Investigation of RNAs 4A and 5 from Cucumoviruses

by

Christopher L. Blanchard
B. App. Sci. (Hons)

A thesis submitted to Charles Sturt University for the degree of Doctor of Philosophy

CSIRO Division of Plant Industry
Canberra, A.C.T.,
Australia.  
February 1996.
Declaration of Originality

This thesis represents the original work of the author apart from the assistance listed in the acknowledgments. To the best of my knowledge this work has not been previously written or published by any other person. The research was done at Charles Sturt University, Wagga Wagga and CSIRO Division of Plant Industry, Canberra under the joint supervision of Dr B. J. Anderson and Dr T. J. Higgins.

C. L. Blanchard
Acknowledgments

I wish to express my appreciation to the many people who assisted me over the past four years. First and foremost I would like to thank the past and present members of the B. J. A. group of scientific excellence. Our esteemed leader, Dr Beau Anderson has always been a tower of strength, and her enthusiasm and constant support has made her the best supervisor a student could ever hope for. I would like to thank my fellow PhD student, Andrew Walter, for bringing a new dimension into our lab, being a good bloke, and keeping us all amused with his tales of a small Victorian town. I would like to acknowledge the tremendous amount of assistance I received from Trish Boyce in the initial characterisation of RNA 5 as well as her friendship over the past few years. To other members of the B. J. A. group of scientific excellence, Sean Hannon and Olaf Reger I would like to express my gratitude for their assistance and for providing an enjoyable working environment.

I wish to thank my other supervisor, Dr T. J. Higgins for warmly welcoming me into his lab at a critical time in my project and thank him for his support and advice. I would also like to thank all the members of T, J.’s group for valuable scientific advice and for providing an exciting and stimulating working environment. I would like to thank my fellow PhD student and partner in crime at CSU, Steven Simpfendorfer, for his support especially during overseas travel. I am also grateful for the technical assistance received from Matthew Gooden in the sequencing of RNA 5 clones. I am especially grateful to Dr Paul Chu for teaching me everything I know about virus purification and inoculation, and for spending many hours assisting me with these procedures.

I would like to thank the support technical staff at Charles Sturt University and CSIRO for assisting me in my work, especially Ken Simpson and all the lab staff at CSU, Roger Mummery for looking after my neglected plants at CSIRO, Cecila Miller for help with the electron microscope, and the CSIRO visual resources staff; you guys are great. Thanks also to Ros Joseph for many technical tips. I would also like to thank the members of the Virology group for useful advice on my project. Thanks also to John Watson, Don Spencer, Kaye Bateman, Amanda Ellery and Andrew Walter for reading my thesis and providing useful comments.

Special thanks to Dr Peter Palukaitis for useful discussions, making available unpublished data on possible subgenomic promoter elements, and for boosting my enthusiasm for plant virus research. I am also grateful to Ron Garret for providing Pri-CMV and Fernando Garcia-Arenal for providing V-TAV and the sequence of V-TAV RNA 3.
Thanks also to my wife for her encouragement and support throughout my PhD.

This work was supported by an Australian Postgraduate Award and a Grains Research and Development Corporation Junior Fellowship. I would also like to acknowledge generous support from the director of the Centre for Conservation Farming Ass. Prof. Jim Pratley for his never ending financial generosity and the South Australian Research and Development Institute for funding overseas travel.
Abstract

Cucumber mosaic virus (CMV) is a serious plant pathogen having a wide host range which includes a number of commercially important crop species. Studies to understand the structure of CMV have been undertaken to enable the design of strategies to genetically engineer CMV-resistant plants using pathogen-derived resistance. CMV has also been well studied as it is the type member of the cucumovirus group. CMV consists of three genomic RNAs (RNAs 1, 2, and 3) as well as two subgenomic RNAs (RNA 4 and 4A) and a small population of 3' coterminal molecules termed RNA 5. RNA 5 is the least-characterised CMV RNA and it is not known from which genomic RNA the molecules are derived, or how they are synthesised. It is also not known how widely distributed RNA 4A and 5 are throughout other cucumovirus strains.

This thesis describes the characterisation of Q-CMV RNA 5 by the sequencing of a number of RNA 5 cDNA clones to determine which genomic RNA RNA 5 is derived from. Sequence analysis indicated that the RNA 5 population consisted of molecules derived from both the 3' terminal 307 nt of RNA 2 and the 3' terminal 304 nt of RNA 3. Primer extension analysis was undertaken on purified RNA 5 to confirm the presence of these two sequence variants in the RNA 5 population and to ensure that there were no molecules derived from RNA 1 present. Similarities between sequences around the start site of RNA 5 on the genomic RNAs and previously-characterised subgenomic RNAs indicate that RNA 5 may be synthesised as a subgenomic RNA despite the fact that RNA 5 does not appear to be capped or encode a protein.

A survey of various cucumovirus strains was carried out to determine if RNA 4A and 5 were commonly associated cucumovirus strains. RNA 4A was found to be present in all strains examined which included strains from both subgroups of CMV and a strain from another cucumovirus, tomato aspermy virus (TAV). This supports the proposal that RNA 4A has an important role in CMV or was important for the evolution of the cucumoviruses. RNA 5 was found to be present in the TAV strain and all the CMV subgroup II strains examined, but only one of the two CMV subgroup I strains examined contained RNA 5. This indicates that RNA 5 may be important in the CMV subgroup II and TAV life cycle but is not essential to all CMV subgroup I strains.

The use of RNA 5 to genetically engineer CMV-resistant plants was investigated by expressing RNA 5 in transgenic tobacco plants. A strategy to successfully engineer plants using RNA 5 was designed after it was found that unmodified RNA 5 expressed in transgenic plants was not able to protect the plants from CMV infection. This
alternative strategy involved the use of a ribozyme to remove extraneous 3' sequences which may interfere with the formation of secondary structures thought to be involved with viral replicase recognition. Also, a 5' terminal viral replicase recognition sequence was included to increase the level of expression of RNA 5 upon infection with CMV or other viruses which have replicase complexes capable of recognising replicating RNA 5.
Table of Contents

CHAPTER 1: LITERATURE REVIEW ........................................................................................................... 1
  1.1. INTRODUCTION.................................................................................................................................... 1
  1.2. CLASSIFICATION OF CUCUMOVIRUSES ....................................................................................... 1
    1.2.1. Classification of Plant Viruses ..................................................................................................... 1
    1.2.2. Cucumber Mosaic Virus ............................................................................................................... 4
    1.2.2.1. Host Range and Symptoms .................................................................................................... 4
    1.2.2.2. Particle Structure .................................................................................................................... 4
    1.2.2.3. Molecular Structure .............................................................................................................. 5
    1.2.2.4. Genomic RNAs ....................................................................................................................... 6
    1.2.2.5. Other RNAs Associated with CMV ...................................................................................... 7
    1.2.2.6. 5' and 3' Terminal Structures on CMV RNAs ..................................................................... 9
    1.2.2.7. Promoter Regions and Viral Replicase Signals .................................................................... 9
    1.2.2.8. Replication ............................................................................................................................. 11
    1.2.3. Other Viruses in the Cucumovirus Group .................................................................................. 12
  1.3. CMV IN LUPINS ............................................................................................................................... 12
    1.3.1. The Problem of CMV in Lupins ................................................................................................. 12
    1.3.2. Approaches to the Control of CMV in Lupins .......................................................................... 13
      1.3.2.1. Management Practices ....................................................................................................... 13
      1.3.2.2. Breeding CMV Resistant Plants ......................................................................................... 13
      1.3.2.3. Genetically Engineering Virus Resistant Plants ............................................................... 13
  1.4. AIM OF PROJECT ............................................................................................................................ 14

CHAPTER 2: MATERIALS AND METHODS ............................................................................................ 16
  2.1. PROPAGATION AND PURIFICATION OF VIRUS AND VIRAL RNA ............................................ 16
  2.2. PLASMID DNA SEQUENCING .................................................................................................... 17
    2.2.1. Manual Sequencing .................................................................................................................... 17
    2.2.2. Automated Sequencing ............................................................................................................. 17
  2.3. SEQUENCING OF RNA 5 CDNA CLONES .................................................................................... 17
  2.4. CLONING AND SEQUENCING THE 3' END OF RNAs 1 AND 2 ................................................ 17
  2.5. REVERSE TRANSCRIPTASE AND POLYMERASE CHAIN REACTION ................................ 18
  2.6. COMPUTER ANALYSES ............................................................................................................... 19
  2.7. PURIFICATION OF Q-CMV RNA 5 ............................................................................................... 19
  2.8. TOTAL RNA EXTRACTION ........................................................................................................... 19
  2.9. IN VITRO RNA TRANSCRIPTION ............................................................................................... 20
  2.10. PRIMER EXTENSION REACTION .............................................................................................. 20
  2.11. TOBACCO TRANSFORMATION ................................................................................................. 21
  2.12. PHOSPHINOTHICIN ACETYLTRANSFERASE (PAT) ASSAY ..................................................... 21
  2.13. VIRUS CHALLENGE ..................................................................................................................... 21
  2.14. SUBGROUP DETERMINATION OF L- AND PRI-CMV ............................................................... 22

CHAPTER 3: CHARACTERISATION OF Q-CMV RNA 5 ............................................................................. 23
  3.1. INTRODUCTION ............................................................................................................................... 23
  3.2. RESULTS .......................................................................................................................................... 26
    3.2.1. Sequencing of 20 RNA 5 cDNA Clones .................................................................................. 26
    3.2.2. Computer Alignment of RNA 5 Sequences .............................................................................. 27
    3.2.3. Optimisation of Primer Extension Reactions as a Technique for Differentiating Similar RNA
      Molecules ............................................................................................................................................ 29
    3.2.4. Primer Extension Analysis of Q-CMV RNA 5 ........................................................................ 32
    3.2.5. Is RNA 5 Derived from a Subgenomic Promoter on RNAs 2 and 3 ? ................................... 34
    3.2.6. Does RNA 5 Encode a Protein ? ............................................................................................... 35
  3.3. DISCUSSION ..................................................................................................................................... 36

CHAPTER 4: SURVEY FOR RNAs 4A AND 5 IN V-TAV AND A NUMBER OF CMV
  SUBGROUP I AND II STRAINS ............................................................................................................. 42
  4.1. INTRODUCTION ............................................................................................................................... 42
  4.2. RESULTS .......................................................................................................................................... 43
    4.2.1. RNA 4A is Present in Both Subgroup I and II Strains of CMV ............................................... 43
    4.2.2. RNA 4A is Present in Another Cucumovirus, V-TAV .......................................................... 45
    4.2.3. Survey of Subgroup I and II Strains of CMV for RNA 5 ......................................................... 46
    4.2.4. There are Two Sequence Variants in the RNA 5 Population in Subgroup II CMV Strains .... 49
    4.2.5. RNA 5 is Present in Another Cucumovirus, V-TAV .............................................................. 51
4.2.6. The CMV Strain Which Does Not Contain RNA 5 Produces More Severe Symptoms Than Those That Do Contain RNA 5 .................................................................................................... 52
4.3. DISCUSSION ............................................................................................................................... 53

