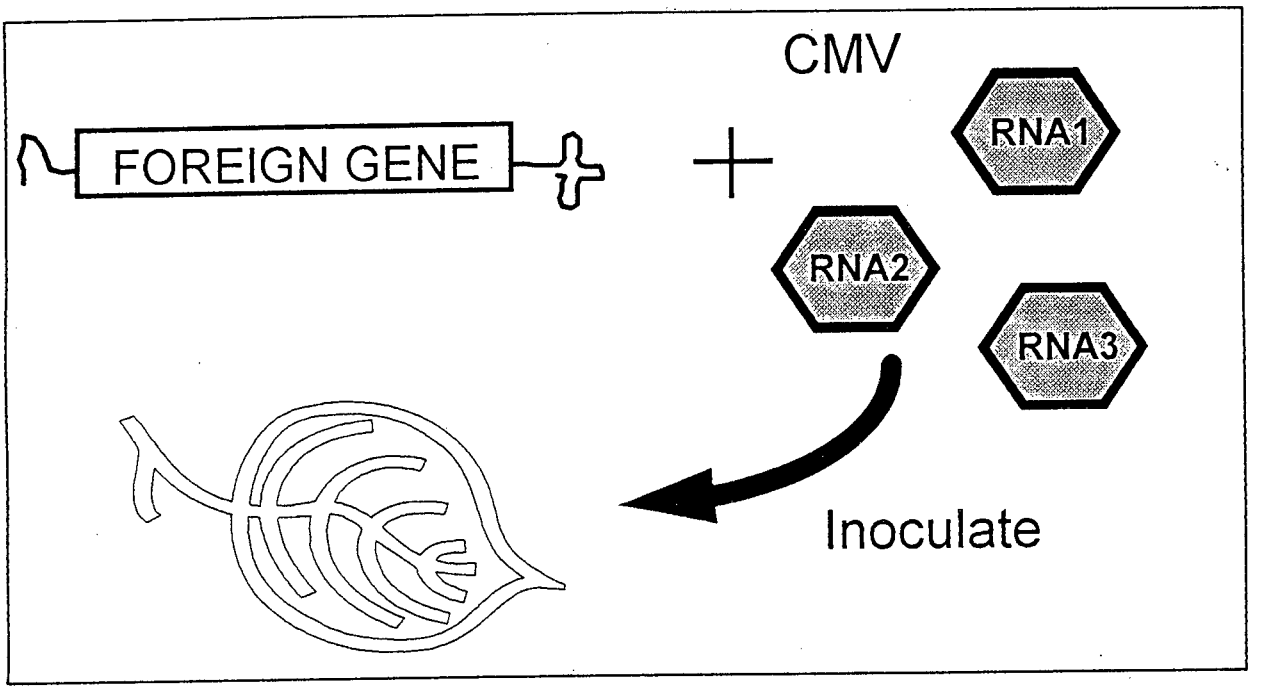

Figure 3.1. Representation of the strategy to express a foreign protein in plants.
A viral-like RNA carrying the gene of interest would be co-inoculated with wild type CMV RNA onto healthy plants and monitored for the production of the foreign protein.

The construction of a full length cDNA clone is a long and tedious process. Generally, the entire sequence is required to produce infectious clones, but infectious transcripts have been produced using incomplete viral cDNA clones (Davis *et al.*, 1989). The presence of non-viral nucleotides at the 5' end of infectious transcripts strongly reduces infectivity (Janda *et al.*, 1987). 3' non-viral extensions do not appear to interfere with viral replication provided the secondary structure of the 3' terminus is not abrogated (Miller *et al.*, 1986).

This chapter describes the cloning and sequencing of the 5' and 3' untranslated regions (UTR) of CMV Y and the insertion of a reporter gene between them to create plasmid p5'cat3'.

MATERIALS AND METHODS.

Virus strains.

CMV-Y was kindly donated for initial studies by Prof. M.B. von Wechmar. This isolate originally was characterised by the late R.I.B. Francki and is a subgroup I CMV member (Palukaitis *et al.*, 1992). CMV Y was propagated on *Nicotiana glutinosa* plants.

Viral double stranded RNA extraction.

Double stranded CMV-Y RNA was isolated from infected *Nicotiana tabacum* cv *xanthi* plants according to the method of Valverde (1990). 5g leaf tissue was ground to a fine powder in liquid nitrogen and added to 8 ml 1xSTE buffer (pH 6.8) in a 50 ml centrifuge tube. 1 ml of 10% SDS, 0.5 ml 2% bentonite and 9 ml of STE saturated phenol was added to the homogenate and the tubes were shaken thoroughly for 30 minutes. After centrifugation at 8000 g for 15 min.(Sorval RC-5 Superspeed), the supernatant fluid (SNF) was transferred to a new 50 ml centrifuge tube and its ethanol concentration adjusted to 16% v/v. SNF was then passed through a cellulose column and then washed with 40 ml 1xSTE containing ethanol at 16% v/v. The ds RNA was eluted by first equilibrating the column with 2.5 ml 1xSTE and washing it out with 10 ml 1x STE. The sample was collected in a 50 ml centrifuge tube, again made up to 16% ethanol v/v and passed through a second cellulose column. This time the sample was eluted in 6ml 1x STE and collected in a new 50 ml centrifuge tube. RNA was precipitated by adding 0.5 ml 3M sodium acetate (pH 5.5) and 20 ml 95% ethanol to the sample and storing it at -20°C overnight. Samples were centrifuged at 8000 g for 30 minutes, the pellets washed in 70% ethanol at -20°C and resuspended in 100µl 1xTE and stored indefinitely at 4°C.

Complementary DNA synthesis.

Four synthetic oligonucleotides were designed based on the previously published sequence of CMV Y (Prof. Rybicki, Microbiology, University of Cape Town). Both oligonucleotides 5vec4 (5'-CCCTCGAGGT AATCTAACCA CCTGTG-3') and 5vec5 (5'-CCCTCGAGGC

CTCGGGAAAT CTAAC-3') were used to initiate first strand cDNA synthesis of the 5' UTR. Two strands of cDNA were synthesised from the ds RNA template used in the reverse transcription reaction (Appendix H). The same primers were used in the ensuing amplification step (Appendix I). Both these primers incorporated an Xho I restriction enzyme site at their 5' ends to facilitate easy cloning of the cDNA fragment. Oligonucleotides 3vec4 (5'-CCGGATCCTT CCCAGAATCC TCCCT-3') and 3vec3a (5'-TGGAGGCCCC CACGAAG-3') were used similarly to obtain a cDNA copy of the 3' UTR. This product had to be amplified a second time with the overlapping 3vec3 oligo (5'-CCGAGCTCTG GTCTCCTTTT GGAGG-3') to produce a full length 311 bp fragment of the 3' UTR. Primer 3vec3 included a Sac I restriction enzyme site at its 5' end while 3vec4 carried a Bam HI site at its 5' end to ease the cloning process. Table 3.1 lists the primer sequences as well as the restriction sites found on each. Figure 3.4 illustrates the PCR strategy employed to amplify the UTRs.

Primer	RE site *	Primer sequence
5vec4	XhoI	5'-CCCTCGAGGTAATCTAACCACCTGTG-3'
5vec5	XhoI	5'-CCCTCGAGGCCTCGGGAAATCTAAC-3'
3vec4	BamHI	5'-CCGGATCCTTCCCAGAATCCTCCCT-3'
3vec3a	none	5'-**TGGAGG**CCCCCACGAAG
3vec3	SacI	5'- CCGAGCTCTGGTCTCCTTT**TGGAGG**-3'

Table 3.1. Table showing the primers used to amplify the 5'- and 3'-UTRs of CMV Y. The sequences of the restriction enzyme sites included in some of the primers are underlined. The PCR strategy is illustrated in figure 3.4. * RE site: restriction enzyme site. Bold bases indicate the 6 base pair overlap between primers 3vec3a and 3vec3.

Attempts were made to obtain a full length cDNA copy of the entire RNA 3 molecule using oligonucleotides 5vec4 and 3vec3a. First attempts were done similarly to the 5' and 3' UTRs (Appendix I) except that the extension time at 72°C in the amplification stage was increased to 180 seconds. Later attempts tried different denaturing conditions, different reverse transcriptase enzymes: Tth polymerase (Boehringer-Mannheim); AMV (Promega) and MMuLV (Promega) and varying amplification conditions.

Complementary DNA cloning.

pCMV5'.

Double stranded DNA of the 5' end was blunted (Appendix Jii) and inserted into XhoI digested and blunted pSK vector (Appendix Ji & iii). This clone contained an extra aberrant amplification fragment of 90 base pairs. The required 132 bp fragment was excised from this clone by utilising the XhoI sites in the oligonucleotides used to amplify the 5' cDNA fragment. The gel-purified 5' XhoI fragment was inserted into the pKS Xho I site to yield pCMV5' (Fig. 3.45. which was subsequently transformed into competent *E.coli* Dh5α cells and sequenced (Appendix Jiv).

pCMV3'.

This clone was constructed by inserting the 3' ds cDNA fragment into SmaI cut and alkaline phosphatase treated pSK (Fig. 3.5) as described in appendix J. This clone was transformed into competent *E.coli* Dh5α cells and sequenced (Appendix Jiv & D).

p5'cat3'.

The 3' UTR was excised from pCMV3' using the BamHI/SacI primer sites and directionally cloned into BamHI/SacI digested pCMV5' to yield clone pCMV5'3' (Fig. 3.5). Oligonucleotides 5vec4 and 3vec3 were used in an amplification reaction to confirm the orientation of the 3' insert.

The *cat* gene conferring resistance to the antibiotic chloramphenicol was excised from pCm1 as a TaqI fragment and inserted into the ClaI site of pCMV5'3' to render p5'cat3' (Fig. 3.5). The orientation of the insert was confirmed by restriction enzyme analysis and end-sequencing across the 5' UTR.

Double stranded RNA infectivity studies.

Double stranded RNA of CMV-Y was heat denatured at 80°C for 10 minutes in a 50% DMSO solution and then placed on ice until inoculation. Denatured ds RNA was inoculated onto 4 celite dusted leaves of *Nicotiana glutinosa* plants at a concentration of 75μg/ml in a

1% bentonite solution. Control plants were similarly inoculated with non-denatured ds RNA. Plants were monitored for symptom development for 22 days post inoculation.

RESULTS:

Double stranded RNA isolations and infectivity studies.

Double stranded RNA was successfully isolated from CMV-Iinfected *Nicotiana tabacum* cv. *xanthi* plants as shown in figure 3.2.

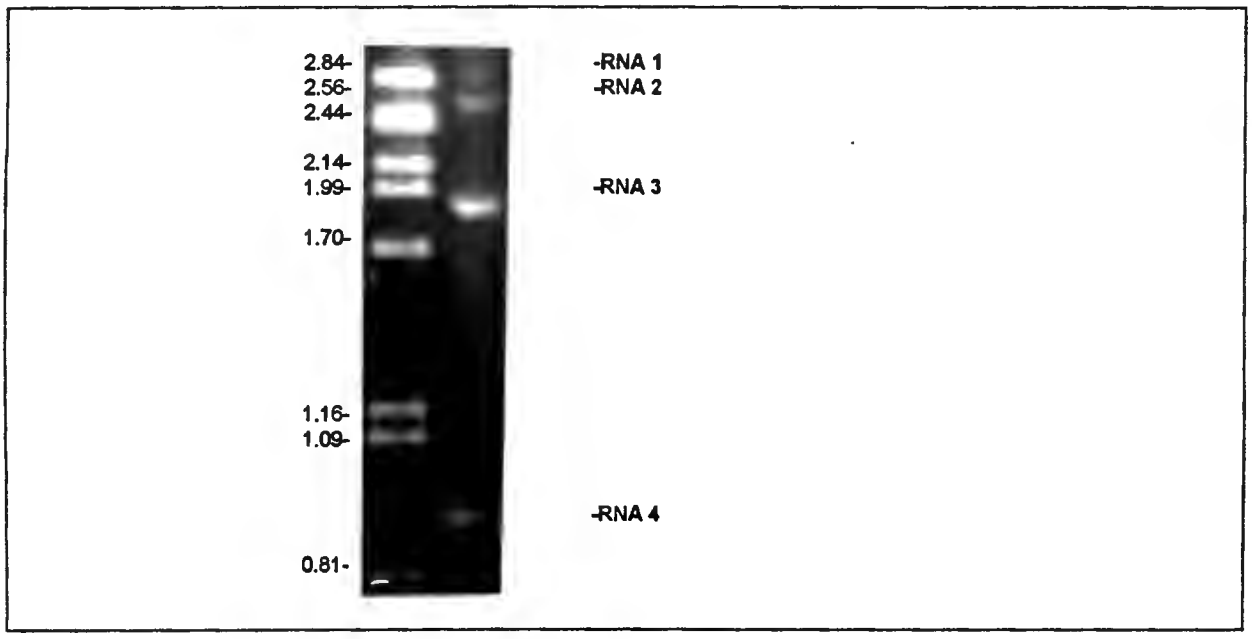

Figure 3.2 Double stranded RNA profile of CMV Y. Lane 1: λ-DNA cut with PstI. The sizes of the fragments are indicated in kilobases; Lane 2: CMV-Y dsRNA.

Four plants inoculated with denatured ds RNA exhibited slight mosaic symptoms three days post inoculation (p.i.) and severe mosaic patterns by day 8 p.i. Control plants showed no symptom development 22 day p.i (results not shown). This clearly demonstrated that fully denatured ds RNA of CMV-Y is infectious.

Reverse transcription and cDNA synthesis.

The extensive secondary structure found in the 5' and 3' UTRs (Figures 3.8 & 3.9) make reverse transcription (RT) and amplification very difficult. Both UTR fragments were obtained by extensively denaturing the ds RNA template in 10% DMSO and doing the RT reaction (Appendix H) at 70°C using a thermostable polymerase, Tth which also has reverse transcriptase activity in the presence of manganese (Boehringer-Mannheim). The cDNA was then amplified (Appendix I) by PCR with the necessary primers to give ds DNA fragments of each UTR (Fig. 3.4). The 5'-UTR was made using primers 5vec4 and 5vec5 for both the RT and PCR reactions (Appendix H & I). The 3'-UTR was reverse transcribed using primers 3vec4 and 3vec3a. Primer 3vec3, complementary to the 3' end of CMV Y, was unable to bind the template RNA during RT so another overlapping primer (Table 3.1), primer 3vec3a, was designed with less secondary structure than 3vec3. This primer yielded a slightly shortened 3'-UTR fragment after RT-PCR. The full-length 3' fragment (311 bp) was obtained by a second round of PCR using the 3vec3 primer (Fig. 3.4). PCR fragments were analysed on a 2% agarose gel (Appendix E) as shown in figure 3.3.