CHAPTER 5: INVESTIGATION OF THE POSSIBLE USE OF Q-CMV RNA 5 TO GENETICALLY ENGINEER VIRUS-RESISTANT PLANTS ............................................................................................................................... 58
5.1. INTRODUCTION .......................................................................................................................... 58
5.2. RESULTS ...................................................................................................................................... 65
  5.2.1. Subcloning of Q-CMV RNA5 cDNA into Binary Vectors .................................................... 65
  5.2.2. Transformation of N. tabacum with the Binary Vector Containing Q-CMV RNA 5 .............. 66
  5.2.3. PAT Assay of Putative Transgenic Plants ............................................................................. 66
  5.2.4. Presence of the RNA 5 Transcript in Transgenic Plants ...................................................... 67
  5.2.5. Virus Challenge of Transgenic Plants .................................................................................. 68
5.3. DISCUSSION ................................................................................................................................ 68

CHAPTER 6: CONCLUSIONS AND FUTURE DIRECTIONS ....................................................................... 75
6.1. CHARACTERISATION OF Q-CMV RNA 5 INDICATES RNA 5 IS DERIVED FROM RNA 2 AND 3 AS A SUBGENOMIC RNA ......................................................................................... 75
6.2. RNA 4A AND 5 OCCUR IN MOST CUCUMOVIRUSES ................................................................ 77
6.3. ENGINEERING VIRUS RESISTANCE WITH RNA 5 .................................................................. 80

LITERATURE CITED ............................................................................................................................... 82

APPENDIX 1: PUBLICATIONS ARISING FROM THESIS ........................................................................ 98
List of Figures

Fig. 1.1: Classification of plant viruses based on the nature of genetic material and presence or absence of an envelope, 2
Fig. 1.2: Structure of CMV RNAs, 5
Fig. 1.3: Alignment of putative RNA 5 subgenomic promoter elements 1 and 2 to ICR2-like motifs, 10
Fig. 1.4: Location of putative subgenomic promoter elements on Q-CMV RNAs 3, 11
Fig. 1.5: Alignment of putative RNA 5 subgenomic promoter element 3 to sequences around the RNA 4 start site, 11
Fig. 3.1: Sequence comparison of BMV and alphavirus subgenomic promoters, 25
Fig. 3.2: Alignment of RNA 5 and genomic RNA cDNA sequences, 28
Fig. 3.3: Alignment of consensus sequences from the three putative groups of RNA 5 sequence variants, 29
Fig. 3.4: Primer binding sites on aligned RNA 5.1 and RNA 5.2 sequences, 30
Fig. 3.5: Primer extension analysis of RNA 5.1 and 5.2 using different primers, 31
Fig. 3.6: Binding regions of RNA 5-specific primers, 32
Fig. 3.7: Primer extension analysis of QCMV RNA 5, 33
Fig. 3.8: Alignment of the postulated alphavirus subgenomic promoter consensus sequence with Q-CMV RNA 1, 2 and 3 sequence, 35
Fig. 4.1: Binding site for the primer P4AII on Q-CMV RNA 4A, 43
Fig. 4.2: Detection of RNA 4A in subgroup II CMV strains by primer extension analysis, 44
Fig. 4.3: Binding site of P4AI to putative subgroup I RNA 4A sequences, 44
Fig. 4.4: Detection of RNA 4A in subgroup I CMV strains by primer extension analysis, 45
Fig. 4.5: Binding site for P4A-TAV on putative RNA 4A from TAV, 45
Fig. 4.6: Detection of RNA 4A in V-TAV by primer extension analysis, 46
Fig. 4.7: Conserved region present in all TAV and CMV strains that was used to detect RNA 5 in cucumoviruses, 47
Fig. 4.8: Binding site of the primer PCMV5 on Q-CMV RNA 5 molecules, 47
Fig. 4.9: Detection of RNA 5 in subgroup I and II CMV strains, 48
Fig. 4.10: Examination of total RNA isolated from L- and Ny-CMV infected plants for the presence of RNA 5, 49
Fig. 4.11: Binding site of P5-doublet on Q-CMV RNAs 5.2 and 5.3, 49
Fig. 4.12: Detection of two sequence variants in the RNA 5 population of subgroup II CMV strains, 50
Fig. 4.13: Binding site of PCMV5 on putative V-TAV RNA 5 molecules, 51
Fig. 4.14: Detection of RNA 5 in V-TAV, 52
Fig. 4.15: Symptoms displayed by N. clevelandii infected with various strains of CMV, 53
Fig. 4.16: Alignment of Q-CMV and V-TAV 3' terminal consensus sequences, 54
Fig. 4.17: Alignment of Q-CMV and Fny-CMV 3' terminal consensus sequences, 55
Fig. 4.18: Location of Q-CMV RNA 5 5' terminal sequence on BNYVV RNA 3, 56
Fig. 5.1: Binary plasmid containing Q-CMV RNA 5 expression cassette, 65
Fig. 5.2: PAT assay on putative transgenic plants, 66
Fig. 5.3: Determination of the presence of RNA 5 in transgenic plants by RT-PCR, 67
Fig. 5.4: CMV RNA 5 construct containing a 3' end ribozyme, 70
Fig. 5.5: Self-cleavage analysis of RNA 5 construct containing a 3' end cis-acting ribozyme, 71
Fig. 5.6: Amplification of RNA 5 using a 5' end promoter, 72
Fig. 5.7: (+) strand synthesis directed by RNA 2 5' terminal sequence, 73
## List of Tables

Table 1.1. Subgroup designation of CMV strains mentioned in this thesis, 3
Table 2.1 Virus isolates employed in these studies, 16
Table 2.2. Computer programs used to analyse sequences, 19
Table 3.1: Sizes of putative primer extension products using different primers on RNA 5.1 and 5.2, 30
Table 3.2: ORFs present in RNA 5.2 (clones a21 and b69) and RNA 5.3 (clone a22), 36
Table 3.3: Summary of RNA 5 sequence variation, 37
Chapter 1: Literature Review

1.1. Introduction

Cucumber mosaic virus (CMV) is a serious pathogen in crop species around the world and as a result has been well characterised. Much of this characterisation has focused on the components of the virus which make up the genome. At an early stage an additional component was found to be responsible for expression of the virus coat protein. Other CMV-associated molecules have been identified recently and preliminary investigations of their structure and function have been undertaken (Ding et al. 1994; Ding et al. 1995b; Blanchard et al. 1991; Palukaitis et al. 1992). Full characterisation of these newly-identified RNAs will further our understanding of the biology of the virus and may provide information on the taxonomy of CMV. This knowledge may also allow us to develop strategies to genetically engineer plants which are resistant to CMV.

1.2. Classification of Cucumoviruses

1.2.1. Classification of Plant Viruses

The first characteristic which is used to classify plant viruses is the composition and nature of their genetic material. This can be composed of either deoxyribonucleic acid (DNA) or ribonucleic acid (RNA). The genomic DNA or RNA can be in the double-stranded (ds) or single-stranded (ss) form. If the nucleic acid is in the ss form then it may be in the positive sense (i.e., virion RNA has the same polarity as mRNA) or negative sense. Further divisions can be made based on the presence or absence of an envelope (a protective covering in addition to the coat protein). Examples of plant viruses with the various types of genomes are illustrated in Fig. 1.1. The genome of cucumoviruses, like most viruses which infect higher plants, is ss positive sense RNA.

Most of the ss positive sense viruses can be grouped into two super families. The first of these, which includes cucumoviruses, contain a cap structure on the 5’ end of the genomic RNAs. The other family consists of viruses which have a small viral encoded protein (VPg) covalently linked to the 5’ end of the genomic RNAs (Cornelissen 1989). Viruses containing a 5’ end cap can be further divided into three groups based on the proteins encoded by the genomic RNAs. The first of these groups can be further divided into tobamoviruses, tobraviruses, and the Bromoviruses (Cornelissen 1989). The Bromoviridae family encompasses alfamo, bromo, cucumo, and ilaraviruses.
Fig. 1.1: Classification of plant viruses based on the nature of genetic material and presence or absence of an envelope. Representative virus groups which reflect the nature of their genome and particle morphology are shown diagrammatically (from Francki et al. 1991).
The type member of the cucumovirus group is cucumber mosaic virus (CMV, van Vloten-Doting et al. 1981). CMV strains can be divided into two subgroups, termed I and II (Owen and Palukaitis 1988). The genomic RNAs from a number of CMV strains from both subgroups have been partially or completely sequenced (Cuozzo et al. 1988; Hayakawa et al. 1989a, b; Noel and Tahar 1989; Quemada et al. 1989; Rezain et al. 1984, 1985; Davies and Symons 1988; Nitta et al. 1988; Rizzo and Palukaitis 1988 1989; Kataoka et al. 1990a,b; Owen et al. 1990; McGarvey et al. 1995). This has allowed comparisons to be made between strains from the two subgroups (Palukaitis et al. 1992). The strains mentioned in this thesis are listed in Table 1.1. Other viruses belonging to the cucumovirus group include bean distortion mosaic virus (BDMiV, White et al. 1995), tomato aspermy virus (TAV) and peanut stunt virus (PSV). The genomic RNAs of the V strain of TAV (Bernal et al. 1991; Moriones et al. 1991; F. García-Arenal, personal communication) and J strain of PSV (Karasia et al. 1991; Karaswa et al. 1992) have been completely sequenced.

<table>
<thead>
<tr>
<th>Subgroup</th>
<th>Strain</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>As, C, D, Fny, i17f, Japanese Y, L, M, nt9, Ny, O, Ix, Try, Y.</td>
</tr>
<tr>
<td>II</td>
<td>B14, Kin, P, Pri, Q, Sn, Wc, Wl.</td>
</tr>
<tr>
<td>Unknown</td>
<td>Ywa</td>
</tr>
</tbody>
</table>

Serological and cross-protection tests were used to establish the relationships between cucumoviruses. Studies with different isolates of CMV and TAV (Hollings, et al. 1968; Stace-Smith and Tremaine 1973; Devergne and Cardin 1975) established a distinct serological relationship between TAV and CMV. However, it has been suggested that this serological relationship should be viewed with some caution as it was established that, in a number of cases, antisera had been prepared to a mixture of the two viruses (Mink et al. 1975). Based on biological, chemical and serological properties of V-TAV and Q-CMV, Habili and Francki (1974a) concluded that these two cucumoviruses were different viruses and not merely strains of the same virus. However, based on comparisons between RNA sequences, it has been proposed that TAV, Q-CMV and Fny-CMV may represent three equivalent subgroups of a taxonomic entity (Bernal et al. 1991). PSV was also shown to be distantly related to both CMV and TAV (Devergne and Cardin 1975). The study of CMV RNA 4A and 5 may provide information on how cucumoviruses should be classified by determining if the viruses in the different taxonomic groups have a different ability to synthesise these RNAs. If a difference is observed between the various cucumovirus groups, the presence of RNA 5 may be used as an indicator of which subgroup they belong to.
1.2.2. Cucumber Mosaic Virus

1.2.2.1. Host Range and Symptoms

CMV has a world-wide distribution and a broad host range. At least 191 plant species from 40 families have been shown to be susceptible to CMV (Murant and Harrison 1979). CMV may cause a number of symptoms in plants which include: mosaic of cucumber and many other cucurbits, celery, and many other species of dicotyledonous and monocotyledonous crops, ornamental and weed plants including some trees and shrubs, blight of spinach and fern leaf of tomato. CMV may also cause plant dwarfing and flower breaking (Murant and Harrison 1979).

1.2.2.2. Particle Structure

Virus particles of Q-CMV contain 18% ribonucleic acid (RNA) which is composed of 22.4% adenine, 24.7% guanine, 22.8% cytosine, and 30.1% uracil. The molecular weight of the virus is approximately 5.8 x 10^6 daltons. When stained in phosphotungstic acid or uranyl acetate, electron micrographs of Q-CMV preparations show isometric particles which are approximately 28 µm in diameter (Francki et al. 1966).

The CMV capsid itself is composed of 180 identical protein subunits (Finch et al. 1967) which have a Mr of 24, 500 daltons (van Regenmortel et al. 1972; Habili and Francki 1974a). They are arranged in pentamer-hexameter clusters with T=3 surface lattice symmetry. Neutron small-angle scattering of CMV particles in solution has shown that the particles have an external diameter of 29 µm and an internal diameter of 16.5 µm (Jacrot et al. 1977).

Virus particles are stabilised by interactions between the coat protein and the RNA contained in the particles (Kaper and Geelen 1971). When these interactions are disrupted the particles readily dissociate into the RNA and protein components. With the removal of the destabilising agent, the components can re-assemble to form biologically active virus particles (Kaper 1969). If there is a mixture of components from more than one virus during the re-assembly process, heterologous particles will be formed. This has been demonstrated with the construction of hybrid virus particles containing RNA and protein from both CMV and TAV (Kaper and Geelen 1971). CMV coat protein has also been shown to encapsidate turnip yellow mosaic virus (Kaper and Geelen 1971) and tobacco mosaic virus (Chen and Francki 1990) even
though these RNAs are both approximately 50% larger than the largest CMV RNA molecule.

1.2.2.3. Molecular Structure

Like other members of the Bromoviridae family, CMV contains a tripartite genome (i.e., the genome consists primarily of 3 separate molecules of RNA with an additional fourth subgenomic RNA). Two additional RNAs (4A and 5) have also been associated with some CMV strains and they contain sequences derived from genomic RNAs. The structure of the CMV RNAs is illustrated diagrammatically in Fig 1.2.

Fig. 1.2: Structure of CMV RNAs. The designated RNA number and size of the RNAs are indicated. Coding regions are indicated by a box and the designated protein names are shown above the boxes. Putative regions from which RNA 5 has been proposed to be derived are indicated and labelled with the reference of the report. RNAs which are
known to contain a 5' cap are labelled with a m$^7$G

1.2.2.4. Genomic RNAs

RNA 1 encodes a single protein with an approximate molecular weight of 111,000 daltons termed 1a (Rezain et al. 1985). Sequence similarity with other viruses (Cornelissen and Bol 1984; Haseloff et al. 1984; Rezain et al. 1985) indicates the 1a protein forms part of the viral RNA-dependent RNA polymerase (RdRp or replicase). Proposed functions for this protein include initiation of (-) RNA synthesis (Hahn et al. 1989), nucleotide binding (Rizzo and Palukaitis 1989), helicase activity (which is needed to unwind duplex RNA, Hodgeman 1988) and methyltransferase activity (which is thought to be involved in the capping of viral RNA, Mi et al. 1989).

The amino acid sequence similarity between the 1a proteins from Fny- and Q-CMV is 85% (Palukaitis et al. 1992) which is significantly higher than the nucleotide sequence similarity. The nucleotide sequence similarity between RNA 1 molecules of the two subgroups varies along the length of the molecule. The sequence similarity between Fny- and Q-CMV RNA 1 in the 5' noncoding region is 81%, whereas the sequence similarity between the 3' terminal noncoding regions is only 64% (Palukaitis et al. 1992). The 3' terminal 180 nucleotides (nt) of the two RNAs fold up to form very similar secondary structures (Rizzo and Palukaitis 1989).

Comparison of amino acid sequences from proteins of related viruses, as well as direct studies on purified RdRp (Hayes and Buck 1994), has provided evidence to indicate that the protein encoded by RNA 2 (2a) also forms part of the viral replicase complex. This is supported by the presence of a glycine-aspartic acid-aspartic acid (GDD) sequence motif characteristic of all RdRp (Koonin 1991). Both the 1a and 2a proteins have been detected in CMV-infected cells and shown to be associated with the replication complex (Hayes and Buck 1994). Isolation of a functional replicase complex from CMV-infected host plants (Hayes and Buck 1990) indicated that in addition to the 1a and 2a proteins, this complex contained one major and several minor host-encoded proteins. Q-CMV RNA 2 also encodes an overlapping protein (2b) which is expressed via a subgenomic RNA (see RNA 4A below).

As with RNA 1, sequence similarity between Fny- and Q-CMV RNA 2 varies along the length of the molecule with the 5' end sequences having higher similarity than the 3' terminal sequences (Palukaitis et al. 1992). The nucleotide sequence similarity between RNA 2 of Fny- and Q-CMV is 71% which is similar to the amino acid sequence similarities of the 2a proteins (73%, Palukaitis et al. 1992).
RNA 3 encodes two proteins. The 5' ORF encodes the 3a protein with a molecular weight of approximately 30,000 daltons. By analogy with other related viruses (Davies and Symons 1988) and recent experimental evidence (Ding et al. 1995a), it has been proposed that the 3a protein facilitates cell-to-cell movement of the virus. This was demonstrated by co-injection of 9.4-kDa fluorescein-conjugated dextran or fluorescently-labelled CMV RNA with unlabelled CMV 3a protein which facilitated movement of the labelled molecules to neighbouring cells. Expression of the 3a protein in transgenic plants also facilitates cell-to-cell movement of labelled dextran and CMV RNAs in these plants. The second protein encoded by RNA 3 is the coat protein which is expressed via a subgenomic RNA (RNA 4).

### 1.2.2.5. Other RNAs Associated with CMV

RNA 4 is a subgenomic RNA identical to the 3' terminal 1 kb region of RNA 3 (Schwinghamer and Symons 1975, 1977; Gould and Symons 1978) and has been sequenced in a number of CMV strains (Cuozzo et al. 1988; Nitta et al. 1988; Davies and Symons 1988; Quemada et al. 1989; Hayakawa et al. 1989a; Owen et al. 1990, Anderson et al. 1995). RNA 4 and the coat protein are useful for subgroup determination of CMV strains by using serological assays with the coat protein (Palukaitis et al. 1992), hybridisation studies with RNA 4 (Owen and Palukaitis 1988) and the use of a subgroup II-specific Eco RI site on RNA 4 (Wylie et al. 1993; Anderson et al. 1995).

A number of studies have indicated the presence of discrete populations of RNA molecules which are encapsidated in addition to the four major RNAs (Peden and Symons 1973; Boccard and Baulcombe 1993; Symons 1978; Gould et al. 1978; Ding et al. 1994; Avila-Rincon 1986). These have been termed RNAs 4A, 5 and 6. These RNAs appear to be unique to cucumoviruses and may account for some of the peculiar features of this virus group. These features include the very wide host range of CMV and the ability to support the replication of satellite RNAs.

Q-CMV RNA 4A has recently been shown to be a subgenomic RNA, 682 nt in length, which encodes a protein (2b, see Fig. 1.1) with a predicted molecular mass of 11,300 daltons found at the 3' end of RNA 2 overlapping the 2a ORF (Ding et al. 1994). When cotyledons of cucumber plants were inoculated with virus mutants lacking this ORF, the virus was unable to spread systemically, suggesting that the 2b protein may be involved in long-distance movement of virus (Ding et al. 1995c). RNA 4A has not been characterised in virus strains other than Q-CMV. However, Ding et al. (1994) predicted
RNA 4A would be present in all cucumoviruses based on conservation of the putative 2b protein and RNA 4A nucleotide sequences in these strains. Further studies are thus required to confirm the presence of RNA 4A in other CMV strains as well as in other cucumoviruses.

Initial studies of RNA 5 indicated that it was a population of molecules with an approximate molecular weight of 0.12 x 10^6 daltons which forms a well-defined band on agarose gels (Peden and Symons 1973; Gould et al. 1978; Symons 1978). There have been two conflicting hypotheses regarding the origin of RNA 5. Blanchard (1991) reported that RNA 5 is almost identical in sequence to the 3' end of RNA 2 whereas Palukaitis et al. (1992) reported RNA 5 to be co-terminal with the 3' end of RNA 3 and, by implication, RNA 4. They, therefore concluded that RNA 5 was somehow derived from RNA 2 and 3, respectively. Both studies had their limitations as Blanchard (1991) only analysed one RNA 5 cDNA clone which did not allow for the possibility of there being a mixed population and Palukaitis et al. (1992) based their conclusion on hybridisation experiments (which are made difficult by the high sequence similarity at the 3' ends of CMV RNAs), direct sequencing of RNA 5 from two CMV strains, and the presence of putative subgenomic promoter sequences on RNA 3 (P. Palukaitis, personal communication). Both of these studies are incomplete and a more complete analysis of RNA 5, which should include analysis of a number of RNA 5 cDNA clones as well as a direct analysis of purified RNA 5, is required to determine the true nature of the RNA 5 population.

CMV strains from subgroups I and II appear to contain a population of small RNAs (RNA 6) consisting of a mixture of plant RNAs and fragments of viral genomic RNAs. The plant derived RNAs are thought to be tRNAs, based on their size and it is not known why they are encapsidated. They may contain sequences or secondary structures capable of initiating encapsidation or they may be encapsidated non-specifically due to their presence where encapsidation is occurring (Palukaitis et al. 1992). The nature of the virus-derived fragments is not known.

RNA 4A and 5 are distinct from the defective (D) RNAs (designated D RNA 3α and D RNA 3β, Graves and Roossinck 1995) and satellite RNAs known to be associated with CMV. The CMV D RNAs are derived from single, in-frame deletions within the 3a open reading frame of RNA 3. Although these D RNAs were derived from a subgroup I CMV strain, other virus strains from both subgroup I and II, as well as peanut stunt virus, were able to support replication of the DI RNA. Symptom production and standard virus yield are not affected by the presence of the D RNAs (Graves and Roossinck 1995). Satellite RNAs are small RNAs which cannot multiply without the
assistance of the virus with which they are normally associated. The virus upon which
the satellite is dependent is called the helper virus and no nucleic acid sequence
similarity exists between the virus and the satellite RNA (Francki 1985). The helper
virus does not require the satellite RNA for replication but satellites have been shown to
either enhance or ameliorate viral symptoms caused by CMV (Palukaitis 1988). A
number of satellite RNAs have been associated with various strains of CMV (Collmer
and Howell 1992; Roossinck et al. 1992) and many of these have been characterised
(Gould et al. 1978; Gordon and Symons 1983; Hidaka et al. 1984; Palukaitis and Zaitlin

1.2.2.6. 5' and 3' Terminal Structures on CMV RNAs

The four major encapsidated RNAs have been shown to contain a cap structure at their
5' termini (Symons 1975). This structure consists of a modified guanosine attached to
the penultimate base through a 5' p-p-p 5' link. These cap structures have been shown to
function in, but are not essential for, in vitro ribosome binding and translation
(Dasgupta et al. 1976). The 5' cap structure also increases the stability of the viral RNA
by protecting it from degradation by exonucleases (Matthews 1991).

Extensive sequence similarity exists at the 3' untranslated region (UTR) of the CMV
RNAs. Sequences in the 3' terminal half of this conserved region can be folded to form
a secondary structure similar to that of a tRNA molecule (Symons 1979). These tRNA-
like structures can be aminoacylated with tyrosine (Kohl and Hall 1974).