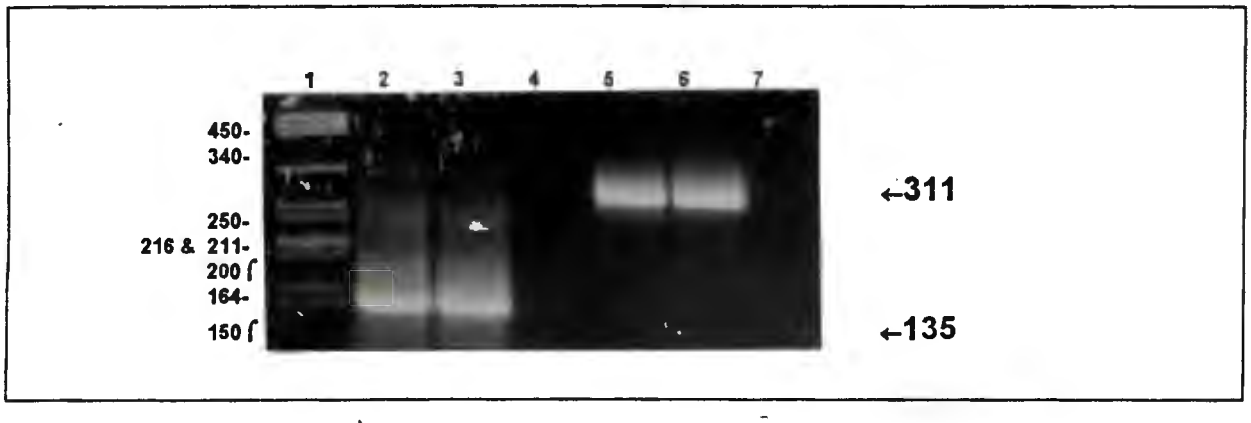

Figure 3.3. PCR amplification of the 5' and 3' untranslated regions. Lane 1: λ DNA cut with PstI. (the sizes of the bands are indicated in base pairs; lanes 2&3: 135 bp PCR fragment of the 5'UTR; lane 4: Control PCR RT-PCR with primers 5vec4 and 5vec5; lanes 5&6: 311 bp PCR fragment of the 3'UTR; lane 7: Control RT-PCR with primers 3vec3 and 3vec4.

Several attempts at obtaining a full-length cDNA clone of the entire CMV RNA 3 were unsuccessful. Neither increasing the DMSO concentration to 50% to fully denature the ds RNA, nor extending the extension times could solve the problem. MuMLV reverse transcriptase (Promega) and AMV reverse transcriptase (Promega) were used according to the manufacturers instructions, but no product was obtained.

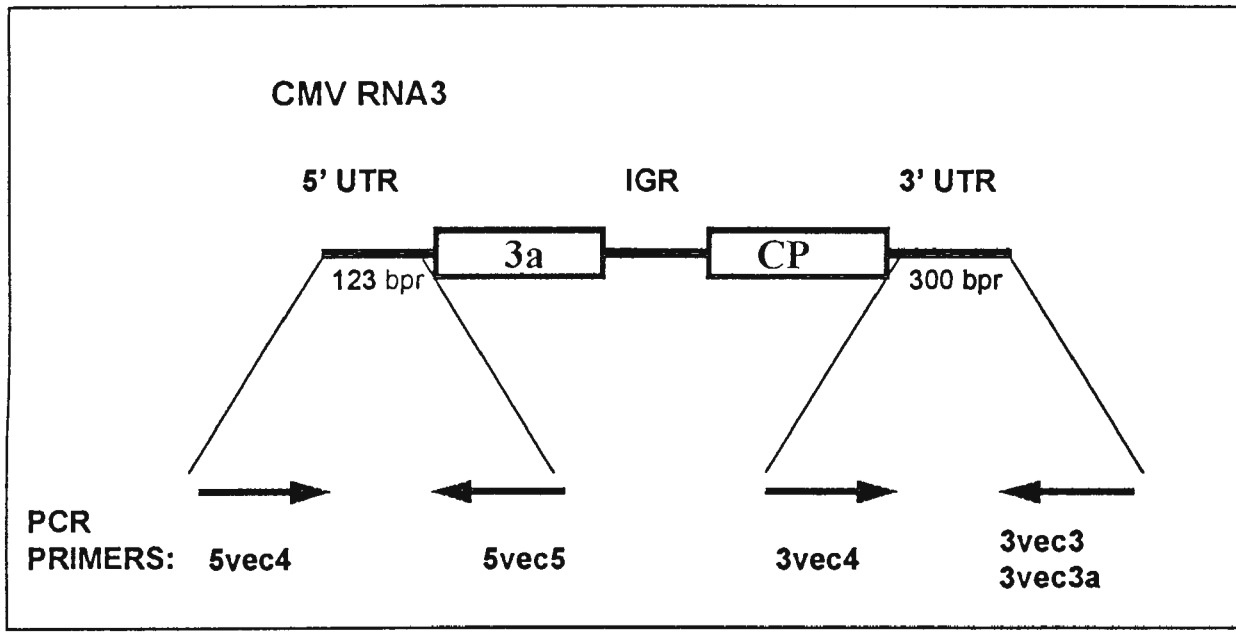

Figure 3.4. Diagram illustrating the PCR strategy for the amplification of the 5' & 3' UTRs as well as full-length RNA 3. A full description of each primer is given in table 3.1.

cDNA cloning and sequence analysis of:

The 5' UTR.

The 5' UTR (135 bp) was cloned into the XhoI site of pKS to give clone pCMV5' (Fig 3.5). The sequence varied from the published sequence of CMV Y at a few positions: a GC insertion at base 23 ; an A to G change at position 34; a G to A at position 51; a T to C at position 67; a deletion of an A at position 81; a change of an A to a T at position 83; a change of a T to an A at base 85; an A to C change at base

87 and the sequence at base 91 was altered from CGTT to ATGT. These variations in the sequence of the 5' UTR of CMV Y were found to occur in other subgroup I CMVs; the sequence of CMV Fny (Owen *et al.*, 1990) is shown in figure 3.6 for comparison.

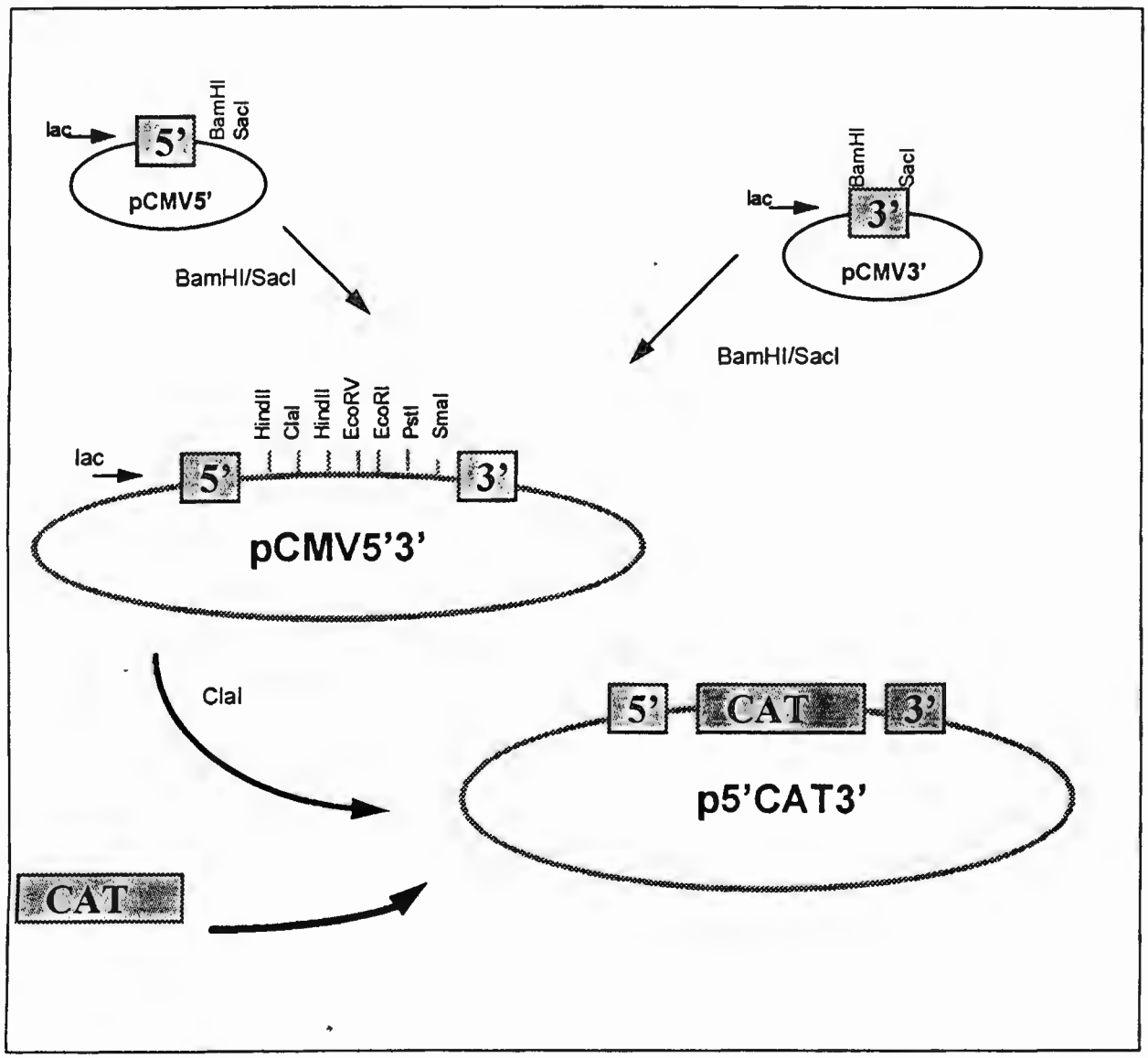

Figure 3.5. Diagram illustrating the cloning procedure to create p5'cat3'.

5': 5' untranslated region; 3': 3' untranslated region; CAT: chloramphenicol acetyl transferase gene TaqI fragment from pCm1; lac: arrow indicates the direction of the β-galactosidase gene. A more extensive description of the cloning strategy is given in the Materials and Methods section.

The 3'UTR.

The 3' UTR was cloned into SmaI cut pKS and the clone pCMV3' was subsequently sequenced. The 3' terminus was more conserved with only two changes being observed: an insertion of a T at position 2003 and a change of a T to a C at position 2065 of CMV Y. The deletion does not fall within the essential 3' tRNA-like region of the 3' UTR (Fig. 3.9) and the T to C change occurs in other strains of CMV; the sequence of CMV Fny is shown in figure 3.7 for comparison.

pCMV5'cat3'.

pCMV5'3' (Fig.3.5) was constructed by inserting the 3'-UTR downstream of the 5'-UTR in pCMV5'. The construct was checked by restriction enzyme analysis and chain reaction using the 5vec4 and 3vec3 primers (Fig 3.4).The cloning procedure created a mini-multiple cloning site (MCS) between the 5'- and 3'-UTRs which would facilitate the insertion of a foreign gene between them. The *E. coli* chloramphenicol acetyl transferase (CAT) gene was cloned as a TaqI fragment into the ClaI site in this mini-MCS. The CAT gene was obtained from pCM1 growing on CAT 100μg/ml. Sequencing of the clone p5'cat3' (Fig. 3.5) showed that the CAT gene was inserted in the correct orientation. Restriction enzyme analysis and PCR amplification using the 5vec4 and 3vec3 primers revealed no deletions in the initial p5'3' clone as result of this cloning step (results not shown).

```
CLUSTAL W(1.5) multiple sequence alignment


Cmv5pr        CTCGAGGTAATCTAACCACCTGTGTGTGTGCGTGTGTGTGTGTCGAGTCGTGTTGTCC
Cmvfny5   --------GTAATCTTACCA-CTGTGTGTGTGCGTGTGTGTGTGTCGAGTCGTGTTGTCC
Cmvy5     --------GTAATCTAACCACCTGTGTGTGT--GTGTGTGTGTATCGAGTCGTGTTGTCC
                 ******* **** **********  ********** ****************

Cmv5pr    ACACATTTGAGTCGTGCTGTCCGCACATAT-TTTATCTTTATGTGTACAGTGTGTTAGAT
Cmvfny5   GCACATTTGAGTCGTGCTGTCCGCACATAT-TTTATCTTT-TGGGTACAGTGTGTTAGAT
Cmvy5     GCACATTTGAGTCGTGTTGTCCGCACATATATATTTATTTCGTTGTACAGTGTGTTAGAT
           *************** ************* * * * ***    ****************

Cmv5pr    TTCCGAGCCTCGAGGG
Cmvfny5   TTCC------------
Cmvy5     TTCC------------
          ****
```

Figure 3.6. Sequence comparison of the 5'UTR with CMV Y (Nitta *et al.*, 1988) and another subgroup I CMV strain, CMV Fny (Owen *et al.*, 1990). Cmv5pr; 5'UTR of the strain of CMV Y used in this study; Cmvy5: 5'UTR of CMV Y (Nitta *et al.*, 1988): Cmvfny5: 5'UTR of CMV Fny (Owen *et al.*, 1990). Extra base pairs in Cmv5pr are from restriction enzyme sites incorporated into the primers used to amplify the 5'UTR.

```
CLUSTAL W(1.5) multiple sequence alignment

Cmv3pr     CCGGATCCTTCCCAGAATCCTCCCTCCGATCTCTGTGGCGGGAGCTGAGTTGGCAGTTCT
Cmvy3      --------TTCCCAGAATCCTCCCTCCGATCTCTGTGGCGGGAGCTGAGTTGGCAGTTCT
Cmvfny3    --------TTCCCAGAATCCTCCCTCCGATCTCTGTGGCGGGACGTGAGTTGGCAGTTCT
                   ********************************** ***************

Cmv3pr     GCTATAAACTGTCTGAAGTCACTAAACG-TTTTTACGGTGAACGGGTTGTCCATCCAGCT
Cmvy3      GCTATAAACTGTCTGAAGTCACTAAACG--TTTTACGGTGAACGGGTTGTCCATCCAGCT
Cmvfny3    GCTATAAACTGTCTGAAGTCACTAAACGTTTTTTACGGTGAACGGGTTGTCCATCCAGCT
           **************************** ******************************

Cmv3pr     TACGGCTAAAATGGTCAGTCGTGGAGAAATCCACGCCAGCAGATTTACAAATCTCTGAGG
Cmvy3      TACGGCTAAAATGGTCAGTCGTGGAGAAATCTACGCCAGCAGATTTACAAATCTCTGAGG
Cmvfny3    TACGGCTAAAATGGTCAGTCGTGGAGAAATCCACGCCAGCAGATTTACAAATCTCTGAGG
           ****************************** ****************************

Cmv3pr     CGCCTTTGAAACCATCTCCTAGGTTTCTTCGGAAGGACTTCGGTCCGTGTACCTCTAGCA
Cmvy3      CGCCTTTGAAACCATCTCCTAGGTTTCTTCGGAAGGACTTCGGTCCGTGTACCTCTAGCA
Cmvfny3    CGCCTTTGAAACCATCTCCTAGGTTTCTTCGGAAGGACTTCGGTCCGTGTACCTCTAGCA
           ************************************************************

Cmv3pr     CAACGTGCTAGTTTCAGGGTACGGGTGCCCCCCCACCTTCGTGGGGGCCTCCAAAAGGAG
Cmvy3      CAACGTGCTAGTTTCAGGGTACGGGTGCCCCCCCACCTTCGTGGGGGCCTCCAAAAGGAG
Cmvfny3    CAACGTGCTAGTTTCAGGGTACGGGTGCCCCCCCACTTTCGTGGGGGCCTCCAAAAGGAG
           ********************************** ***********************

Cmv3pr     ACCAGAGCTCGG
Cmvy3      ACCA--------
Cmvfny3    ACCA--------
           ****
```

Figure 3.7. Sequence comparison of the 3'UTR with CMV Y (Nitta *et al.*, 1988) and another subgroup I CMV strain, CMV Fny (Owen *et al.*, 1990). Cmv3pr; 3'UTR of the strain of CMV Y used in this study; Cmvy3: 3'UTR of CMV Y (Nitta *et al.*, 1988): Cmvfny3: 3'UTR of CMV Fny (Owen *et al.*, 1990). Extra base pairs in Cmv3pr are from restriction enzyme sites incorporated into the primers used to amplify the 3'UTR.