1.2.2.7. Promoter Regions and Viral Replicase Signals

The viral replicase recognition signals for the initiation of (-) strand synthesis are
thought to be contained in the 3' tRNA-like structure of CMV RNAs (Gargouri-Bouzid
1991). The length of the region required for efficient RNA accumulation in RNA 3 was
demonstrated to be 275 nt in Kin-CMV (Boccard and Baulcombe 1993). The sequence
similarity at the 3' end between CMV RNAs from different subgroups is only 60-65%,
however, almost identical secondary structures are formed (Palukaitis et al. 1992). Near
identical 3' RNA secondary structure is also observed between CMV and bromoviruses
despite low primary sequence similarity (Ahlquist 1981). In fact, when the 3' terminal
186 nt of Fny-CMV RNA 3 was incorporated into full length brome mosaic virus
(BMV) RNA 3, the BMV replicase was able to recognise the hybrid molecule allowing
normal accumulation (Rao and Grantham 1994). This indicates that secondary structure
rather than primary sequence may be important for replicase recognition.
Conserved sequences also occur in the 5' UTR of CMV RNAs 1 and 2 which are thought to be involved with replicase recognition. These sequences contain regions that are similar to the promoter regions (internal control regions ICR 1 and 2) of eukaryotic tRNA genes (Sharp et al. 1985, Marsh et al. 1989; Pogue and Hall 1992). Studies undertaken on BMV RNA indicate that the ICR2-like motif is required for (+) strand synthesis (Pogue et al. 1990 1992). The CMV motifs are contained in stem-loop structures which are also thought to be important in (+) strand replication (Pogue and Hall 1992).

RNA synthesised by a subgenomic promoter which allows transcription of RNA 4 to a level higher than any other CMV RNA. The nucleic acid sequence required for large-scale production of the subgenomic RNA 4 extends 70 nt upstream and 20 nt downstream from the RNA 4 initiation point on RNA 3 (Boccard and Baulcombe 1993). A putative subgenomic promoter element has been found in the region upstream of the RNA 4 initiation site and is an ICR2-like motif, similar to those found at the 5' ends of RNAs 1 and 2 (Boccard and Baulcombe 1993). However, a similar ICR2-like motif is also found in the intercistronic region of BMV RNA 3 which does not form part of the proposed RNA 4 subgenomic promoter (French and Ahlquist 1988; Pogue et al. 1992). An ICR-like motif has also been identified upstream of RNA 4A on CMV RNA 2 implicating this motif in the synthesis of RNA 4A (Ding et al. 1994).

Studies on the way RNA 5 is synthesised has included a search for the presence of putative subgenomic RNA promoters upstream from the start point of RNA 5 on the genomic RNAs. P. Palukaitis (personal communication) identified a number of nucleotide sequences on RNA 3 which may form part of a subgenomic promoter which could be involved with the synthesis of RNA 5. The first two elements have similarity with the ICR 2 motif and are shown aligned to the ICR 2-like motifs in RNA 3 from QC-MV and BMV in Fig. 1.3. The first element (see Fig. 1.3, element 1) is located at nucleotides 1466-1478 and the second element (see Fig. 1.3, element 2) is located at nucleotides 1777-1790 on RNA 3 (Fig. 1.4). A third element (Fig. 1.5) was identified in the region immediately upstream from the start point of RNA 5 which was similar in sequence to a region immediately upstream of the RNA 4 start point. It is not known whether these elements are capable of directing transcription of RNA 5 or if there are any other elements present which have this function. No putative RNA 5 subgenomic promoter elements could be found on RNA 1 or 2.
Chapter 1: Literature Review

ICR 2-like motif from BMV RNA 3      GGUUCAuuCCCuu
ICR 2-like motif from Q-CMV RNA 3      GGUUCAuuCCuUuu G
Element 1 upstream of RNA 5      GGUUCuA uUuu G
Element 2 upstream of RNA 5      GGUuAucu CuCgaaaG
Consensus sequence      GGUUCA uCCCUu G

Fig. 1.3: Alignment of putative RNA 5 subgenomic promoter elements 1 and 2 to ICR2-like motifs. Sequences are aligned to a consensus sequence shown on the bottom row. Conserved nucleotides appear in upper case.

Fig. 1.4: Location of putative subgenomic promoter elements on Q-CMV RNAs 3. The two ORFs on RNA 3 are indicated by horizontal boxes and the region corresponding to RNA 5 is enclosed in a vertical box. The position of putative subgenomic elements are indicated (*).

Fig. 1.5: Alignment of putative RNA 5 subgenomic promoter element 3 to sequences around the RNA 4 start site. Nucleotide positions with respect to RNA 3 are indicated. Nucleotides conserved in both sequences are shown in upper case. The start position of RNA 4 and 5 are indicated (*).

1.2.2.8. Replication

The replication cycle of CMV is not well characterised although it has been proposed that CMV replicates in a similar manner to BMV (Palukaitis et al. 1992). Three distinct phases are thought to occur in the replicative cycle of BMV (Dreher and Hall 1988c). The first step involves the synthesis of minus (-) strand copies of the genomic RNAs. The next step is the synthesis of daughter plus (+) strand genomic RNAs and subgenomic RNAs using the (-) strand copies as templates. To carry out the replication of CMV RNAs, the replicase must be able to recognise 5', 3' signals and subgenomic promoter regions. The same replicase complex must be capable of recognising signals and promoters with differing primary sequence and secondary structure. Three models have been proposed for the synthesis of subgenomic RNAs (Miller et al. 1985 and references therein). The first model proposed internal initiation by the replicase on (-) strand genomic RNAs. The second would involve premature termination during (-)
strand synthesis followed by independent replication of the subgenomic RNA. The third model proposed that the full genome-length RNA is processed by nuclease cleavage to yield subgenomic RNAs. In the case of BMV, there is strong evidence (Miller et al. 1985) that the first model is correct and that subgenomic RNAs arise through internal initiation on the (-) strand of the genomic RNA. The subgenomic promoter for BMV RNA 4 synthesis is well characterised (see Chapter 3, Fig. 3.1). The sequence required for RNA 4 synthesis in CMV has also been characterised (Boccard and Baulcombe 1993).

1.2.3. Other Viruses in the Cucumovirus Group

The genomic structure of TAV is very similar to CMV. It has a tripartite, positive-sense ss RNA genome. RNAs 1, 2, 3 and 4 have molecular weights of $1.26 \times 10^6$, $1.10 \times 10^6$, $0.9 \times 10^6$, and $0.43 \times 10^6$ daltons respectively (Canady 1995). TAV and CMV have indistinguishable particle morphology, sedimentation coefficients, RNA base ratios and buoyant densities (Habili and Francki 1974a). However, they differ in their nucleotide sequence and in the conditions required for their stability during purification (Habili and Francki 1974b 1974c). The identification of RNA 4A and 5 in TAV would provide further evidence that TAV and CMV are closely related.

1.3. CMV in Lupins

1.3.1. The Problem of CMV in Lupins

In Australia, one of the most economically-important hosts of CMV is lupins, in which CMV is the second most important virus disease after bean yellow mosaic virus (Jones and McLean 1989). CMV is a major threat to the economic viability of the lupin industry as infection of most cultivars with CMV may cause a loss in yield of between 50 and 100% (Jones 1987). CMV in lupins is difficult to control due to the way in which it is transmitted. The two most important mechanisms of infection in lupins are seed-borne infections and current season infections (Jones 1987). Seed-borne infection is the transmission of the virus from the parent plant to the daughter plant through the persistence of virus in the seed. Current season infection occurs when previously healthy lupins are infected with the virus via aphid transmission. The fact that CMV has many alternative hosts compounds the problem of controlling CMV as weed species located close to lupin crops may act as secondary reservoirs of infection. This was illustrated at a heavily infected lupin breeding site where fifteen species of weeds were found to be infected (Jones 1987).
1.3.2. Approaches to the Control of CMV in Lupins

1.3.2.1. Management Practices

Certain management practices have been developed in an attempt to control yield losses due to CMV infection. The first of these is the use of a seed certification system to ensure the percentage of contaminating CMV-infected seeds which are sown is low. Studies with various levels of infected seeds have indicated that the level of infected seed sown should be less than 0.5% to ensure no significant yield loss is caused by CMV (Bwyne et al. 1994). The presence of CMV in lupin seeds is detected using an assay which involves amplification of viral sequences using the polymerase chain reaction (Wylie et al. 1993).

High seeding rates also decrease the number of infected plants by formation of a canopy which shades out stunted infected plants. Earlier sowing allows an earlier canopy formation, and hence earlier shading, to increase the chances of killing infected plants which act as infection sources. However, this strategy is not effective when aphid populations arrive early (Bwyne et al. 1994). Another practice which may reduce the susceptibility of lupins to infection with CMV is inoculation with bacteria that cause root nodulation (Wahyuni and Randles 1993). This effect is partially reversed by the addition of nitrate which causes a reduction of nodulation. The mechanism of this action is not known but it is thought to be an example of systemic acquired resistance, (Matthews 1991) possibly mediated by a phytoalexin produced in response to tissue invasion by the root nodulating bacteria (Wahyuni and Randles 1993).

1.3.2.2. Breeding CMV Resistant Plants

Plant breeding programs to develop virus-resistant cultivars involves the transfer of plant-derived genes which confer viral resistance to otherwise susceptible host species. Plant-derived viral resistance genes that have been characterised include \( Tm-1 \), \( Tm-2 \) and \( Tm-22 \) in tomato, \( Rx \), and \( Ry \) in potato and \( N \) and \( N' \) in tobacco. However, the vast majority of plant-derived resistance genes remain uncharacterised. Some cultivars have been identified as being moderately resistant but none have been found to be highly resistant or immune to CMV (Jones and Cowling 1995).

1.3.2.3. Genetically Engineering Virus Resistant Plants

A recently-developed technique for the production of virus-resistant plants involves the expression in the plant of genes derived from the pathogen itself (pathogen-derived
resistance, PDR) or genes which have a direct effect on the pathogen upon entry to the host cell (Wilson 1993). The idea of genetically engineering virus-resistant plants has existed for some time (Sandford and Johnston 1985). The concept of using pathogen-derived genes evolved from the observation that prior infection with a mild, symptomless or attenuated strain of the virus protected crops from subsequent infection by severe strains of the virus (Wilson 1993).

Many of the plant virus genomes have been well characterised and several virus-derived genes have been successfully used to engineer resistance. The first example of pathogen-derived resistance was the use of the TMV coat protein gene to engineer virus-resistant plants (Powell-Abel et al. 1986). In the case of CMV, genes used to engineer resistance include the coat protein gene (Cuzzo et al. 1988; Quemada et al. 1991; Namba et al. 1991), modified RNA 2 sequence that encodes a truncated 2a protein (Anderson et al. 1992) and satellite RNAs (Harrison et al. 1987). Virus-derived sequences which have been successfully used to genetically engineer resistance against other viruses include the movement protein (Cooper et al. 1995; Beck et al. 1994), defective interfering RNAs (Kollar et al. 1993) and sequences derived from the 3' end of genomic RNAs (Zaccommor et al. 1993). Advantages and disadvantages of these strategies are discussed in Chapter 5.

1.4. Aim of Project

Compared to many other viruses, CMV is well characterised. However, additional research is required to obtain a full understanding of the origin, role and replication of individual components associated with this virus. CMV has been well characterised for two reasons: a) it has been used as a model system for the study of (+) strand ss RNA viruses, and b) to understand the structure and replication of this virus to design strategies, such as PDR, which may reduce the damage caused to economically-important crop species. CMV studies have generally focused on RNAs 1, 2, 3, and 4, consequently little is known about the associated RNAs, 4A and 5. Although, the origin and function of RNA 4A is now known, it has not been demonstrated that RNA 4A is a common feature of cucumoviruses. It is also not known how RNA 5 is generated or what function it plays in the life cycle of CMV. A number of strategies have been used to engineer transgenic plants which are resistant to CMV infection as other methods of reducing damage caused by CMV are in some ways ineffective. A strategy which used RNA 5 to genetically engineer virus resistance may provide an additional approach which could have advantages over some of the strategies currently used.

This dissertation describes the study of RNAs 4A and 5 found in cucumovirus strains,
and explores the potential use of RNA 5 in genetically engineered virus resistance. The genomic RNAs from which RNA 5 is derived were determined and the proposal that RNA 5 is generated as a subgenomic RNA was investigated. The presence of RNA 4A and 5 was surveyed in a number of cucumovirus strains to assess how widespread these RNAs are in cucumoviruses. Transgenic tobacco plants expressing RNA 5 were constructed to determine if genes expressing this RNA could confer resistance to CMV.

The results obtained have provided new information on the nature and distribution of RNA 4A and 5 and the biology of cucumoviruses. The presence of RNA 4A in all cucumovirus strains tested indicates that RNA 4A plays an important role in the life cycle of cucumoviruses or that it may have played a role in the evolution of cucumoviruses. Sequence analysis of purified RNA 5 and RNA 5 cDNA clones indicated that RNA 5 is derived from both RNAs 2 and 3. The presence of sequences on the genomic RNAs 2 and 3 indicated that RNA 5 may be synthesised in the same way as subgenomic RNAs although the possibility that RNA 5 is generated by specific cleavage of RNAs 2, 3, 4, and 4A can not be completely discounted. The function of RNA 5 is still unclear but it does not appear to include the production of a protein product. It is conceivable that RNA 5 may have a role in symptom modulation. RNA 5 was shown to be a common feature of CMV subgroup II strains and that the RNA 5 populations in all these strains appeared to consist of at least two sequence variants. The presence of RNA 5 in TAV indicated that RNA 5 is not limited to CMV. The absence of RNA 5 in a CMV subgroup I strain indicates that RNA 5 is not essential for the survival of all subgroup I CMV strains and that in relation to the ability to generate RNA 5, CMV subgroup II strains may be more similar to TAV than they are to CMV subgroup I strains. Transgenic tobacco plants expressing RNA 5 from a complex dimeric gene construct were susceptible to infection with CMV, indicating that this strategy may not be useful in engineering virus resistance. However, based on the knowledge that we now have of RNA 5, a new gene construct is proposed which may protect plants from damage by CMV.
Chapter 2: Materials and Methods

2.1. Propagation and Purification of Virus and Viral RNA

All virus strains were maintained in *Nicotiana glutinosa* and propagated in *Nicotiana clevelandii*. Plants were lightly dusted with Carborundum and mechanically inoculated by rubbing virus particles or RNA suspension on the leaves. Excess inoculum was washed off with tap water. Plants were maintained in an insect proof, air-cooled glasshouse. The origin of virus strains are described in references listed in Table 2.1.

Table 2.1: Virus isolates employed in these studies

<table>
<thead>
<tr>
<th>Virus</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q-CMV</td>
<td>Francki <em>et al.</em> 1966</td>
</tr>
<tr>
<td>L-CMV</td>
<td>Jones and McKirdy 1990</td>
</tr>
<tr>
<td>P-CMV</td>
<td>Tochihara and Tamura 1976</td>
</tr>
<tr>
<td>Pri-CMVa</td>
<td>Wahyuni <em>et al.</em> 1992</td>
</tr>
<tr>
<td>B14-CMV</td>
<td></td>
</tr>
<tr>
<td>Wc-CMV</td>
<td>Jones and McKirdy 1990</td>
</tr>
<tr>
<td>Ny-CMV</td>
<td>Anderson <em>et al.</em> 1995</td>
</tr>
<tr>
<td>V-TAVb</td>
<td>Habil and Francki 1974a</td>
</tr>
</tbody>
</table>

aPri-CMV was a gift from Ron Garrett (Agriculture Victoria, Melbourne).

bV-TAV was kindly provided by Fernando García-Arenal (Depatamento de Patología Vegital, Agrónomos, Ciudad Universitaria, Spain).

Q-CMV virus particles were purified from systemically-infected tissue essentially as previously described (Peden and Symons 1973). Leaf tissue was homogenised in a Waring blender in 2 volumes (w/v) of 0.5M sodium citrate buffer, pH 6.5, containing 0.5% thiolglycollic acid and 5mM EDTA. The extract was strained through a double layer of muslin and shaken briefly with an equal volume of chloroform. After centrifugation at 10,000 x g for 10 minutes, PEG was added to the clarified supernatant to 10% (w/v) and the mixture stirred for 45-60 minutes at 40C. Precipitated virus was pelleted by centrifugation at 10,000 x g for 10 minutes and suspended in re-suspension buffer (5 mM sodium borate, 0.5 M EDTA, pH 9.0) containing 1% Triton X-100 (50 mL/100 g leaf material) by stirring for at least 30 minutes. The suspension was clarified by centrifugation at 15,000 x g for 20 minutes and subjected to 2 cycles of high - low speed centrifugation at 50,000 x g for 75 minutes and 5,000 x g for 10 minutes, respectively. Virus pelleted by ultra centrifugation was suspended in 20 mM phosphate buffer, pH 7.4. Isolation of RNA from purified virus was performed by removing viral protein with sodium dodecyl sulphate (SDS) and phenol as described by Peden and Symons (1973).
Chapter 2: Materials and Methods

2.2. Plasmid DNA Sequencing

2.2.1. Manual Sequencing

Plasmid DNA was prepared by CsCl gradient purification (Sambrook et al. 1989) and the DNA sequence was determined by chain termination dideoxy sequencing using a TaqTrack sequencing kit (Promega) or a Sequenase sequencing kit (USB) according to the instructions provided by the manufacturer.

2.2.2. Automated Sequencing

Plasmid DNA was prepared by CsCl gradient purification (Sambrook et al. 1989) or by using the Wizard DNA purification system (Promega, Mezei 1991) according to the instructions supplied. DNA sequence was determined using dye primers with the Applied Biosystems DNA sequencing system according to the instructions provided by the manufacturer (USB).

2.3. Sequencing of RNA 5 cDNA Clones

20 RNA 5 cDNA clones constructed using RT PCR and RNA 5 purified by agarose gel electrophoresis (Blanchard 1991) and were either completely or partially sequenced. Only partial sequencing was required as only the regions specific to RNA 5.2 or RNA 5.3 were required to determine which genomic RNA the molecule was most similar to. Primers were designed to amplify RNA 5 cDNA by RT-PCR. The 5' primer contained the T7 RNA polymerase promoter sequence for the transcription of RNA 5 from cDNA clones. The 3' primer contained a Bam HI site for linearising cDNA clones before transcription. Double stranded cDNA was amplified in a polymerase chain reaction using the 3' and 5' primers mentioned above, gel purified and ligated to pCR1000 or end-filled and ligated to Sma I digested pUC18. Recombinant plasmids were transferred into E. coli strains INVα F' or DH5α.

The sequence of five RNA 5 cDNA clones (b37, a21, b69, a22, and b82) were determined on both strands using manual ds sequencing and another 15 clones (a8, b47, b32, b40, b74, b14, b31, b76, b55, b36, a18, b79, b20, b17, and a6) were completely or partially sequenced on one strand using automatic sequencing.

2.4. Cloning and Sequencing the 3' end of RNAs 1 and 2

The revised RNA 1 and 2 sequences determined by Ding et al. (1995b) were not
available at the initiation of this project. To determine the 3' end sequence of these RNAs, cDNA clones were generated and partially sequenced. Double stranded cDNA corresponding to the 3' half of the CMV genomic RNAs 1 and 2 were amplified using RT-PCR with the 3' primer QCMV3' (5'-CCGGATCCTGGTCTCTTATGGAGAACC-3') and the 5' primers Q1-5' (for RNA 1, 5'-CCACTTTCGTCACCTTTAACGAGGAG-3': nucleotides 1570 - 1595) and Q2-5' (for RNA 2, 5'-GCAGTAGCAGAACGGTTTCGTCTCTCG-3': nucleotides 1221 - 1247). Agarose gel purified ds cDNA was ligated to pCR1000 using T4 DNA ligase (Promega, Weiss et al. 1968) and transformed into *E coli* INVαF' competent cells as outlined in the TA cloning kit instruction manual (Invitrogen). Putative transformants were screened by restriction enzyme analysis of plasmid DNA isolated by the alkaline lysis method (Sambrook et al. 1989). The 3' ends of the cDNA inserts were sequenced using manual dideoxy sequencing (Taq Track, Promega) or automated sequencing (ABI).

2.5. Reverse Transcriptase and Polymerase Chain Reaction

First-strand cDNA was syntheses by extension of 3' end-complementary oligonucleotides using Moloney murine leukemia virus reverse transcriptase which has been genetically altered to remove the associated ribonuclease H activity (MMLV H-RT, superscript, BRL, Tanese and Goff 1988) according to the instructions supplied by the manufacturer. RNA templates had been previously treated with DNase I (Bresatec, Markey 1984) to remove traces of DNA. This cDNA was used as a template to produce second-strand cDNA in the first round of a polymerase chain reaction. The double-stranded cDNA was amplified in further rounds of the polymerase chain reaction using *Taq* DNA Polymerase (Perkin Elmer, Chien et al. 1976) as described in Sambrook et al. (1989). PCR conditions were as follows: 94°C for 5 min., 50°C for 2 min., 72°C for 3 min. (one cycle); 94°C for 1 min., 50°C for 2 min., 72°C for 3 min. (29 cycles); 94°C for 1 min., 50°C for 2 min., 72°C for 10 min..
2.6. Computer Analyses

Computer programs used in the analysis of RNA and DNA sequences are listed in Table 2.2.

Table 2.2: Computer programs used to analyse sequences

<table>
<thead>
<tr>
<th>Analysis performed</th>
<th>Program</th>
<th>Referencea</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sequence alignment</td>
<td>Pileup</td>
<td>GCG</td>
</tr>
<tr>
<td></td>
<td>Lineup</td>
<td>GCG</td>
</tr>
<tr>
<td></td>
<td>Ugly</td>
<td>GCG</td>
</tr>
<tr>
<td></td>
<td>Bestfit</td>
<td>GCG</td>
</tr>
<tr>
<td></td>
<td>Lalign</td>
<td>(Huang and Miller 1991)</td>
</tr>
<tr>
<td>Motif search</td>
<td>Fasta</td>
<td>(Altschul et al. 1990)</td>
</tr>
<tr>
<td></td>
<td>Blastn</td>
<td>(Altschul et al. 1990)</td>
</tr>
<tr>
<td>ORF analysis</td>
<td>Frames</td>
<td>GCG</td>
</tr>
<tr>
<td>RNA secondary structure analysis</td>
<td>Fold</td>
<td>GCG</td>
</tr>
<tr>
<td></td>
<td>Mfold</td>
<td>GCG</td>
</tr>
<tr>
<td></td>
<td>Squiggles</td>
<td>GCG</td>
</tr>
</tbody>
</table>


2.7. Purification of Q-CMV RNA 5

CMV RNA 5 isolated from virus particles was separated from larger RNAs using agarose gel electrophoresis. Individual RNAs were identified by ethidium bromide staining and RNA 5 was excised and purified with the RNaid kit (BIO 101, Vogestein and Gillepie 1979) according to the instructions provided.