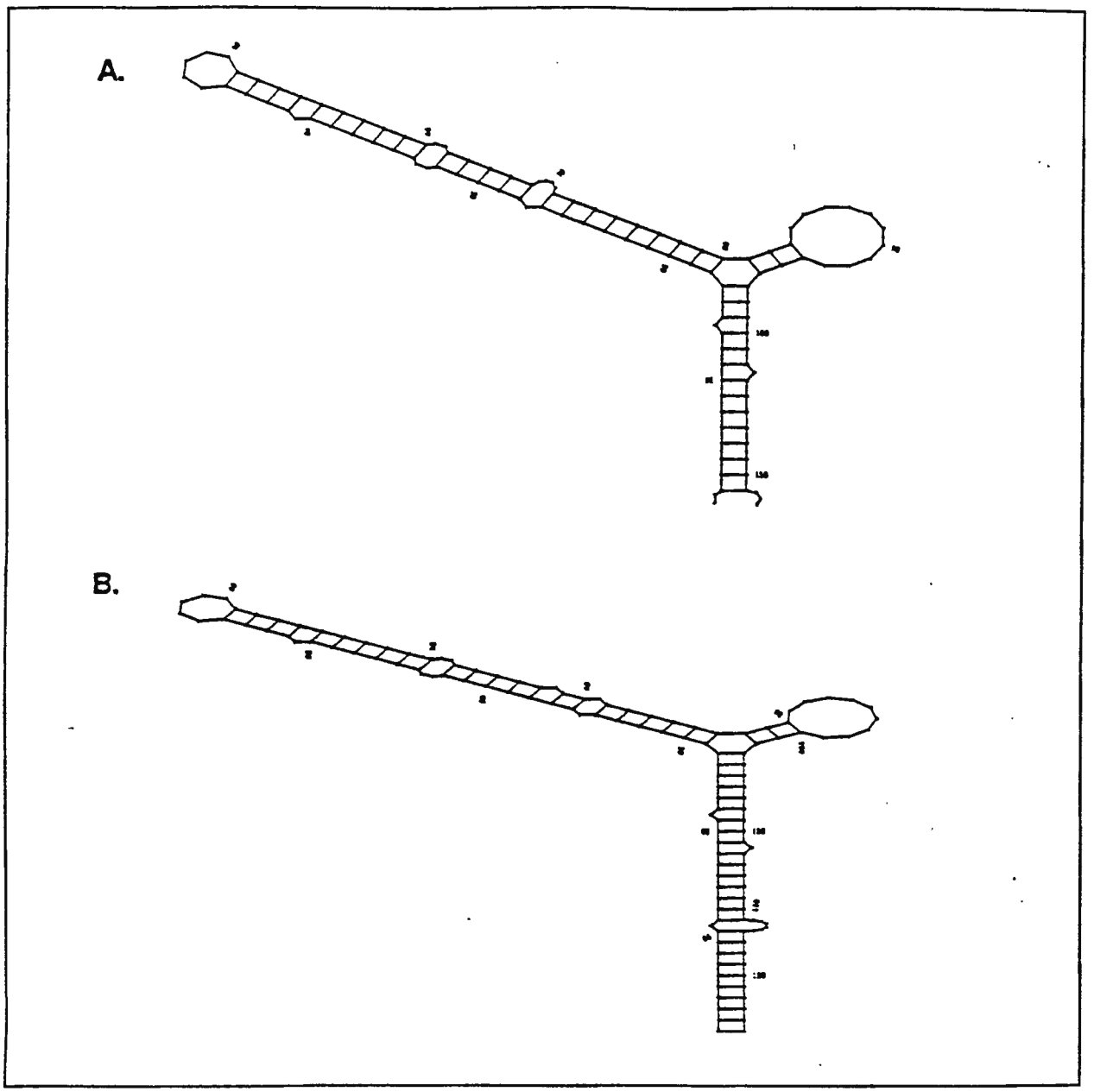

Figure 3.8. A comparison of the secondary structures of the 5'UTR of: A. the published CMV Y strain (Nitta et al., 1988) and B: The CMV Y strain variant used in this study. The program Squiggles was used to generate these representations from the Fold programme of the GCG Wisconsin sequence analysis package.

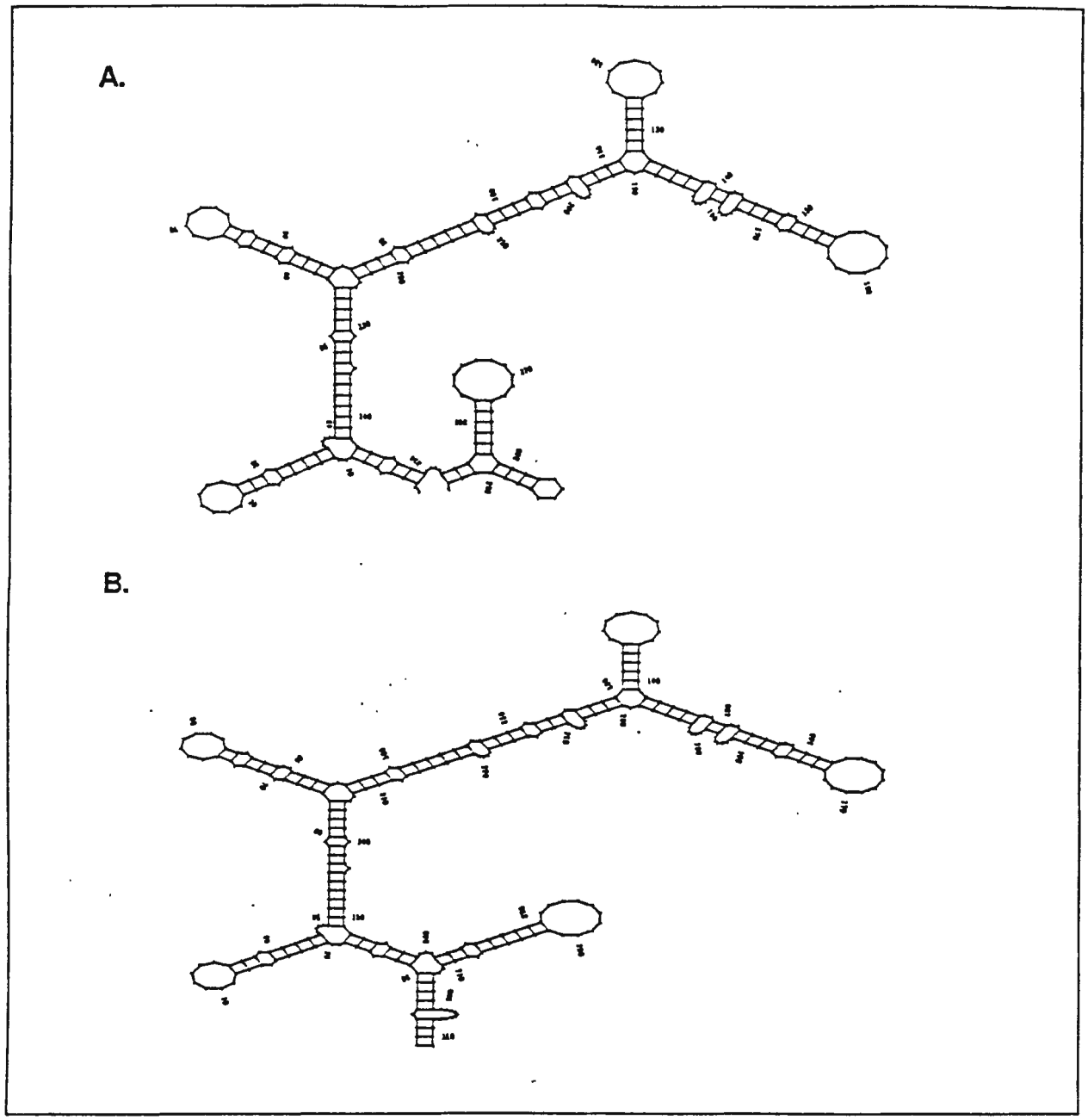

Figure 3.9. A comparison of the secondary structures of the 3'UTR of: A. the published CMV Y strain (Nitta et al., 1988) and B: The CMV Y strain variant used in this study. The program Squiggles was used to generate these representations from the Fold programme of the GCG Wisconsin sequence analysis package.

DISCUSSION.

The inability to make a full-length cDNA copy of the RNA 3 of CMV Y was problematic. The initial aim was to create a full-length cDNA from which a biologically active RNA 3 could be transcribed, the intention being to test the infectivity of the RNA 3 transcript by providing the replication functions in the form of fully denatured, gel purified ds RNA 1 & 2. The infectivity studies done on fully denatured double stranded RNA clearly showed that this form of CMV Y is fully infectious. The fully functional 5' and 3' UTRs would then have been excised from the infectious RNA 3 cDNA clone for construction of the expression vector. Several attempts using previously tried and tested protocols proved ineffective. A possible explanation could be that the ds RNA was not fully denatured, thereby stalling the reverse transcriptase in its first strand synthesis. Treating the ds RNA with methyl-mercuric hydroxide (Dawson *et al.*, 1986) to fully denature the RNA before RT PCR may obviate the problem. Another option is to utilise inner primers that will amplify an overlapping section of the RNA which also contains a unique restriction enzyme site. This site could be used to ligate the two shorter cDNA fragments of RNA 3 to create a full-length clone (Suzuki *et al.*, 1991).

The 5'- and 3'-UTRs that had been obtained directly from the ds RNA isolated from CMV Y were subsequently used to construct the viral vector without first being able to test their biological activities. The sequence analysis would suggest that these clones were fully functional, showing no serious deletions or changes in secondary structure (Figures 3.8 and 3.9).

Denatured ds RNA fom CMV-Y was shown to be fully infectious on tobacco plants. This result is novel as far as we know, as ds RNAs of ss RNA plant viruses have hitherto been thought to be non-infectious. The success of this experiment would suggest that RNA viruses could be stored and transported in the ds RNA form which is more stable than intact virions containing ss genomic RNA (Rybicki, 1995). The possibility of escape of the virus during transport is also reduced, as ds RNA needs to be fully denatured before it is infectious.

CHAPTER 4

CONSTRUCTION AND TESTING OF THE RNA & DNA EXPRESSION VECTORS CONTAINING THE CLONED CMV FRAGMENTS.

INTRODUCTION.

The creation of a viral expression vector is a logical progression from making infectious clones corresponding to the genomes of RNA viruses, which can be expressed in the plant either as infectious cDNA or as *in vitro*-transcribed RNA. Both techniques will be reviewed below.

IN PLANTA TRANSCRIBED RNA.

The first approach to expressing *in vitro* synthesised RNA in plants entails cloning the viral sequence downstream of the CaMV 35S promoter (Mori *et al.*, 1991; Dessens & Lomonossoff, 1993; Macfarlane *et al.*, 1992; Ding *et al.*, 1995). The DNA clone is inoculated onto the plant where it is targeted to the plant cell nucleus. The host transcription mechanism recognises the viral CaMV 35S or other plant promoter and makes mRNA from the DNA clone (Fig. 4.1). The mRNA is exported to the cytoplasm where it is translated to produce the relevant viral proteins (Van Bokhoven *et al.*, 1993) necessary for the normal viral life cycle in the host.

A full length cDNA copy of the viral RNA genome can be usually produced by reverse transcription utilising a primer complementary to the sequence at the 3'-OH end of the molecule. The ssDNA can then be made double stranded and amplified by PCR utilising a second primer complementary to the 5' end of the viral genome (Dessens & Lomonossof, 1993). These authors were also able to demonstrate that a polyadenylation signal is not required for the viral RNA to be

exported from the nucleus and translated in the cytoplasm: they inoculated cloned cDNAs of cowpea mosaic virus (CPMV) RNAs 1 & 2 onto cowpea seedlings, with the plasmid DNA linearised immediately after the viral 3' end, thereby preventing the synthesis of transcripts longer than genome-length. No CPMV symptoms were formed on cowpea plants inoculated with equal amounts of non-linearised plasmid DNA, confirming the importance of preventing the synthesis of over-sized transcripts (Dessens & Lomonosoff, 1993).

Yamaya *et al.* (1988) were able to express TMV RNA in transgenic plants generated by agroinfection with *Agrobacterium tumefaciens* expressing the viral sequence. They placed the viral cDNA sequences under the control of the CaMV 35S promoter in a specially designed plasmid vector. The CaMV 35S promoter was modified so that the 5' end of the *in vivo* transcripts would coincide with the 5' end of the viral RNA, without incorporating any extra non-viral nucleotides, thereby generating an RNA that resembled the true viral genomic RNA.