2.8. Total RNA Extraction

Approximately 0.5 g of leaf tissue was frozen in liquid nitrogen and ground in a mortar with 0.6 mL of extraction buffer (0.1 M NaCl, 10 mM Tris-HCl pH 8.0, 1 mM EDTA, 1.0% SDS). The slurry was extracted with an equal volume of phenol:chloroform:isoamyl alcohol (25:24:1). After centrifugation, the nucleic acids were precipitated from the aqueous phase with ethanol and resuspended in sterile distilled water. The RNA was then precipitated by adding 0.25 volumes of 10 M lithium chloride (LiCl) and maintaining the solution on ice for at least 1 hour. Following centrifugation, the pellet was dissolved in sterile distilled water. The RNA was then re-precipitated with ethanol before being washed in 70% ethanol. The pellet
was finally dissolved in water and the RNA concentration determined by measuring absorbance at 260 nm.

2.9. In vitro RNA Transcription

CMV RNA 5 transcripts used in primer extension reactions were synthesised in vitro by run-off transcription using T7 RNA polymerase (Promega, Davanloo et al. 1984) according to the manufacturer's instructions. Briefly, RNA 5 cDNA clones were linearised at the 3' end of the RNA 5 cDNA insert with Bam HI. Reactions mixtures (20 µL) containing 40 mM Tris-HCl (pH 7.9), 6 mM MgCl2, 2 mM spermidine, 10 mM dithiothreitol, 15 mM each dNTP, 5000 units of T7 RNA polymerase, 40 units of Ribonuclease Inhibitor (RNasin, Promega, Polakowski et al. 1993) and 0.5 µg of DNA template were incubated for 1 hour at 37°C. After transcription, samples were treated with RNase-free DNase I (Bresatec, Markley 1984) for 15 minutes at 37°C and the reaction stopped by adding EDTA to a final concentration of 10 mM.

2.10. Primer Extension Reaction

All primers used in primer extension reactions were phosphorylated using polynucleotide kinase (Promega) and γ−32P-labelled ATP (370 MBq/mL; Amersham) as described by Sambrook et al. (1989). Reactions with total RNA isolated from virus-infected *N. clevelandii* contained approximately 2 µg of RNA, reactions with RNA transcripts from in vitro transcription reactions contained approximately 20 ng of RNA and those on encapsidated viral RNA preparations contained 2 µg of RNA. Fifty ng of labelled primer were incubated with the RNA template in reverse transcriptase buffer (50 mM Tris-HCl, pH 8.3, 75 mM KCl, 3 mM MgCl2, 50 mM dithiothreitol) at 70°C for 10 min followed by rapid cooling on ice. Annealed primer was then extended by adding 1mM each of all four dNTPs, 30U RNasin (Promega, Polakowski 1993) and 200U of MMLVH- reverse transcriptase (superscript, BRL, Tanese and Goff 1988) in 20 µL of reverse transcriptase buffer and incubated at 37°C for 30 min. Samples were immediately denatured by adding 5 µL of formamide loading buffer (95% formamide, 10 mM EDTA, 0.1% bromophenol blue, 0.1% xylene cyanol) and heating at 80°C for 5 min. Up to 5µL of the reaction mixtures were electrophoresed through polyacrylamide gels containing 7M urea and the extension products were visualised by autoradiography. Sequencing ladders to determine the size of the extension products were generated by manual double-stranded sequencing of various CMV cDNA clones.
2.11. Tobacco Transformation

Binary vectors were transferred from *Escherichia coli* to *Agrobacterium tumefaciens* (AGL1) by triparental mating (Lazo et al. 1991). The structure of the plasmids in *Agrobacterium* was verified by restriction enzyme analysis of plasmid DNA isolated using the method outlined by Dhaese et al. (1979). Tobacco leaf discs (cv. Wisconsin 39) were transformed using *Agrobacterium* by the procedure of Horsch et al. (1985) as modified by Higgins et al. (1988).

2.12. Phosphinothricin Acetyltransferase (PAT) Assay

To determine whether putatively transformed plants contained the gene encoding phosphinothricin acetyltransferase (PAT), the PAT activity of protein extracts was determined. Leaf tissue was ground in extraction buffer (100 mM Na₂PO₄ [pH 7.0], 20 mM NaCl, 1 mM phenyl methyl sulfonyl fluoride, 1 mg/mL BSA) in a ratio of 1:1.5 (w/v). After centrifugation, 16 µL of supernatant was added to 1 mL (0.02 µCi) of [1-14C]acetyl-CoA (50-60 µCi/mmol; Amersham) and the tubes were incubated at 37°C for 30 minutes. Tubes were centrifuged and 10 µL aliquots of the supernatant were spotted onto silica gel TLC plates (Merk). Ascending chromatography was carried out in a 3:2 (v/v) mixture of propan-1-ol and NH₄OH (25% NH₃). Plates were dipped in a solution containing 0.4% (w/v) diphenyl oxazole in 1-methylnaphthalene (α) and dried. 14C-labelled compounds were detected by fluororadiography (Schroeder et al. 1993).

2.13. Virus Challenge

For every viral inoculum level with each transgenic *N. tabacum* line, six replicates were challenged. The inoculum levels were either 0.5, 1 and 5 µg/mL (lines 1, 2, 5, and 6) or 0.2 and 1 µg/mL (lines 7, 14, 5, and 16). The first fully-expanded leaf as well as the next-oldest and next-youngest leaves were selected for inoculation and dusted with a thin layer of Carborundum. A 25 µL aliquot of virus inoculum (virus particles diluted in 20 mM phosphate buffer, pH 7.4) was added to each half of the selected leaves and rubbed five times. Excess inoculum was washed off with tap water. Each week until the experiment ended (at the commencement of flowering) plants were inspected for the presence of local lesions and symptoms of systemic infection and compared with uninoculated transgenic plants and inoculated non-transgenic plants. Leaf disk samples were taken at 1-week intervals to assess virus levels at various stages of infection using the enzyme linked immunosorbent assay (ELISA) described by Clark et al. (1977).
2.14. Subgroup Determination of L- and Pri-CMV.

To confirm that L-CMV belonged to CMV subgroup I and Pri-CMV belonged to subgroup II, the presence or absence of a diagnostic Eco RI site (Wylie et al. 1993; Anderson et al. 1995) on RNA 4 was determined. Two oligonucleotide primers, PCMV-Eco (5'-TATGATAAGAGCTTGTTCGCG-3') and PCMV5 (see Chapter 4), were used to amplify cDNA of a 527 nt region toward the 3' end of L- and Pri-CMV RNA 4 by RT-PCR. The presence of the diagnostic Eco RI site (indicative of a subgroup II strain) was indicated by restriction enzyme digestion of the 527-bp fragment into 342-bp and 185-bp fragments which were visualised by ethidium bromide after electrophoresis on a 1% agarose gel. The absence of cleavage with L-CMV PCR product indicated that L-CMV belonged to subgroup I. Complete cleavage of the Pri-CMV PCR product indicated that it belonged to subgroup II.
Chapter 3: Characterisation of Q-CMV RNA 5

3.1. Introduction

CMV is generally thought of as consisting of three genomic RNAs and a single subgenomic RNA derived from RNA 3 and which encodes the coat protein. However, there have been recent reports of additional molecules associated with Q-CMV which are not present in any other member of the Bromoviridae family. These include defective RNAs and additional 3' co-terminal RNAs. The function of CMV D RNAs is not known. The additional 3' co-terminal molecules, RNAs 4A and 5, appear to be exact copies of the 3' ends of one or more of the genomic RNAs.

RNA 4A was sequenced from seven cDNA clones and shown to be almost identical to the 3' terminal 682 nt of RNA 2 (Ding et al. 1994). The sequence variation observed in RNA 4A was thought to reflect sequence variation in the RNA 2 population. A similar variation was also observed between RNA 2 and the RNA 5 sequence determined by Blanchard et al. (1991) which may also be explained by the natural sequence variation within the RNA 2 population. The sequence of Q-CMV RNAs 1, 2 and 3 have all been revised since the original publications (Davies and Symons 1988; Ding et al. 1995b).

In a preliminary study of Q-CMV RNA 5, Blanchard (1991) showed that the sequence of the 5' end of RNA 5 was identical to the 5' end of a conserved region in all the 3' ends of the genomic RNAs of CMV. Furthermore, the 3' end of RNA 5 was shown to be 3' coterminal with the genomic RNAs of CMV by northern blot hybridisation analysis using a probe derived from the conserved 3' end common to all larger Q-CMV RNAs. This combined information was used to design primers for RT-PCR cloning of RNA 5. The sequence of one clone was determined and found to be the same as the 3' terminal 307 nt of RNA 2. Palukaitis et al. (1992) showed that RNA 5 was derived from the 3' end of RNA 3. Their conclusions were supported by the presence of a putative subgenomic promoter upstream of the region corresponding to RNA 5 on RNA 3 (P. Palukaitis, personal communication).

At least three different possibilities exist for the way 3' co-terminal molecules may be produced. The first would be specific degradation of the larger RNAs. This could possibly be due to the presence of a site that is highly susceptible to ribonuclease (possibly due to the secondary structure formed in this region) such as the one present at 161 nt from the 3' end of BMV RNAs (Dasgupta and Kaesberg 1977). Large amounts of this BMV 3' terminal RNA can be produced by ribonuclease cleavage at the susceptible site. These RNA molecules are resistant to further degradation. Secondly,
3' co-terminal molecules could be produced after premature termination of genomic RNA (-) strand synthesis followed by (+) strand synthesis of the truncated RNA molecule. This is similar to the way in which DI RNAs are produced in that the viral replicase becomes detached from the template, possibly due to secondary structure interference (Huang and Baltimore 1970). The difference with DI RNAs is that synthesis normally resumes at another position on the template as indicated by the presence of a mosaic of sequences generally found in DI RNAs. Finally, 3' co-terminal molecules could be produced by initiation of (+) RNA synthesis on full-length (-) RNA at an internal promoter. This is the mechanism by which the subgenomic RNA 4 of CMV and subgenomic RNAs of other viruses are synthesised (Matthew 1991). For instance, RNA molecules with a similar size to RNA 5 are thought to be synthesised in barley yellow dwarf virus (BYDV) in this manner (Kelly 1994). These BYDV molecules do not appear to encode proteins and by a process which is not understood, are thought to have a regulatory function in the virus life cycle.

The synthesis of subgenomic RNAs is an essential function of RNA viruses. The production of subgenomic RNAs allows ribosomes access to the AUG of internal and 3' proximal genes (Matthews 1991). This strategy also allows regulation of gene expression by controlling the amount of mRNA that is transcribed. Full-length (+) strand synthesis in BMV, CMV and cowpea chlorotic mottle virus (CCMV) requires conserved sequence motifs at the (+) strand 5' and 3' termini (Marsh and Hall 1987; Marsh et al. 1989). These motifs resemble internal promoters used for the transcription of eukaryotic tRNA genes by RNA polymerase III. The promoters are termed internal control regions (ICRs) 1 and 2 or, alternatively, the A and B boxes. Mutational analysis of these sequence motifs in BMV indicated that they are required for (+) strand synthesis (Pogue et al. 1990; Pogue et al. 1992) and form part of a stem-loop structure which is also critical for efficient synthesis (Pogue and Hall 1992). ICR-like sequences found upstream from CMV subgenomic RNA sequences on genomic RNAs are thought to play a role in the synthesis of CMV subgenomic RNAs (Boccard and Baulcombe 1993). The sequences required for subgenomic promoter activity consist of 70 nt upstream and 20 nt downstream of the RNA 4 start point. Deletion analysis indicates that progressive removal of upstream sequences gradually reduced promoter activity (Boccard and Baulcombe 1993). When the subgenomic promoters were duplicated, several subgenomic RNAs were produced with the smallest one being synthesised at higher concentrations. This indicates that it is possible to produce more than one subgenomic RNA from a single genomic RNA and that smaller subgenomic RNAs are preferentially synthesised. Further studies are required to fully understand the synthesis of subgenomic RNAs.
The subgenomic promoter for BMV RNA 4 has been characterised (Marsh et al. 1989) and shares conserved sequences with the subgenomic promoter regions of the alphaviruses (Ou et al. 1982; Levis et al. 1990) as well as those of other RNA plant viruses (Marsh et al. 1989). The similarities between the postulated BMV subgenomic promoter and the postulated alphavirus subgenomic promoter are shown in Fig. 3.1.

*BMV* | GUAUCUGCUAAUUCAGC | GUAAUUAUAUAUAAU | A-U tract
---|---|---|---
*alphavirus* | ACCUCUACGCCGCUCCUA | AUAGUUGCGUUA | consensus

Fig. 3.1: Sequence comparison of BMV and alphavirus subgenomic promoters. The core region and the 3' A-U tract (bold) is compared to the consensus of the alphavirus subgenomic promoter. The conserved sequence blocks between alphaviruses and BMV are underlined and the transcription start sites are indicated (*) (from Marsh et al. 1989).

The 3' non-translated region of (+) strand RNA plant viruses often contains secondary structures such as a tRNA-like configuration or a poly(A) stretch. In cucumoviruses, a tRNA-like structure exists at the 3' termini of all genomic RNAs (Joshi et al. 1983). Mapping studies on similar tRNA-like structures in BMV have indicated that it is involved with recognition by viral replicase to initiate (-) strand synthesis (Bujarski et al. 1985, 1986; Hall et al. 1987; Dreher and Hall 1988; Dreher et al. 1989). As 3' termini of CMV and BMV have very similar secondary structures, it is likely that the tRNA-like structures at the 3' end of CMV RNAs are also involved with replicase recognition (Boccard and Baulcombe 1993). This is supported by the fact that CMV replicase can replicate a BMV RNA 3 mutant which contains the 3' terminal tRNA-like structure derived from a CMV strain (Rao and Grantham 1994).

In addition to tRNA-like structures, some viruses have been shown to contain a pseudoknot region upstream of the tRNA-like structure (Duggal et al. 1994). Pseudoknots form when bases in the loop of a stem-loop structure are paired with those outside the stem-loop (Schimmel 1989). Potential roles of such pseudoknot structures in viral RNA may include a function in viral replication and virus spread (Takamatsu et al. 1990) or in the binding of host proteins (Leathers et al. 1993). Regions upstream from the tRNA-like structure on CMV RNAs are able to form stem-loop structures (Rezain et al. 1985). It is not known if these stem-loops form pseudoknot structures.

The extensive sequence similarity at the 3' termini of Q-CMV genomic RNAs was first recognised by Symons (1979). Comparisons of the three sequences indicated that RNA 1 and 2 are more similar to each other in this region than either are to RNA 3. More
recent analyses of the Q-CMV 3’ terminal conserved regions have shown this region to encompass 308 nt in RNA 1, 307 nt in RNA 2 and 304 nt in RNA 3 (Wood 1991; Ding et al. 1995b). The RNA 5 sequence determined by Blanchard (1991) is identical to the 307 nt conserved region at the 3’ end of RNA 2. Subgroup I CMV strains do not show such an extensive region of similarity at the 3’ terminus (Symons 1985). Only comparison of the terminal 189 nt of the TAV genomic RNAs have been made which show almost complete sequence similarity in this region (Wilson and Symons 1981). More extensive studies of sequence similarity within this region on CMV and between CMV and TAV is presented in this chapter and Chapter 4.

A number of methods have been developed to differentiate nucleic acid sequences which differ at single nucleotide positions (Hoebee et al. 1988; Sorrentino et al. 1991). Primer extension reactions have been used to detect nucleotide differences in DNA molecules by the substitution of mobility-shifting analogues for natural nucleotides during the extension process (Livak et al. 1992). Primer extension also has application in the characterisation of RNA molecules (Sambrook et al. 1989). The 5’ ends of RNA molecules can be precisely mapped by annealing a primer to the genomic RNA downstream from the 5’ end of the molecule. Synthesis of cDNA from this primer with reverse transcriptase results in the production of extension products which terminate at the 5’ end of the RNA molecules. By accurately determining the size of extension products, the distance between the 5’ end of the molecule and the primer annealing site can be determined. The sizes of primers used in the characterisation of RNA molecules by primer extension analysis generally range from 30 to 40 nt (Sambrook et al. 1989).

This chapter describes the characterisation of Q-CMV RNA 5. This is done by sequencing of RNA 5 cDNA clones constructed by Blanchard (1991) and comparison of these sequence with genomic sequences to determine from which genomic RNA RNA 5 is derived. This result is confirmed by direct analysis of RNA 5 molecules using a new method based on primer extension analysis which can differentiate between molecules with little sequence variation.

3.2. Results

3.2.1. Sequencing of 20 RNA 5 cDNA Clones

The sequence data derived from RNA 5 cDNA clones provides strong evidence as to from which genomic RNA(s) the RNA 5 molecules are derived. However, confirmation is required due to the possibility of cloning degraded genomic RNA which may co-migrate and be co-purified with RNA 5 molecules. The cDNA of the 3’ ends of these genomic RNAs may then be amplified during the polymerase chain reaction because the
The sequence of 20 RNA 5 cDNA clones was determined and aligned to the genomic RNA 3 sequence (Davies and Symons 1988) and the sequence of RNAs 1 and 2 determined by cloning and sequencing the 3' end of agarose gel purified RNA 1 and 2. The sequence of the 3' end RNA 1 cDNA clones were identical to the recently revised RNA 5 population (Davies et al. 1995a). Alignment of the sequence was carried out using the Pileup and Lineup programs in the GCG package and is shown in Fig. 3.2.

3.2.2. Computer Alignment of RNA 5 Sequences

The sequence of 20 RNA 5 cDNA clones was determined and aligned to the genomic RNA 3 sequence (Davies and Symons 1988) and the sequence of RNAs 1 and 2 determined by cloning and sequencing the 3' end of agarose gel purified RNA 1 and 2.

The sequence of the 3' end RNA 1 cDNA clones were identical to the recently revised RNA 1 and 2 sequence (Ding et al. 1995a). Alignment of the sequence was carried out using the Pileup and Lineup programs in the GCG package and is shown in Fig. 3.2.
Fig. 3.2: Alignment of RNA 5 and genomic RNA cDNA sequences. The 3’ end genomic RNA 1, 2, and 3 cDNA sequences are labelled r1, r2 and r3, respectively. Individual RNA 5 cDNA clones are indicated. A gap in the aligned sequence is denoted by (.) and N refers to nucleotide positions that could not be resolved. Positions of deletions and substitutions with respect to the genomic sequences are indicated above the aligned sequences (*).

The sequences aligned in Fig. 3.2 could be grouped into three classes based on the genomic RNA which they most closely resemble. The three types of molecules will be referred to as RNA 5.1 (most similar to the 3’ end of RNA 1), RNA 5.2 (most similar to the 3’ end of RNA 2) and RNA 5.3 (most similar to the 3’ end of RNA 3). The first group has one member (RNA 5.1: b37), which is 308 nt in length. The second group has 15 members (RNA 5.2: a21, b69, a8, b47, b32, b40, b74, b14, b31, b76, b55, b36 and a18) which are either 306 (b69) or 307 (all others) nt in length. The five members of the third group (RNA 5.3: a22, b79, b82, b20 and b17) are all 304 nt long.

To assess the sequence variability between the putative RNA 5 molecules, the consensus sequence from each RNA group was aligned (see Fig. 3.3) using the Pileup, Lineup and Pretty programs from the GCG package.
Chapter 3: Characterisation of Q-CMV RNA 5

Fig. 3.3: Alignment of consensus sequences from the three putative groups of RNA 5 sequence variants. Gaps (.) have been introduced to enable optimal alignment between the sequences. Non-conserved nucleotides appear in lower case.

The alignment of sequences in Fig. 3.3 illustrates the extensive sequence similarity between these molecules. Variation between the RNA 5 sequences reflected the sequence variation in the 3' termini of the Q-CMV genomic RNAs (Wood 1991).

Comparison of sequences in Fig. 3.3 indicates that the RNA 5.1 and RNA 5.2 sequences are more similar (97%) to each other than they are to the RNA 5.3 sequence. RNA 5.2 and RNA 5.3 were 89% similar while RNA 5.1 and RNA 5.3 were 90%.

3.2.3. Optimisation of Primer Extension Reactions as a Technique for Differentiating Similar RNA Molecules

To determine whether RNA 5 contains the three sequence variants (5.1, 5.2 and 5.3), a method was developed to analyse the RNA 5 molecules directly. Primer extension reactions were used to confirm the identity of the molecules found in the RNA 5 cDNA populations. To achieve this, primers specific for the three possible RNA 5 sequence variants were designed. These primers had to be long enough to bind to the RNA molecules and produce radioactive extension products giving a signal strong enough to be detectable in a reasonable time period (16 hours), but short enough to retain specificity for the individual sequences.

To assess the optimum size of an oligonucleotide primer required for differentiation of putative CMV RNA 5 molecules using primer extension analysis, RNA 5.1 and 5.2 were used as they were the most difficult molecules to differentiate due to the small variation between their sequences. Five different primers were designed to bind to RNA 5.1 or RNA 5.2 sequences (see Fig. 3.4). Primers specific for each RNA bound to the same site but differed in their length. Two of the primers were designed to bind RNA 5.2 (but not RNA 5.1) at a position where the RNA 5.2 sequences differed from RNA 5.1 by one nucleotide. This position of variation corresponded to the binding point of the 3' terminal nucleotide of both primers. The first primer was 18 nt long (P2-18: 5'-CGACTTCGATAGTGTGATT-3') and the second was 10 nt long (P2-10: 5'-GATAGTGTGATT-3'). Three primers of varying sizes were designed to bind to RNA 5.1 at a region where RNA 5.1 differs from RNA 5.2 by two nucleotides. One of the
positions of variation in the RNA 5.2 sequence corresponded to the binding point of the 3′ terminal nucleotide of the primers and the other position of variation corresponded with the binding point of the fourth nucleotide from the 3′ end of the primer. The three primers vary in their length and have the following sequences: P1-17, 5′-CCGTTTCGCAACCAATAC-3′; P1-10, 5′-CAACCAATAC-3′; P1-8, 5′-ACCAATAC-3′. The binding regions of all the primers are illustrated in Fig. 3.4.