A contrary result has been demonstrated by Commandeur *et al.* (1991) for barley yellow dwarf virus (BYDV). cDNAs of BYDV RNAs 3 & 4 that were cloned between the CaMV 35 S promoter and a polyadenylation signal sequence, were found to be biologically active despite having extra non-viral nucleotides at their 5' and 3' ends. The extra nucleotides - up to 40 at the 5' end (only 15 extra were tolerated on RNA 3) and several hundred at the 3' end - were not included in the sequence of the *in vivo* produced transcripts.

There are cases of viral RNAs being synthesised from cDNAs lacking RNA polymerase sequences as was found with a cDNA copy of AlMV RNA3 (Dore & Pinck, 1988). It is thought that the promoter activity could be provided by "cryptic promoter-like" sites on the DNA or by integration into the host cell genome, downstream of an active promoter (Boyer & Haenni, 1994).

The expression of infectious viral RNAs through *in vivo* transcription of cDNA-containing vectors has several advantages. DNA is more stable than RNA and

therefore less likely to be degraded in the inoculum. The expense and technical difficulties of RNA transcription and capping are also circumvented by using cDNA clones of viruses (Ding *et al.*, 1995). Transcripts are also continuously provided from the cDNA in the nucleus of the inoculated cells (Mori *et al.*, 1991) to the cytoplasm for inclusion in the viral replication cycle.

IN VITRO TRANSCRIBED RNA.

Infectious viral RNA can be transcribed *in vitro* from a cDNA clone placed downstream of a T7 phage promoter using T7 RNA polymerase. These run-off RNA transcripts are then inoculated onto plants which are monitored for symptom development as illustrated in figure 4.1. This has been done successfully for CMV (Hayes & Buck, 1990b; Suzuki *et al.*, 1991 and Boccard and Baulcombe, 1992) as well as BMV (Ahlquist *et al.*, 1987; Mori *et al.*, 1991; Mori *et al.*, 1993), TMV (Dawson *et al.*, 1986; Donson *et al* 1991; Hamamoto *et al.*, 1993), alfalfa mosaic virus (Van der Vossen *et al.*, 1993.) and BBMV (Pogany *et al.*, 1994).

The first transcription vector designed by Ahlquist *et al.* (1984) allowed the insertion of a viral cDNA sequences downstream of a bacteriophage λ-promoter. *E. coli* RNA polymerase was used in the *in vitro* transcription reactions to make infectious RNA transcripts. There were a couple of disadvantages of the *E. coli* RNA polymerase based system: the polymerase produced a large proportion of premature termination products which resulted in lower than expected yields of transcript, and the commercially obtained *E. coli* RNA polymerases were also subject to great batch variability. The T7 RNA polymerase promoter sequence is preferred in the construction of cDNA clones for RNA transcription as T7 RNA polymerase gives much higher yields of transcript and its level of activity is maintained between different batches of the enzyme (Ahlquist *et al.*, 1987; Boyer & Haenni, 1994).

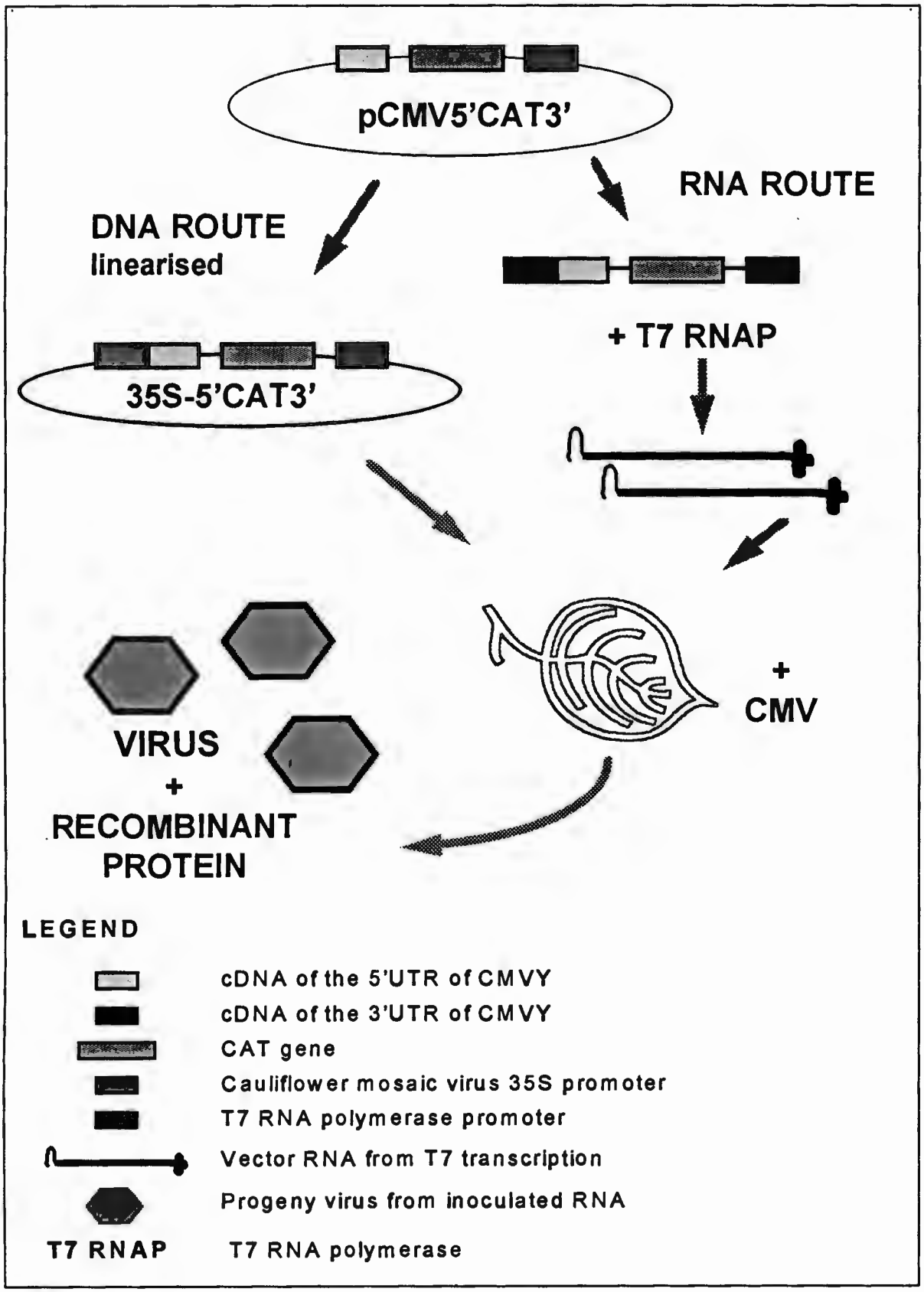

Figure 4.1 Diagram illustrating the inoculation strategy of the artificially produced RNA and cDNA copies of the viral RNAs onto plants to produce infectious virions.

An oligonucleotide containing the T7 promoter sequence directly linked to the 5'-end nucleotides of the viral sequence can be used to prime second strand synthesis of the viral cDNA. This procedure ensures the promoter is fused to the 5' end of the viral sequence without the incorporation of extra non-viral nucleotides. The aim of the cloning process is to obtain RNA transcripts resembling the wild type RNA as closely as possible, as previous studies have demonstrated that addition of 5' non-viral nucleotides or interference with the 3' end secondary structure interferes with viral infection efficiency (Miller *et al.*, 1986; Janda *et al.*, 1987).

The advantage of this approach is that one does not need to have cDNA copies of all the viral genome segments in order to study the segment of interest. The other segments needed for infection can be supplied from a total RNA extraction from a plant infected with the wild type virus. The wild type RNA would then support the replication of the transcribed viral RNA segment. The RNA is immediately able to be expressed and produce the protein of interest. The disadvantages are that the inoculum RNA is easily degraded and vast quantities are required for successful inoculations. Preparation of RNA transcripts is very labour intensive as all procedures must be done in an RNase free environment and the transcripts also have to be capped with 7-methyl guanosine to be biologically active in the plant cell.

This chapter describes the construction and subsequent testing of both DNA and RNA versions of the CMV derived construct, 5'cat3'. The DNA plasmid vector carried the 5'cat3' construct immediately downstream of the CaMV 35S promoter and was linearised at the 3' end of the viral sequence prior to inoculation onto CMV infected tobacco plants. RNA transcripts of the construct were made by utilising the T7 RNA polymerase promoter sequence fused to the 5' end of the 5'cat3' construct. T7 run-offs were inoculated onto CMV infected tobacco plants and monitored for CAT expression.

MATERIALS AND METHODS.

Construction of the DNA vector.

The 5'cat3' segment of p5'cat3' was amplified using the 3vec3 primer and a modified 5vec4 primer (Appendix I). The 5vec4c primer (5'-GTAATCTAACCACCTGTGTG-3'OH) lacks the XhoI site found in primer 5vec4: its first 5' nucleotide is the first base of the RNA 3 sequence. The resultant PCR product was blunt-end ligated into the StuI site of pUC18-35S (Dessens & Lomonossof, 1993) as per appendix J. Clonal orientation was determined by restriction enzyme analysis. A clone carrying the 5'cat3' construct in the correct orientation was sequenced using the 5vec5 primer (Appendix D).

Construction of the 5'cat3' RNA constructs.

A primer incorporating the T7 RNA polymerase sequence, primer 5vec6b (5'-ATTAATACGA CTCACTATAG TAATCTAACC ACC-3'OH) was used together with the 3vec3 primer (Table 3.1) to amplify the vector construct from p5'cat3' (Appendix I). Repeated attempts at cloning this product proved futile; the resultant clones all lacked the T7 promoter sequence. RNA transcripts were therefore made using the PCR product as template. Transcription was performed as per the Boehringer-Mannheim T7 RNA polymerase protocol on two microlitres of PCR product. 5µl aliquots of run-off product were incubated with 2 units of DNase or 10µg of RNase respectively, incubated at 37°C for 30 minutes and run on RNase free 1xTBE (Fig. 4.2) gels with untreated transcription products (Appendix E). The gels were blotted onto Hybond N^+ nylon membrane in 20x SSC as per the manufacturers instructions. Blots were probed in high SDS buffer (Appendix F) with digoxigenin labeled CAT probe (Appendix Gi) and detected using chemiluminescence as per the Boehringer-Mannheim protocol (Fig 4.2). A 2µl aliquot was taken from the correctly sized RNA band and used in a reverse transcription reaction (Appendix H) using primer 3vec3. Second strand synthesis was performed with the 5vec4 primer. The cDNA was amplified with the same primers and viewed on a 1xTBE gel (Fig. 4.3), (Appendix E). Transcripts for inoculation were synthesised with the addition of an m^7Gppp cap (Epicentre).

Inoculation of CMV Y RNA on plants.

Nicotiana tabacum cv. *xanthi* seedlings at the 3-4 leaf stage were inoculated with total RNA extraced from a *Nicotiana glutinosa* plant systemically infected with CMV Y. RNA was isolated essentially following the method given by Verwoerd *et al.* (1989). Briefly; 2 x 5g of leaf tissue was ground to a fine powder in liquid nitrogen and transferred to 2 50ml centrifuge tubes. 12ml hot extraction buffer [phenol: 0,1 M $LiCl_2$, 100mM Tris-HCl pH 8, 10 mM EDTA, 1% SDS (1:1)] was added to each tube and the sample vortexed. 6ml chloroform was added to each tube and subsequently vortexed. After 30 minutes shaking at room temperature, the samples were centrifuged at 13000 rpm for 10 minutes in a Sorval RC-5 Superspeed centrifuge. 15ml supernatant fluid was drawn off each sample and 7.5ml $LiCl_2$ was added per sample. The RNA was precipitated overnight at 4°C and pelleted the following morning by centrifuging at 10000 rpm for 30 minutes at 4°C. Pellets were resuspended in 5ml water and reprecipitated with 500µl 3M sodium acetate, pH 4.7 and 11ml 100% ethanol. The samples were gently mixed and left at -20°C for 4 hours. RNA was pelleted as before and resuspended in 1ml water. All steps were taken to ensure that reagents and equipment were RNase-free. RNA was inoculated at a concentration of 250µg/ml in a 2% bentonite solution made up in 1xSTE, pH 6,8. A volume of 25µl was inoculated per leaf of the plant. Seedlings were systemically infected with CMV Y 8 days post infecton (p.i.) and were ready for vector inoculations.

Vector Inoculations.

The pUC18-35S plasmid containing the 5'cat3' construct in the correct orientation was linearised after the viral 3' sequence by utilising the SacI site in the 3vec3 primer. This would generate an *in vivo* RNA transcript having 5 additional non-viral nucleotides at the 3'-end. Linearised DNA was inoculated at a concentration of 133µg/ml (in a 2% bentonite solution made up in 1xSTE pH 6.8) onto three celite dusted leaves of *Nicotiana tabacum* cv *xanthi* plants that were all ready systemically infected with CMV Y. A total of 10µg of DNA was inoculated onto each plant. Circularised plasmid was inoculated similarly at a concentration of 266µg/ml or 20µg per plant. Plants were at the 8 leaf stage at the time of inoculation and the top three leaves of each plant were inoculated with the constructs. m^7Gppp capped T7 transcripts were inoculated at a total concentration of

133µg/ml which meant that approximately 5µg of the correctly sized transcript was inoculated per plant.

CAT ELISA analysis of inoculated plants.

16 mm punches were taken from the inoculated leaves at 3 and 5 days post inoculation. Leaf discs were ground up in 400µl extraction buffer (0,25 Tris, pH 7.8, 5mM EDTA, 1% β-mercapto-ethanol) and centrifuged at 4°C for 15 minutes. The supernatant fluid was drawn off and the protein concentration was determined by a standard Bradford assay (Appendix K). A volume representing 50µg of total protein was used in the Boehringer-Mannheim CAT ELISA Detection Kit. A negative control plant leaf punch was spiked with 100 pg CAT enzyme to serve as a positive control, as no positive control CAT-expressing plants were available.

Northern blot analysis of inoculated plants.

400 mg of an inoculated leaf was ground up in liquid nitrogen and added to 1ml of Trizol (Life Sciences) reagent. Total RNA was extracted as per manufacturers instructions and run on an RNase free gel (Fig. 4.4 & 4.5) as per Appendix E. Gels were run in duplicate, blotted and probed with either a DIG-labelled CAT RNA probe (Appendix Gii) or dig-labeled CMV RNA 3' RNA probe (Appendix Giii), (Fig 4.4 & 4.5). Detection was according to the Boehringer-Mannheim chemiluminescence detection protocol.

RESULTS.

CONSTRUCTION OF THE DNA VECTOR.

The 5'cat3' construct could not be excised from p5'cat3' as the XhoI site situated on the 5'-end of the viral 5' sequence was repeated in the 5'-end of the 5vec5 primer initially used to clone the 5' UTR. It was therefore necessary to amplify the fragment using a new 5' primer, 5vec4c, and the old 3vec3, and blunt clone it into the specially constructed StuI site situated at the transcription initiation site of the

CaMV 35 S promoter found in pUC18-35S (Fig 4.1) (Dessens & Lomonossoff, 1993).

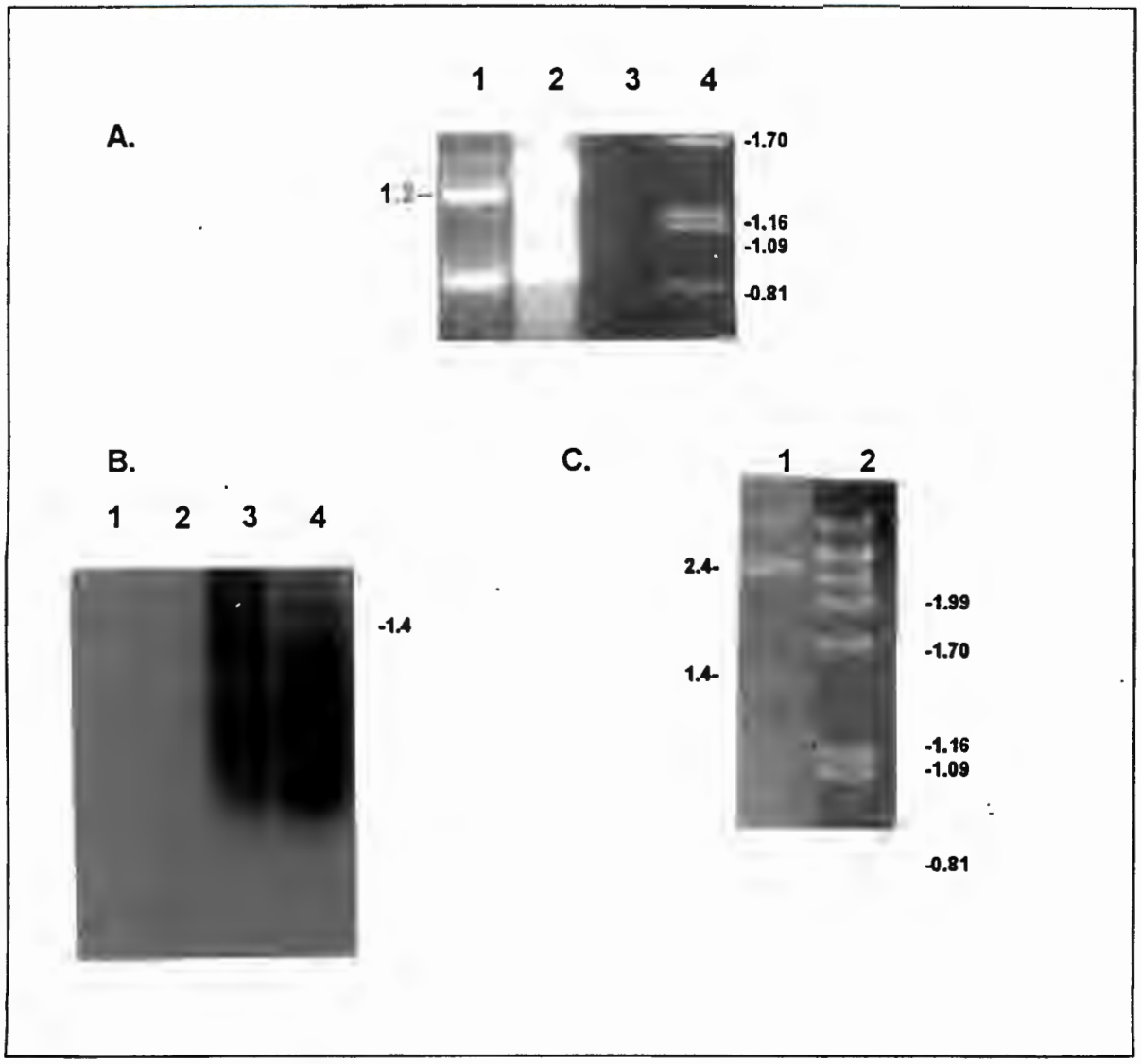

Figure 4.2. Analysis of the vector transcripts. (A.) Lane 1: 8 µl of a transcription reaction including m^7Gppp capping reagent; lane 2: 8µl of a standard T7 RNA polymerase reaction from PCR product of the 5'cat3 construct bearing the T7 promoter sequence at its 5'-end and lane 4: DNA cut with Pst I (the sizes of the fragments are indicated in kb in the illustration). The expected 1.2 RNA transcript is also indicated. (B.) Northern blot analysis of the transcripts using a DIG-labelled DNA fragment specific for the CAT gene. Lane 1:RNase treated transcripts; lane 3: Positive PCR control of the 1.2 kb 5'cat3' fragment and lane 4: DNase treated transcripts. The size of the Gibco BRL RNA molecular weight marker is indicated alongside the autoradiograph in kb. (C.) Gel showing the relative sizes of λ-DNA cut with Pst I and RNA molecular weight marker from Gibco BRL on a 1% agarose gel.

Primer 5v4c was made to eliminate the Xho I site in primer 5vec4 and ensure that the viral sequence started at the start of the inserted sequence. Sequencing over the junction site confirmed that no extra non-viral nucleotides were inserted between the 5' end of the construct cDNA and the transcription initiation site of the promoter.

CONSTRUCTION OF THE RNA TRANSCRIPTS.

The 5'cat3' construct was amplified from p5'cat3' using a primer containing the T7 RNA polymerase promoter sequence and the 3' primer, 3vec3. Several unsuccessful attempts were made at cloning the product into blunted vectors. On each of numerous occasions that a clone of the correct size and orientation was obtained, sequencing of the 5' end revealed that the T7 promoter sequence was missing. It was however possible to obtain RNA transcripts by using PCR product as template in a T7 RNA polymerase transcription reaction. Gel analysis of the products revealed two bands, one at the expected size of 1.2 kilobase pairs and another smaller product (Fig 4.2). DNase and RNase treatment of the resultant products indicated that the products were definitely RNA (Fig 4.2). Both of the run-offs contained the CAT gene as demonstrated by hybridisation to the CAT probe (Fig. 4.2). Reverse transcription of the products demonstrated that the larger product was indeed the correct sized RNA (Fig. 4.3). The smaller RNA could not be reverse transcribed by the primer complementary to the 3'-end of the construct. Once the integrity of the transcripts had been determined, transcripts were done in reaction mixes containing m^7Gppp and lower concentrations of GTP to make capped RNA transcripts of the vector (Fig. 4.2).

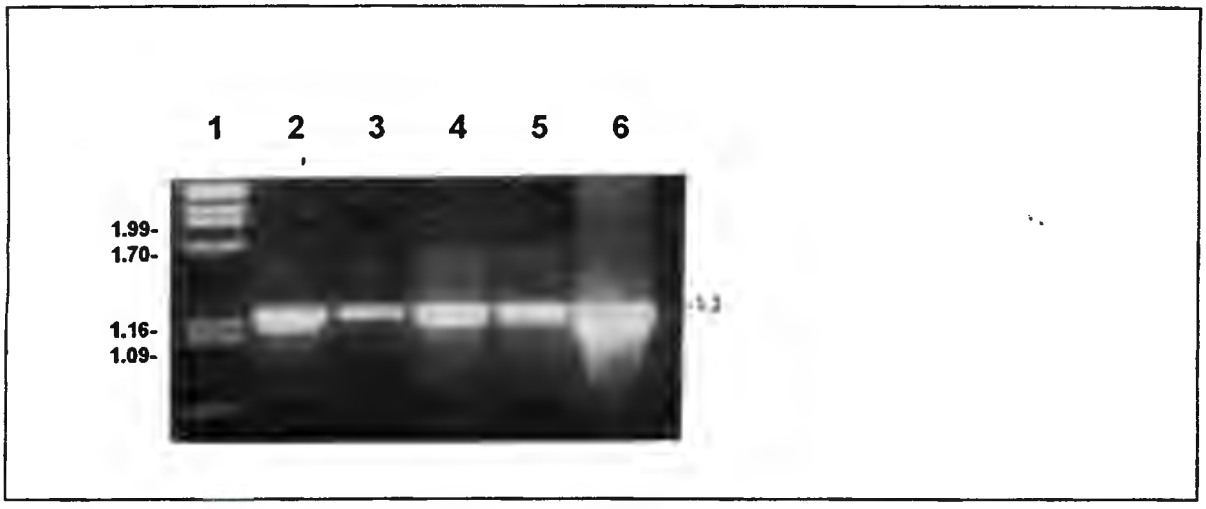

Figure 4.3. Analysis of the reverse transcription and amplification of the larger T7 transcript generated from the PCR templates. Lane 1: DNA cut with Pst I (the sizes of the fragments are indicated in kilobasepairs); lanes 2 - 4: product of the reverse transcription/amplification of the transcript; lane 5; positive control amplification of p5'cat3' utilising the same primers as in the reverse transcription/amplification steps and lane 6: product from the reverse transcription and amplification of the smaller transcription product.

CAT ELISA DETECTION.

No CAT protein was detected in any of the inoculated plants at 3 or 5 days post inoculation. There was no reduction in the activity of the enzyme used to spike a healthy leaf sample at the start of CAT isolation from the leaves. The ELISA could be used to quantify CAT when present in extracts at 100 pg/ml or 20 pg in 50 µg of total protein (400 pg/mg protein).

NORTHERN BLOT ANALYSIS OF THE INOCULATED PLANTS.

No CAT RNA could be detected in total RNA extracts from leaves inoculated with either the DNA or RNA vector (Fig.4.4). $7{,}34 \times 10^{-16}$ moles of the target sequence could be detected using the CAT RNA probe. Northern blot analysis of the inoculated leaves demonstrated replicating CMV as detected by a probe complementary to the 3' UTR of CMV Y RNA 3 (Fig. 4.5). This probe was able to detect down to 73,4 femtomoles of the target sequence.

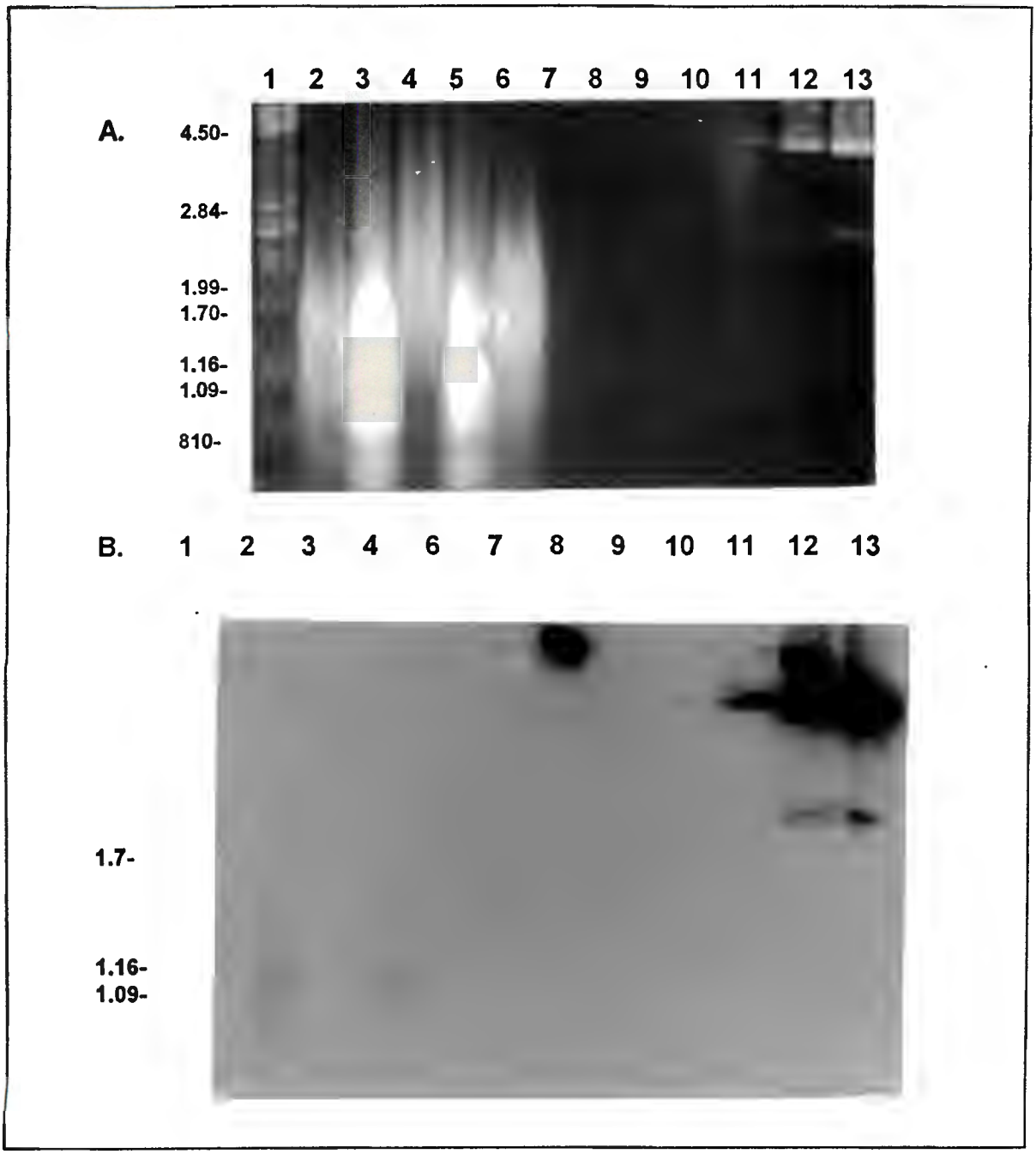

Figure 4.4. Analysis of the production of chloramphenicol acetyl transferase (CAT) in inoculated plants. (A.) Total RNA isolated from inoculated and control plants; 10µg RNA was loaded per lane. Lanes :DNA molecular weight markers, λ-DNA cut with PstI (the sizes of the fragments are indicated in kilobasepairs at the side of the photograph; see fig. 4.3 for a comparison of DNA and RNA molecular weight markers); lanes 2 & 3: RNA from CMV infected plants inoculated with the RNA vector transcripts; lane 4: RNA from CMV infected plants inoculated with the linearised DNA vector; lane 5: RNA from a CMV infected plant and lane 6: RNA from a healthy *Nicotiana tabacum* cv. *Xanthi* plant. Lanes 8-13 are SacI linearised p5'cat3', size 4.1 kb, run as positive controls in increasing concentrations of 10 pg, 100pg, 1ng, 10ng, 100ng and 200ng. The lower band is uncut plasmid (B.) Northern blot analysis of the same gel using the DIG-labeled RNA probe for negative sense CAT RNA. The sizes of the fragments of λ-DNA cut with Pst I are indicated alongside the photograph in kb.

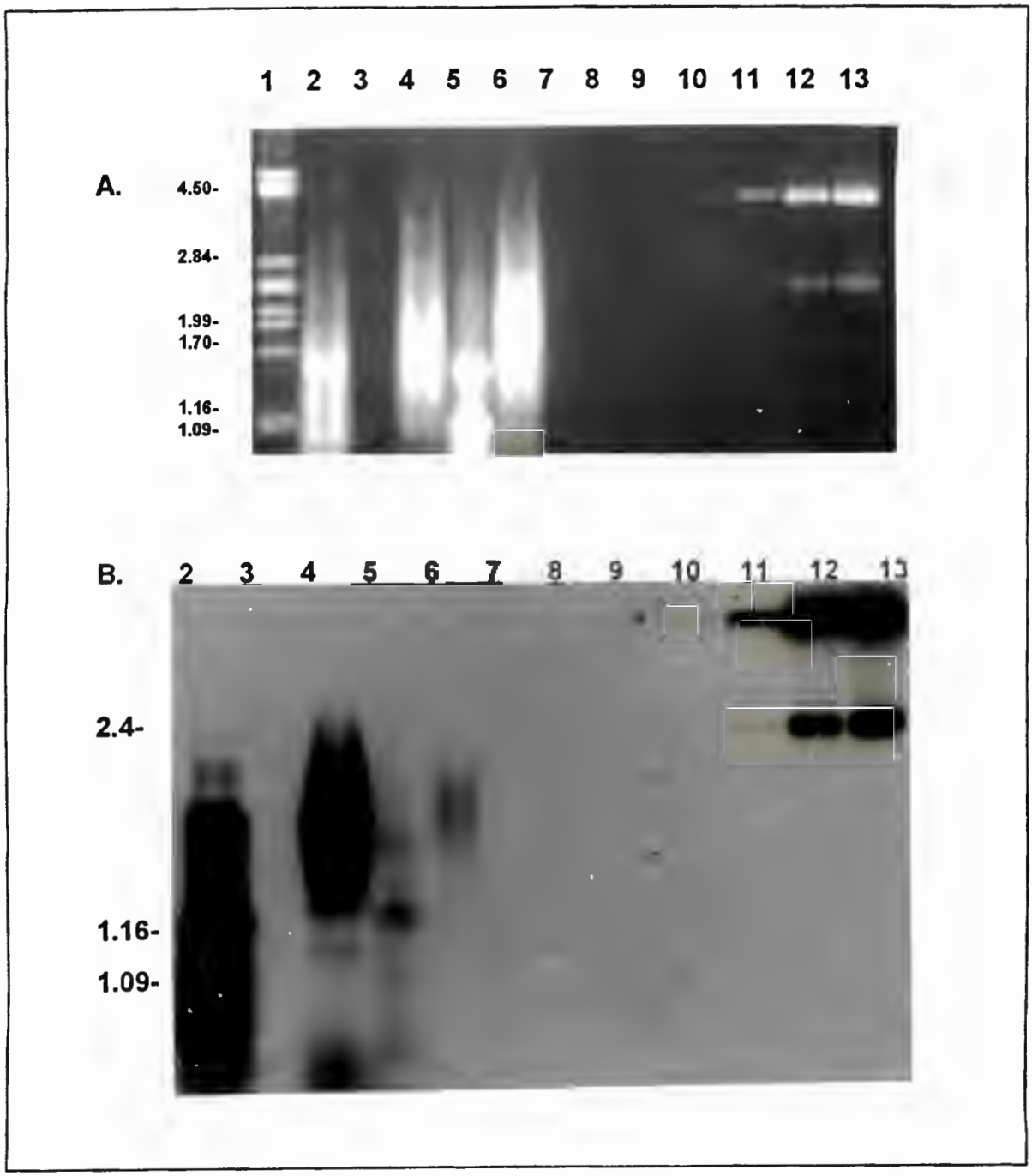

Figure 4.5. Detection of CMV RNA in the inoculated plants. (A.) Total RNA isolated from inoculated and control plants; 10µg RNA was loaded per lane but only 2µg was loaded in the CMV infected control plant lane. Lane 1 :DNA molecular weight markers, λ-DNA cut with PstI (the sizes of the fragments are indicated in kilobasepairs at the side of the photograph; see fig. 4.3 for a comparison of DNA and RNA molecular weight markers); lane 2: RNA from a CMV infected plant inoculated with the RNA vector transcript; lane 4: RNA from a CMV infected plant inoculated with the linearised DNA vector; lane 5: RNA from a CMV infected plant and lane 6: RNA from a healthy *Nicotiana tabacum* cv. *Xanthi* plant. Lanes 8-13 are SacI linearised p5'cat3', size 4.1 kb, run as positive controls in increasing concentrations of 10 pg, 100pg, 1ng, 10ng, 100ng and 200ng. The lower band is uncut plasmid (B.) Northern blot analysis of the same gel using the DIG-labeled RNA probe for negative sense CAT RNA.

DISCUSSION

The expression of a foreign protein in a plant is a difficult task. A suitable assay system is essential to detect both the expressed protein and its RNA. Reliable positive controls are also required. The CAT ELISA results indicated that the CAT could have been produced at a level lower than 100 pg/ml of plant extract. This is 20 to 200 times lower than the 70 to 739 ng CAT/mg total protein obtained in transgenic plants carrying the *cat*-gene under the control of a CaMV 35S promoter (Gendloff *et al.*, 1990). This level of expression would be of no use in the construction of a high level plant expression vector as too much plant material would have to be generated for large-scale protein extractions. A suitable positive control could have confirmed that the assay system was functional. CAT has been expressed in *N. tabacum* previously (Dawson *et al.*, 1989; Gendloff *et al.*, 1990) and is known not to be toxic to the plant.

Northern blots indicated that no detectable *cat* RNA was being transcribed (Fig. 4.4) despite the high sensitivity of the blot - as little as 734 *f*mol of the target sequence could be detected. This would seem to indicate that the problem of *cat* expression was at the level of replication of the vector and not due to a translational problem.

As little as 7 *f*mol of CMV RNA could be detected on the total RNA extracts from CMV infected plants inoculated with the constructs as well as the CMV control plants (Fig. 4. 5). A fully functional, replicating vector would have shown up as a 1.2 kb RNA band on the northern blot using the CMV 3'-end as a probe. No band was visualised at this size, even at shorter exposure times (not shown), suggesting that the vector was not replicating in the CMV infected plants.

There are several possible reasons as to why the vector construct was not recognised by the viral replication machinery and amplified in the plant host. The first is that the 5' and/or 3' UTR's are not biologically active. This could be overcome by first isolating a full-length biologically active cDNA clone of CMV

RNA 3 and then excising the relevant control regions for further manipulations. Sequence analysis of the 5'- and 3'-untranslated regions did not however suggest that there were any serious deletions or insertions in the sequences other than the normal variations found amongst subgroup I CMV's, as shown in figures 3.5 and 3.6. There did not appear to be any major disruptions of the secondary structures of these control regions either, as demonstrated in figures 3.8 and 3.9.

The other possible reason as to why the vector was not replicated could lie in the fact that the 5' and 3' regions are taken from CMV RNA 3, which in wild type viruses also carries a third UTR in the intergenic region between the coat protein gene and the 3a movement protein gene (Fig 3.3). This region has been found to play an important role in the replication of BMV RNA 3. (Marsh *et al.*, 1991). Deletion of sections of the internal control region of BMV decreased RNA 3 replication in protoplasts to less than 15% of the wild type levels (Pogue *et al.*, 1992). Boccard and Baulcombe (1993) were able to demonstrate that mutants in the CMV RNA 3 of the Kin strain that resulted in truncated coat protein being produced, were still able to replicate to levels of 20% of the wild type virus. Mutants containing major deletions in the coat protein **and** in the intergenic region were unable to replicate at all. They were able to show that a 250-nucleotide region in the intergenic region was required for CMV RNA 3 accumulation (Boccard and Baulcombe, 1993). Inclusion of the IGR of CMV Y RNA 3 into the vector construct as well as the insertion of a second reporter gene to approximate the size of CMV Y RNA3 may result in an effective, high level expression vector for the production of proteins in plants.

CONCLUSION

The use of a viral expression vector for foreign protein production in plants is a relatively new and complicated concept. Although this attempt at constructing a viral expression vector from CMV RNA 3 proved unsuccessful it was not entirely fruitless.

It would appear that without the IGR, the CMV RNA3 cannot be replicated, thus confirming the results of Boccard and Baulcombe (1993) that this region is essential for replication of the virus. This study has also shown that PCR can be used to amplify short control regions effectively. No unacceptable base pair changes occured in the 5' and 3' UTR's cloned in this project (Fig. 3.6 & 3.7) and the secondary structures were maintained (Fig. 3.8 & 3.9). Conformation of the functionality of these regions would be to place the IGR of CMV-Y into the mini multiple cloning site of construct p5'cat3', downstream of the *cat*-gene (Fig. 3.5). A second reporter gene could be cloned between the inserted IGR and the 3'UTR. RNA transcripts could be generated as before and inoculated onto CMV-Y infected plants to test for expression of the reporter genes. The results of Boccard and Baulcombe (1993) would suggest that such a construct should be recognised by the viral replication machinery and replicated together with the viral genomic RNAs.

In trying to repeat this endeavour, it would be advisable to obtain a full-length biologically active cDNA copy of the entire RNA3 molecule of the CMV Y genome. The 3a and coat protein genes could be excised and replaced by two reporter genes of similar coding capacities to try and maintain the original size of the wild type RNA3.

The discovery that completely denatured ds RNA of CMV Y is fully infectious is, as far as can be determined, a novel result. It has previously been thought that ds RNA of ss RNA viruses is non-infectious. This study demonstrated that ds RNA as such is non-infectious, but once it is denatured by heat and DMSO it is

infectious enough to cause symptoms in all the plants tested. A preparation of ds RNA thereofor represents a viable alternative for the long-term storage of ss RNA plant viruses, especially labile viruses such as CMV and AlMV.

Tobacco plants containing the HPV-16 capsid protein gene have been made. Whether the L1 gene exists in these 17 plants as an integrated insert or as an *A. tumefaciens* contaminant remains to be determined. Further studies need to be conducted on these plants at the DNA, RNA and protein levels to characterise the possible expression of the L1 protein.

In conclusion, the field of viral expression vectors is still growing rapidly, with new discoveries being made regularly. This project has illustrated the complexities of constructing a plant viral expression vector system, the functioning of which we know little. All knowledge obtained about the system, be it positive or negative, contributes to the eventual success of the goal:

THE PRODUCTION OF PROTEINS IN PLANTS USING A VIRAL EXPRESSION VECTOR.

APPENDIX

A: Bacterial media, antibiotics and bacterial strains:

All bacterial cultures were grown in Luria-Bertani medium containing the relevant antibiotics as indicated in Sambrook *et al.* (1989).

Unless otherwise stated, all DNA was transformed into and isolated from *E.Coli* Dh5α .

Mini-prep DNA was isolated from 1ml cultures grown up overnight in 2ml Eppendorf tubes at 37°C, shaking vigourously. Maxi-prep DNA was obtained from 100ml cultures grown up in 250 ml flasks in a similar fashion.

B: DNA isolation procedures:

Maxi-prep DNA was isolated using the Nucleobond AX PC-Kit 100 (Machery-Nagel). Minipreps were processed as in Sambrook *et al.* (1989) but utilising solutions 1, 2 and 3 as made in the Nucleobond kit. Miniprep DNA was resuspended in 40 µl TE buffer and 5 µl was routinely in a restriction enzyme digest.

C: Restriction enzyme analysis of DNA:

DNA was routinely digested by the required enzyme under the conditions stated by the supplier (Boehringer-Mannheim). Two units of enzyme was used per microgram of DNA and digested for 1,5 hours at the recommended temperature.

D: Sequencing of cloned DNA templates:

DNA was sequenced according to the protocol of the TaqTrack sequencing system of Promega. The forward sequencing primer was the 24mer primer #1212 with sequence 5'-CGCCAGGGTTTCCCAGTCACGAC-3' while the reverse primer was the 24mer #1233 with sequence 5'-GAGCGGATAATTTCACACAGG-3'.

The junction site of the DNA vector was sequenced using the 5vev5 primer used for PCR at a concentration of 20ng/µl.

Sequence analysis was performed on the University of Wisconsin GCG package. Sequence comparisons were done on the ClustalW analysis program.

E: Gel electrophoresis of RNA and DNA:

The initial RT-PCR products of the 5'- and 3'-UTRs were run on 3% agarose gels in 1xTBE. All other agarose gel electrophoresis was done in 1% agarose gels in 1xTBE as indicated in Sambrook *et al.* (1989). Gels contained ethidium bromide at a concentration of 0,5 µg/ml.

RNA gels were run similarly except that the electrophoresis aparatus was soaked in 2% H_2O_2 overnight and rinsed thoroughly in DEPC-treated water and that all the reagents were prepared RNase free.

DNA was sized against λ-DNA digested with PstI to give 29 fragments ranging from 15 bp to 11497 bp: 15, 72, 87, 94, 150, 164, 200, 211, 216, 247, 264, 339, 448, 468, 514, 805, 1093, 1159, 1700, 1986, 2140, 2443, 2459, 2560, 2838, 4507, 4749, 5077 and 11497 bp.

RNA was sized against the RNA molecular weight marker I of Boehringer-Mannheim.

F: Hybridisation and detection of DNA/RNA blots:

DNA gels were first denatured and then neutralised indicated in the Hybond N^+ manual. Both DNA and RNA gels were blotted in 20xSSC; overnight for DNA gels and 6 hours for RNA gels. The nucleic acids were fixed to the membranes by exposing them, through clingwrap, to UV light for 3 minutes, nucleic acid side down. The membranes were then rinsed in 2xSSC before proceding to the hybridisation step.

DNA blots were prehybridised in the standard DNA prehybridisation buffer as given in the DIG System User's guide of Boehringer-Mannheim. Denatured DIG-labelled DNA probe was added to the prehybridisation buffer at a concentration of 25ng/ml. Membranes were incubated overnight at 68°C and washed the next day as per the DIG protocol. The hybridised probe was detected using the CSPD chemiluminescence detection system of Boehringer-Mannheim.

RNA blots were treated similarly except that the blots were probed with DIG-labelled RNA probes in high percentage SDS prehybridisation buffer containing 50% formamide (DIG systems user's guide). Probe was denatured at 60°C for 10 minutes and added to the prehybridisation buffer at a concentration of 30ng/ml. The lower probe concentration appeared to reduce the background signal. Blots were incubated at 68°C overnight and treated in the same manner as the DNA blots the following day.

G: Probe preparation:

i) CAT DNA probe

A DIG-labelled DNA fragment of chloramphenicol acetyl transferase (CAT) gene was obtained by PCR labelling using the PCR DIG Probe Synthesis Kit

(Boehringer-Mannheim). The CAT sequence specific primers were 18mers designed by K. E. Palmer and gave a 400 base pair product. 100ng of plasmid pCm1 (D. Rawlings) was used in a PCR DIG-labelling reaction using the following cycles: 3 minutes at 92°C, 30 seconds at 50°C and 1 minute at 72°C.

ii) CAT RNA probe

DIG-labelled RNA probes were obtained usng the DIG RNA labelling Kit (Boehringer-Mannheim) except that T3 RNA polymerase (Boehringer-Mannheim) and its 10x buffer were used in the reaction mix.

The positive sense CAT RNA probe was synthesised from XhoI linearised pCAT using T3 RNA polymerase. pCAT was synthesised by inserting the TaqI fragment of the CAT gene into the ClaI site of pBluescript SK. This probe would bind negative sense, replicating CAT RNAs.

iii) CMV RNA 3 3'UTR probe

The negative sense DIG-labelled probe for the 3' UTR was made as a T3 RNA polymerase run-off from PstI linearised pCMV3'. This probe would bind genomic or messenger sense viral RNA.

H: Reverse transcription reaction:

i) RNA denaturation.

2µl RNA template (about 1.5 µg of ds RNA) was mixed with 2µl H_2O, 5µl DMSO and 1µl RNase inhibitor (Boehringer-Mannheim), heated at 65°C for 10 minutes and placed on ice.

ii) Reverse transcription.

The 10µl denatured RNA solution was added to 40µl of reverse transcription reaction mix containing the following: 10 mM Tris-HCL at pH 8.9, 90 mM KCl, 200µM of each dNTP, 0,9 mM $MnCl_2$, 0,5µM of each primer and 4U Tth DNA polymerase (Boehringer-Mannheim). The volume was made up to 50µl with H_2O.

The reaction was performed at 70°C for 30 minutes after which EGTA was added to a concentration of 0,75 mM.

Primers 3vec3b and 3vec4 (table 3.1) were used to reverse transcribe the 3' UTR of CMV Noel green pepper while primers 5vec4 and 5vec5 were used for the 5' UTR.

2µl of the T7 RNA transcription product was pierced from both the upper and lower bands in an agarose gel and used for reverse transcription. Only the 3vec3b primer was used to prime the reaction.

I: Polymerase chain reaction (PCR):

5µl of the reverse transcription reaction was used in a 100µl PCR reaction. The conditions were: 0,5 µM of each primer, 1.5mM $MgCl_2$, 200µM of each dNTP, 1 unit Taq DNA polymerase (Advanced Biotechnologies) using the 10x buffer supplied (200mM $(NH_4)_2SO_4$, 750 mM Tris-HCl pH 9.0 & 0,1% (w/v) Tween). All the primers used in this project were used under the following cycling conditions: 94°C for 2 minutes for initial denaturation of the template; 25 cycles of 93°C for 1 minute, 48°C for 30 seconds and 72°C for 90 seconds followed by a final round of 72°C for 3 minutes.

J: Blunt-end cloning of PCR products :

i) Vector preparation.

5µg of vector DNA was typically digested with the required blunt end cutter enzyme for 1,5 hours, after which it was run on an agarose gel to check for complete digestion. Cut DNA was then extracted with phenol:chloroform to remove the enzyme. The DNA was precipitated with 0,5 vol. 5.25M ammonium acetate and 2 vol. 100% ethanol, mixed gently and left at -70°C for 30 minutes. DNA was recoverd by centrifugation in a benchtop centrifuge at 4°C for 15 minutes and resuspended in H_2O to a concentration of 100ng/µl. The digested

DNA was incubated with 24 units calf intestinal phosphatase (CIP) for 30 minutes at 37°C to remove the 5' phosphate groups at the cut ends of the vector. The reaction was spiked with another 24 units of CIP enzyme and incubated a further 15 minutes at 45°C. CIP was inactivated by heating at 70°C for 10 minutes. An equal volume of TE buffer was added to the mixture before being extracted once with phenol:chloroform. The DNA was precipitated as before and resuspended inTE buffer to a concentration of 100ng/µl.

ii) Preparation of the insert.

An equal volume of chloroform was added to the PCR reaction, vortexed and centrifuged 3 minutes. The aqueous phase was transfered to a new 1.5 ml tube and boiled for 10 minutes to destroy any residual taq DNA polymerase. The reaction mix was allowed to slowly cool down to room temperature. The PCR products were treated with polynucleotide kinase (PNK) to add phosphate groups to their 5'-ends. The reaction conditions were: 50mM Tris-HCl (pH 7,6), 10 mM $MgCl_2$, 10 mM β-mercaptoethanol, 2mM ATP and 2 Units PNK (Amersham). The reaction was incubated at 37°C for 30 minutes, spiked with an equal volume of ATP and incubated for a further 30 minutes at 37°C. The PNK was inactivated at 70°C for 10 minutes. The magnesium concentration was adjusted to 5mM, dNTPs were added to 100µM and 2 units Klenow was added to the reaction. Incubation at room temperature (22°C) resulted in the PCR products being completely blunted. The insert was treated once with phenol:chloroform, the DNA being precipitated from the aqueous phase with 0,5 vol. ammonium acetate and 2 vol. 100% ethanol. After precipitation at room temperature for an hour, the DNA was pelleted at 4°C for 15 minutes. The pellet was washed with 70% ethanol and resuspended to give a concentration of approximaetely 100ng/µl.

iii) Ligations.

Ligations were routinely performed in 20µl with 2 units T4 DNA ligase (Boehringer-Mannheim) at 15°C overnight. The vector:insert ratio for small PCR products of less that 500 base pairs was 1:6 with the DNA concentration not exceding 800ng. The ratio for larger inserts was 1:3 under similar conditions.

iv) Transformations.

E.Coli cells were made competent for transformation using the rubidium chloride protocol of Chung & Miller (1988). 5µl of the ligation mix was routinely transformed into competant *E.Coli* Dh5α. Cells were plated onto agar plates containing the necessary antibiotics and incubated overnight at 37°C. Transformants were screened for inserts by digesting mini-prep DNA as indicated above.

K: Bradford assay to determine protein concentration:

i) Preparation of Bradford reagent: 100mg of Coomassie Blue G-250 was dissolved in 50 ml 95% ethanol (v/v). 100 ml of 85% phosphoric acid (w/v) was added and the volume adjusted to one litre with distilled water.

ii) Protein determination: 5µl of total protein extract was added to 95 µl water and then added to 900 µl Bradford reagent. The solution was gently mixed to avoid bubble formation and the absorbance of the sample was read at 595 nm. The measurements were done between 5 and 45 minutes after the addition of the Bradford reagent to the sample. The protein concentration was calculated by comparing the absorbance of the sample with a calibration curve of BSA ranging from 10-100 µg/ml.

REFERENCES

Agracetus, Inc. (1995). Plant bioreactor systems program overview. May, 1-5.

Ahlquist, P., French, R., Janda, M. and Loesch-Frisch, L. S. (1984). Multicomponent RNA plant virus infection derived from cloned viral cDNA. *Proc. Natl. Acad. Sci. USA* **81**, 7066-7070.

Ahlquist, P., French, R. and Bujarski, J. J. (1987). Molecular studies of brome mosaic virus using infectious transcripts from cloned cDNA. *Adv. Virus Res* **32**, 215-242.

Boccard, F. and Baulcombe, D.C. (1992). Infectious *in vitro* transcripts from amplified cDNAs of the Y and Kin strains of cucumber mosaic virus. *Gene* **114**, 223-227.

Boccard, F. and Baulcombe, D.C. (1993). Mutational analysis of cis-acting sequences and gene function in RNA3 of cucumber mosaic virus. *Virology* **193**, 563-578.

Boyer, J-C. and Haenni, A-L. (1994). Infectious and cDNA clones of RNA virses. *Virology* **198**, 415-426.

Breitburd, F., Kirnbauer, R., Hubbert, N. L., Nonnenmacher, B., Trin-Dinh-Desmarquet, C., Orth, G., Schiller, J. T. and Lowy, D. R. (1995). Immunisation with virus-like particles from cottontail rabbit papillomavirus (CRPV) can protect against experimental CRPV infection. *J. Virology* **69**, 3959-3963.

Brisson, N., Paszkowski, J., Penswick, J. R., Gronenborn, B., Potrykus, I. and Hohn, T. (1984). Expression of a bacterial gene in plants by using a viral vector. *Nature* **310**, 511-514.

Campo, M. S., Grindlay, G. J., O'Neil, B. W., Chandcrachud, L. M., McGarvie, G. M. And Jarrett, W. F. H. (1993). Prophylatic and therapeutic vaccination against a mucosal papillomavirus. *J. Gen. Virol.* **74**, 945-953.

Cason, J., Khan, S. A. and Best, J. M. (1993). Towards vaccines against human papillomavirus type-16 genital infections. *Vaccine* **11**, 603-611.

Chen, B. and Francki, R. I. B. (1990). Cucumovirus transmission by the aphid *Myzus persicae* is determined solely by the viral coat protein. *J. Gen. Virol.* **71**, 939-944.

Chung, C. T. and Miller, R. H. (1989). A rapid and convenient method for the preparation and storage of competent bacteriaal cells. *Nucl. Acids Res.* **16**, 3580.

CMI 1979. Cucumber mosaic virus. *CMI/AAB Description of Plant Viruses* #213.

Coghlan, A. (1995). Grow your own vaccine. *New Scientist* **21 Jan.**, 23.

Commandeur, U., Jarausch, W., Koenig, Y. L. R. and Burgermeister, W. (1991). cDNAs of beet necrotic yellow vein virus RNAs 3 & 4 are rendered biologically active in a plasmid containing the cauliflower mosaic virus 35S promoter. *Virology* **185**, 493-495.

Davis, N. L., Willis, L. V., Smith, J. F. and Johnston, R. E. (1989). *In vitro* synthesis of infectious Venezuelan equine encephalitis virus RNA from a cDNA clone: Analysis of a viable deletion mutant. *Virology* **171**, 189-204.

Davies, C. and Symons, R. H. (1988). Further implications for the evolutionary relationships between tripartite viruses based on cucumber mosaic virus RNA 3. *Virology* **165**, 216-224.

Dawson, W. O., Beck, D. L., Knorr. D. A. and Grantham, G. L. (1986). cDNA cloning of the complete genome of tobacco mosaic virus and production of infectious transcripts. *Proc. Natl. Acad. Sci. USA* **83**, 1832-1836.

Dawson, W. O., Lewandowski, D.J., Hilf, M. E., Bubrick, P., Raffo, A. J., Shaw, J. J., Grantham, G. L. and Desjardins, P. R. (1989). A tobacco

mosaic virus hybrid expresses and loses an added gene. *Virology* **172**, 285-292.

Dessens, J. T. and Lomonossoff, G. P. (1993). Cauliflower mosaic virus 35S promoter-controlled DNA copies of cowpea mosaic virus RNAs are infectious on plants. *J. Gen. Virol.* **74**, 889-892.

De Zoeten, G. A., Penswick, J. R., Horisberger, M. A., Ahl, P., Schultze, M. and Hohn, T. (1989). The expression, localisation and effect of human interferon in plants. *Virology* **172**, 213-222.

Ding, B., Li, Q., Nguyen, L., Palukaitis, and Lucas, W. J. (1995) Cucumber mosaic virus 3a protein potentiates cell-to-cell trafficking of CMV RNA in tobacco plants. *Virology* **207**, 345-353.

Ding, S-W., Anderson, B. J., Haase, H.R. and Symons, R. H. (1994). New overlapping gene encoded by the cucumber mosaic virus genome. *Virology* **198**, 593-601.

Ding, S-W., Rathjen, J. P., Li, W-X., Swanson, R., Healy, H. and Symons, R. H. (1995). Efficient infection from cDNA clones of cucumber mosaic cucumovirus RNAs in a new plasmid vector. *J. Gen. Virol.* **76**, 459-464.

Donson, J., Kearney, C. M., Hilf, M. E. and Dawson, W. O. (1991). Systemic expression of a bacterial gene by a tobacco mosaic virus-based vector *Proc. Natl. Acad. Sci. USA* **88**, 7204-7208.

Dore, J-M. and Pinck, L. (1988). Plasmid DNA containing a copy of RNA3 can substitiute for RNA3 in alfalfa mosaic virus inocula. *J. Gen. Virol* **69**, 1331-1338.

Falco, S.C., Guida, T., Locke, M., Mauvais, J., Sanders, C., Ward, R.T. and Webber, P. (1995). Transgenic canola and soybeans with increased lysine. *Bio/technology* **13**, 577-582.

Finnegan, J. and McElroy, D. (1994). Transgene inactivation: Plants fight back! *Bio/technology* **12**, 883-888.

French, R., Janda, M. and Ahlquist, P. (1986). Bacterial gene inserted in an engineered RNA virus: efficient expression in monocotyledonous plant cells. *Science* **231**, 1294-1297.

Gendloff, E. H., Bowen, B. and Buchholtz, W. G. (1990). Quantitation of chloramphenicol acetyl transferase in transgenic tobacco plants by ELISA and correlation with gene copy number. *Plant Mol. Biol.* **14**, 575-583.

Ghim, S-J., Jenson, B. and Schlegel, R. (1992). HPV-1 protein expression in cos cells displays conformational epitopes found on intact virions. *Virology* **190**, 548-552.

Gilmer, D., Allmang, C., Ehresmann, C., Guilley, H., Richards, K., Jonard, G. and Ehresmann, D. (1993). The secondary structure of the 5'-non-coding region of beet necrotic yellow vein virus RNA3: evidence for the role in viral RNA replication. *Nucleic Acids Res.* **21**, 1389-1395.

Gleaves, A. P. (1992). A versatile binary vector system with a T-DNA organisational structure conducive to efficient integration of cloned DNA into the plant genome. *Plant Molecular Biology* **20**, 1203-1207.

Habili, N. and Symons, R. H. (1989). Evolutionary relationships between luteoviruses and other plant RNA viruses based on sequence motifs in their putative RNA polymerases and nucleic acid helicases. *Nucleic Acids Res.* **17**, 9543-9555.

Hamamoto, H., Sugiyama, Y., Nakagawa,N., Hashida, E., Matsunaga, Y. Takemoto, S., Watanabe, Y. and Okada, Y. (1993). A new tobacco mosaic virus vector and its use for the systemic production of angiotensin-I-converting enzyme inhibitor in transgenic tobacco and tomato. *Bio/technology* **11**, 930-932.

Haq, T. A., Mason, H. S., Clements, J. D. and Arntzen, C. J. (1995). Oral immunisation with a recombinant bacterial antigen produced in transgenic plants. *Science* **268**, 714-719.

Hayes, R. J. and Buck, K. W. (1990a). Complete purification of a eukaryotic virus RNA *in vitro* by a purified RNA-dependent RNA polymerase. *Cell* **63**, 363-368.

Hayes, R. J. and Buck, K. W. (1990b). Infectious cucumber mosaic virus RNA transcribed *in vitro* from clones obtained from cDNA amplified using the polymerase chain reaction. *J. Gen. Virol.* **71**, 2503-2508.

Herbers, K., Wilke, I. and Sonnewald, U. (1995). A thermostable xylanase from *clostridium thermocellum* expressed at high levels in the apoplast of transgenic tobacco has no detrimental effects and is easily purified. *Bio/technology* **13**, 63-66.

Hines, J. F., Ghim, S. J., Christensen, N. D., Kreider, J. W., Barnes, W. A., Schlegel, R. And Jensen, A. B. (1994). Role of conformational epitopes expressed by human papillomavirus major capsid proteins in the serological detection of infection and prophylactic vaccination. *Gynecologic Oncology* **55**, 13-20.

Horsh, R. B., Fry, J. E., Hoffman, N. L., Eichholtz, D., Rogers, S. D. & Fraley, R. T. (1985). A simple and general method for transferring genes into plants. *Science* **227**, 1229-1231.

Hooykaas, P. J. J. (1988). Agrobacterium molecuuar genetics. *Plant Molecular Biology Manual* **A4**, 1-13.

Janda, M., French, R. and Ahlquist, P. (1987). High efficiency T7 polymerase synthesis of infectious RNA from cloned brome mosaic virus cDNA and effects of 5' extensions on transcript infectivity. *Virology* **158**, 259-262.

Jarrett, W. F. H., Smith, K. T. Neil, B. W., Gaukroger, J. M., Chundrachud, L. M., Grindlay, G. J., McGarvie, G. M. And Campo, M.S. (1991). Studies on vaccination against papillomaviruses: prophylatic and therapeutic vaccination with recombinant structural proteins. *Virology* **184**, 33-42.

Joshi, R. L., Joshi, V. and Ow, D.W. (1990). BSMV genome mediated expression of a foreign gene in dicot and monocot plant cells. *EMBO J.* **9**, 2663-2669.

Kamer, G. and Argos, P. (1984). Primary structural comparison of RNA-dependent polymerases from plant, animal and bacterial viruses. *Nucleic Acids Res.* **12**, 7269-7281.

Kirnbauer, R., Booy, F., Cheng, N., Lowy, D. R. And Schiller, J. T. (1992). Papillomavirus L1 major capsid protein self-assembles into virus-like particles that are highly immunogenic. *Proc. Natl. Acad. Sci. USA* **89**, 12180-12184.

Kumagai, M. H., Turpen, T. H., Weinzettl, N., Della-Cioppa. G., Turpen, A. M., Donson, J., Hilf, M. E., Grantham, G. L. , Dawson, W. O., Chow, T.P., Piatak, M. Jr and Grill. L. K. (1993). Rapid, high-level expression of biologically active α-trichosanthin in transfected plants by an RNA viral vector. *Proc. Natl. Acad. Sci. USA* **90**, 427-430.

Lebens, M. Johansson, J. Osek, J. Lindblad, M. and Holmgren, J. (1993). Large scale production of Vibrio cholera toxin B subunit for use in oral vaccines. *Bio/technology* **11**, 1574-1578.

Lin, Y-L., Borenstein, L.A., Selvakumar, R. Ahmed, R. and Wettstein, F. O. (1992). Effective vaccination against papilloma development by immunisation with L1 or L2 structural protein of cottontail rabbit papillomavirus. *Virology* **187**, 612-619.

MacFarlane, S. A., Gilmer, D. and Davies, J. W. (1992). Efficient inoculation with CaMV 35 S promoter-driven DNA clones of the tobravirus PEBV. *Virology* **187**, 829-831.

Marsh, L. E., Pogue, G. P. And Hall, T.C. (1989). Similarities among plant virus (+) and (-) RNA termini imply a common ancestry with promoters of eukaryotic tRNAs. *Virology* **172**, 415-427.

Marsh, L. E., Huntly, C. C., Pogue, G. P., Connell, J. P. and Hall, T.C. (1991).Regulation of (+):(-)-strand asymmetry in replication of brome mosaic virus RNA. *Virology* **182**, 76-83.

Mason, H. S., Lam, D. M-K. and Arntzen, C. J. (1992). Expression of hepatitis B surface antigen in transgenic plants. *Proc. Natl. Acad. Sci. USA* **89**, 11745-11749.

Mason, W. S.m, Taylor, J. M. and Hull, R. (1987). Retroid virus genome replication. *Adv. Virus Res.* **32**, 35-96.

Miller, W. A., Bujarski, J. J., Dreher, T. W. and Hall, T. H. (1986). Minus-strand initiation by brome mosaic virus replicase within the 3' tRNA-like structure of native and modified RNA templates. *J. Mol. Biol.* **187**, 537-546.

Mise, K. Tsuge, S., Nagao, K., Okuno, T. and Furusawa, I. (1992). Nucleotide sequence responsible for the synthesis of a truncated coat protein of brome mosaic virus strain ATCC66. *J. Gen. Virol.* **73**, 2543-2551.

Mori, M., Kazuyuki, M., Kobayashi, K., Okuno, T. and Furusawa,I. (1991). Infectivity of plasmids containing brome mosaic virus cDNA linked to the cauliflower mosaic virus 35S RNA promoter. *J. Gen. Virol.* **72**, 243-246.

Mori, M., Zhang, G-H., Kaido, M., Okkuno, T. and Furusawa, I. (1993). Efficient production of human gamma interferon in tobacco protoplasts by genetically engineered brome mosaic virus RNAs. *J. Gen. Virol.* **74**, 1255-1260.

Nelson, T. (1994). Preparation of DNA and RNA from leaves: expanded blades and separated bundle sheath and mesophyll cells. *In* "The Maize Handbook" (Freeling, M. and Walbot, V., eds.), Chapter **91**, 541-545. Springer Verlag, New York..

Nitta, N., Masutaa, C., Kuwata, S. and Takanami, Y. (1988). Comparative studies on the nucleotide sequence of cucumber mosaic virus RNA 3 between Y and Kin strain. *Ann. Phytopathol. Soc. Jpn.* **54**, 516-522.

Odhiambo, T. R.(1994). Biotechnology: Africa's beachhead for entry into the global economy. *Bio/technology* **12**, 1400.

O'Reilly,.D., Thomas, C. J. R. and Coutts, R. H. A. (1991). Tomato aspermy virus has an evolutionary relationship with other tripartite RNA plant viruses. *J. Gen. Virol.* **72**, 1-7.

Owen, J. and Palukaitis, P. (1988). Characterisation of cucumber mosaic virus I: Molecular hetrogenaity mapping of RNA 3 in eight CMV strains. *Virology* **166**, 495-502.

Owen, J., Shjintaku, M., Aeschleman, P. , Tahar, S. B. and Palukaitis, P. (1990). Nucleotide sequence and nucleotide relationships of cucumber mosaic virus strains: CMV RNA 3. *J. Gen. Virol.* **71**. 2243-2249.

Palukaitis,P., Roossinck, M. J., Dietzgen, R. G. and Francki, R. I. B. (1992). Cucumber mosaic virus. *Adv. Virus. Res.* **41**. 281-348.

Penarrubia, L., Kim, R., Giovanni, J., Kim, S-H. and Fischer, R. L. (1992). Production of the sweet protein monellin in transgenic plants. *Bio/technology* **10**, 561-564.

Pogany, J., Huang, Q., Romero, J., Nagy, P. D. and Bujarski, J. J. (1994). Infectious transcripts from PCR-amplified broad bean mottle bromovirus cDNA clones and variable nature of leader regions in RNA3 segment. *J. Gen. Virol.* **75**, 693-699.

Pogue, G. P., Marsh, L.E., Connell, J. P. And Hall. T. C. (1992). Requirement for ICR-like sequences in the replication of brome mosaic virus genomic RNA. *Virology* **188**, 742-753.

Poirer, Y., Nawrath,C. and Comerville, C. R. (1995). Production of polyhydroxy- alkanoates, a family of biodegradable plastics and elastomers, in bacteria and plants. *Bio/technology* **13**,142-150.

Porta, C., Spall, V. E., Loveland, J., Johnson, J. E., Barker, P. J. and Lomonossoff, G. P. (1994). Development of cowpea mosaic virus as a

high yielding system for the presentation of foreign peptides. *Virology* **202**, 949-955.

Rezaian, M. A., Williams, R. H. V. and Symons, R. H. (1985). Nucleotide sequence of cucumber mosaic virus RNA 1: presence of a sequence complementary to part of the viral satellite RNA and homologies with other viral RNAs. *Eur. J. Biochem.* **150**,331-339.

Rizzo, T. M. and Palukaitis, P. (1988). Nucleotide sequence and evolutionary relationships of cucumber mosaic virus strains: CMV RNA 2. *J. Gen. Virol.* **69**, 1777-1787.

Rizzo, T. M. and Palukaitis, P. (1989). Nucleotide sequence and evolutionary relationships of cucumber mosaic virus strains: CMV RNA 1. *J. Gen. Virol.* **70**, 1-11.

Rose, R. C. *et al.* (1994). Serological differentiation of human papillomavirus types 11, 16 and 18 using recombinant virus-like particles. *J. Gen. Virol.* **75**, 2445-2449.

Rybicki, E. P. (1995). Bromoviridae. pp. 450-457 in Virus Taxonomy: Sixth Report of the International Committee on Taxonomy of Viruses," Eds. F. A. Murphy et al., Springer-Verlag, Wien.

Sambrook, J. , Fritsch, E. F. and Maniatis, T. (1989). "Molecular cloning: A Laboratory Manual," 2nd ed. Cold Spring Harbour Laboratory Press, Cold Spring Harbour, NY.

Shen, W-H. and Hohn, B. (1994). Amplification and expression of the β-glucuronidase gene in maize plants by vectors based on maize streak virus. *Plant J.* **5**, 227-236.

Shintaku, M. H., Zhang, L. and Paluksitis, P. (1992). A single amino acid substitution in the coat protein of cucumber mosaic virus induces chlorosis in tobacco. *Plant Cell* **4**, 751-757.

Sijmons, P. C., Dekker, B. M. M., Schrammeijer, B., Verwoerd, T. C., Van den Elzen, P. J. M. and Hoekema, A. (1990). Production of correctly

processed human serum albumin in transgenic plants. *Bio/technology* **8**, 217-221.

Somerville, C. R. (1993). Production of industrial materials in transgenic plants. *Phil. Trans. R.Soc. Lond.* **342**,251-257.

Stanley, J. (1993). Geminiviruses: plant viral vectors. *Current Opinion in Genetics and Development* **3**, 91-96.

Suzuki, M., Kuwata, S., Kataoka, J., Masuta, C., Nitta, N. and Takanami, Y. (1991). Functional analysis of deletion mutants of cucumber mosaic virus RNA3 using an *in vitro* transcription system. *Virology* **183**, 106-113.

Takahashi, H., Goto, N. and Ehara, Y. (1994). Hypersensitive response in cucumber mosaic virus inoculated *Arabidopsis thaliana. Plant Journal* **6**, 369-377.

Thanavala, Y., Yang, Y-F., Lyons, P., Mason, H. S. and Arntzen, C. (1995). Immunogenicity of transgenic plant derived hepatitis B surface antigen. *Proc. Natl. Acad. Sci. USA* **92**, 3358-3361.

Turpen, T. H., Reinl, S. J., Charoenvit, Y., Hoffman, S. L., Fallarme, V. and Grill, L. K. (1995). Malarial epitopes expressed on the surface of recombinant tobacco mosaic virus. *Bio/technology* **13**, 53-56.

Usha, R., Rohll, J. B., Spall, V. E., Shanks, M., Maule, A. J., Johnson, J. E. and Lomonossoff, G. P. (1993). Expression of an animal virus antigenic site on the surface of a plant virus particle. *Virology* **197**, 366-374.

Vacquero, C., Turner, A. P., Demangeat, G., Sanz, A., Serra, M. T., Roberts, K. and Garcia-Luque, I. (1994)The 3a protein from cucumber mosaic virus increases gating capacity of plasmodesmata in transgenic tobacco plants. *J. Gen. Virol.* **75**, 3193-3197.

Valverde, R. A. (1990). Analysis of double-stranded RNA for plant virus diagnosis. *Plant Disease* **74**, 255-258.

Van Bokhoven, H., Verver, J., Wellink, J. and Van Kammen, A. (1993). Protoplasts transiently expressing the 200K coding sequence of cowpea mosaic virus B-RNA support replication of M-RNA. *J. Gen. Virol.* **74**, 2233-2241.

Vandekerckhove, J., Van Damme, J., Van Lijsebettens, M., Botterman, J., De Block, M., Vandewiele, M., De Clerq, A., Leemans, J., Van Montagu, M. and Krebbers, E. (1989). Enkephalins produced in transgenic plants using modified 2S seed storage proteins. *Bio/technology* **7**, 929-932.

Van der Vossen, E. A. G., Neeleman, L. and Bol. J. F. (1993). Role of the 5' leader sequence of alfalfa mosaisc virus RNA 3 in replication and translation of the viral RNA. *Nucl. Acids Res.* **21**, 1361-1367.

Vervoerd, T. C., Dekker, B. M. M. and Hoekma, A. (1989). A small-scale procedure for the rapid isolation of plant RNAs. *Nucl. Acids Res.* **17**, 2362.

Von Bodman,S. B., Domier, L. L. And Farrand, S. K. (1995). Expresson of multiple eukaryotic genes from a single promoter in *Nicotiana*. *Bio/Technology* **13**, 587-591.

Yamaya, J., Yoshioka, M., Meshi, T., Okada, Y. and Ohno, T. (1988). Expression of tobacco mosaic virus RNA in transgenic plants. *Mol. Gen. Genet.* **211**, 520-525.

Zaccomer, B., Haenni, A-L and Macaya, G. (1995). The remarkable variety of plant RNA virus genomes. *J. Gen. Virol.* **76**, 231-247.

Zur Hausen, H. And De Villiers, E-M. (1994). Human Papillomaviruses. *Annu. Rev. Microbiol.* **48**, 427-447.