Fig. 3.4: Primer-binding sites on aligned RNA 5.1 and RNA 5.2 sequences. The 5′ terminal 110 and 109 nt of RNA 5.1 (R5.1) and RNA 5.2 (R5.2), respectively, have been aligned using the Bestfit program from the GCG package. Identical nucleotides are linked with a vertical line. A gap (.) has been introduced into the RNA 5.2 sequence to allow optimal alignment. The regions corresponding to the binding sites of the five primers are indicated below the sequence.

All five primers were used in primer extension reactions to determine the optimum primer size required to achieve specificity while still retaining a signal that would be detectable. The predicted sizes of the extension products produced if the primers were able to bind to the RNA templates are shown in Table 3.1.

Table 3.1: Sizes of putative primer extension products using different primers on RNA 5.1 and 5.2.

<table>
<thead>
<tr>
<th>Primer a</th>
<th>RNA 5.1 b</th>
<th>RNA 5.2 c</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1-17</td>
<td>102 nt</td>
<td>101 nt</td>
</tr>
<tr>
<td>P1-10</td>
<td>95 nt</td>
<td>94 nt</td>
</tr>
<tr>
<td>P1-8</td>
<td>93 nt</td>
<td>92 nt</td>
</tr>
<tr>
<td>P2-18</td>
<td>76 nt</td>
<td>75 nt</td>
</tr>
<tr>
<td>P2-10</td>
<td>68 nt</td>
<td>67 nt</td>
</tr>
</tbody>
</table>

a The name of the primer indicates the RNA for which it was designed (P1 is specific for RNA 5.1 and P2 is specific for RNA 5.2) and the length of the primer in nt (last number).
b Predicted size of putative extension products when RNA 5.1 is successfully used as the template
c Predicted size of putative extension products when RNA 5.2 is successfully used as the template
The difference in the size of the putative extension products produced with the two different templates is due to a one base deletion in the RNA 5.2 sequence with respect to RNA 5.1 (position 52, Fig. 3.4). To accurately determine the size of primer extension products, they were compared to dideoxy sequencing products from an RNA 5.1 cDNA clone (b37) using the P1-17 primer. The size of P1-17/RNA 5.1 extension product should correspond to the residue complementary to the first nucleotide of the RNA 5.1 sequence. The size of the other extension products were determined by comparison with bands lower down in the sequencing ladder.

RNA was transcribed from the RNA 5.1 (b37) and RNA 5.2 (a21) cDNA clones using T7 RNA polymerase from the T7 promoter in the 5’ end primer. Equal amounts (approximately 20 ng) of transcribed RNA were used as templates with the five different primers labelled at the 5’ end with $\gamma^{32}$P-ATP. The ability of the primers to initiate cDNA synthesis on the RNA templates was assessed by observing the autoradiographs for the presence of bands corresponding to extension products of the predicted size. The autoradiograph of the gel containing the primer extension reaction is shown in Fig 5.

![Fig. 3.5: Primer extension analysis of RNA 5.1 and 5.2 using different primers. The primer extension reactions were electrophoresed on 6% polyacrylamide gels which were subsequently dried and autoradiographed for 16 hours. Lanes A, C, G and T are products of dideoxy sequencing of the RNA 5.1 cDNA clone b37 with the P1-17 primer. The C residue corresponding to the first nucleotide of the RNA 5.1 sequence is marked with an arrow. Lanes corresponding to the primer extension reaction are labelled with their respective primers. Lanes labelled 1 or 2 correspond to primer extension reactions on RNA 5.1 or RNA 5.2 transcripts, respectively.](image)
Extension products were evident with both templates when the 17- and 18-nt primers were used. The 10-nt primers produced extension products only with the template to which they were designed to bind. The 8-nt primer also produced an extension product only with the template to which it was designed to bind. However, this band was barely visible after 16 hr exposure.

### 3.2.4. Primer Extension Analysis of Q-CMV RNA 5

The primers P1-10 and P2-10 were used to detect the presence of RNA 5.1 and 5.2 in purified Q-CMV RNA5 as they were the optimal length to retain specificity for the template while still producing a readily detectable product. A third primer was designed to detect the presence of RNA 5.3 (P3-20, 5’-GGTTCCCCGCGCAAACACAG-3’). This primer was longer than the first two primers as the sequence variation between the RNA 5.3 and the other RNA 5 sequences is greater than that which exists between RNA 5.1 and RNA 5.2 (see Fig. 3.3). The binding sites of all three primers are shown in Fig. 3.6.

Fig. 3.6: Binding regions of RNA 5-specific primers. The consensus sequences of the three putative RNA 5 sequence variants have been aligned using the Pileup and Pretty programs from the GCG package. Gaps (.) have been introduced to allow optimal alignment of the sequences. Nucleotides which are not conserved in at least two of the sequences are shown in lower case. The binding sites of the primers are shown below the sequence (-----) and on the sequences (underlined).

The alignment of primers to the putative RNA 5 sequences in Fig. 3.6 shows P1-10 has 2 mismatches with RNA 5.2 (including the 3’ terminal nucleotide of the primer) and 5 mismatches with the RNA 5.3 sequence (including a mismatch with the 3’ terminal nucleotide of the primer and an insertion in the RNA 5.3 sequence). P2-10 has one mismatch with RNA 5.1 sequence (the 3’ terminal nucleotide of the primer) and three mismatches with the RNA 5.3 sequence. P3-20 has 10 mismatches with both RNA 5.1 and RNA 5.2 (which includes the four nucleotides that are 3’ terminal of the primer).

Primer extension reactions with each primer were performed on RNA transcribed from
RNA 5.1 cDNA clone b37 (Fig. 3.7-lanes 1), RNA 5.2 cDNA clone a21 (Fig. 3.7-lanes 2) and RNA 5.3 cDNA clone a22 (Fig. 3.7-lanes 3). All three primers were shown to be specific for the templates for which they were designed (Fig. 3.7). The primers were also used to detect the presence of the three putative RNA 5 molecules in purified RNA 5 (Fig. 3.7, lanes 5). Primer extension reactions on purified RNA 5 with P1-10 resulted in no extension product of the predicted size, even after long exposure times (1 week). P2-10 resulted in the production of an extension product approximately the predicted size, and primer extension with P3-20 also resulted in the production of an extension product of the approximate predicted size. This indicates the RNA 5 population contains products corresponding to the RNA 5.2 and 5.3 but not the RNA 5.1 products or that RNA 5.1 levels are below the detection threshold of this technique. The sizes of the extension products were determined by comparison with the products of dideoxy sequencing of RNA 5 cDNA clones. Very faint products with a molecular weight lower and higher than that of the expected extension product was observed in the lanes corresponding to the primer extension reaction with P1-10 and P2-10, respectively and RNA 5. The most likely explanation for the presence of these bands is that they are produced by binding to cryptic binding sites on the RNA templates. The small size of the primers increased the possibility of primers binding to such sites. The extension products in the lanes containing reactions with purified RNA 5 were approximately one nucleotide larger than those in the corresponding reactions with transcribed RNA. This may be due to the presence of a cap or some other chemical modification at the 5' end of the viral molecule which causes the addition of a non-templated nucleotide to the extension product during primer extension.

Fig. 3.7: Primer extension analysis of QCMV RNA 5. Reactions were electrophoresed on a 6% polyacrylamide gel which was subsequently dried and autoradiographed. Lanes
corresponding to the various reactions are labelled with their respective primers. Primer extension reactions with transcribed RNA 5.1, RNA 5.2 and RNA 5.3 are labelled 1, 2 and 3, respectively. Primer extension reactions on purified RNA 5 are labelled 5. Lanes corresponding to dideoxy sequencing products are labelled A, C, G, and T. The nucleotide corresponding to the 5’ terminal nucleotide of RNA 5.2 and RNA 5.3 are indicated with an arrow. The RNA 5.2 cDNA clone a21 was used as the template for sequencing with the P2-10 primer and the RNA 5.3 cDNA clone a22 was used as the template with the P3-20 primer.

Sequence analysis of the RNA 5 cDNA clones and primer extension analysis indicates RNA 5 is derived from RNA 2 and 3 but not RNA 1. With this knowledge, the sequence of RNA 5 and the sequence surrounding the start point of RNA 5 on the genomic RNAs can be analysed to determine how RNA 5 is produced and if it has a function.

3.2.5. Is RNA 5 Derived from a Subgenomic Promoter on RNAs 2 and 3?

To investigate the possibility that RNA 5 is derived as a subgenomic RNA, the regions surrounding the initiation point of RNA 5 on the genomic RNAs 2 and 3 were searched for the presence of sequences similar to previously characterised subgenomic promoters. Using the local alignment program Lalign (Huang and Miller 1991), the following sequence were used to search for putative promoter sequences around the start site of RNA 5.2 and RNA 5.3 on the genomic RNAs; postulated promoter sequences from BMV (Marsh et al. 1988), the consensus sequence of the postulated alphavirus subgenomic promoter (Ou et al. 1982) and the tRNA ICR consensus sequence (Sharp et al. 1985). The match with the highest similarity was with a region around the start of RNA 5.3 on RNA 3. When these sequences were aligned with the alphavirus subgenomic promoter consensus sequence with the Gap program from the GCG package (Fig. 3.8, C) a value of 64% was calculated for the percent similarity.

When the alphavirus subgenomic promoter sequence is aligned to the same region surrounding the start point of RNA 5.1 and 5.2 on RNAs 1 and 2, the sequence similarity is lower than that seen on RNA 3 (see Fig. 3.8, A and B). The percent similarity for the RNA 1 sequence is 59% and the similarity for the RNA 2 sequence is 50%. No other regions of high sequence similarity between CMV sequences and the alphavirus subgenomic promoter sequence were found.
Fig. 3.8: Alignment of the postulated alphavirus subgenomic promoter consensus sequence (alpha) with Q-CMV RNA 1 (A, R1), 2 (B, R2) and 3 (C, R3) sequence. Vertical lines indicate matches in the sequence. The start site of the alphavirus subgenomic RNA and RNA 5 are indicated (*).

To investigate the possibility that a putative subgenomic promoter may involve RNA secondary structure, the regions 100 nt upstream from the 3’ conserved region on RNAs 1, 2 and 3 were analysed for the presence of common secondary structures. This region was folded into a number of suboptimal secondary structures (data not presented) using the GCG program Mfold and displayed using the GCG program Squiggles. No similarity between the secondary structures was observed in the regions 100 nt upstream from the conserved regions corresponding to the putative initiation points of RNA 5.

### 3.2.6. Does RNA 5 Encode a Protein?

To investigate the possibility that RNA 5 may encode a protein, a search for open reading frames (ORFs) on the RNA 5 cDNA clones sequenced on both strands was performed using the Frames program from GCG. The consensus sequence of RNA 5.2 contained three major open reading frames on the (+) strand in the three different reading frames. The size of these ORFs in the cDNA clones a21 and b69 were calculated (Table 3.2). The biggest ORF present in RNA 5.3 cDNA clone encoded a polypeptide of 19 aa. No ORFs larger than 5 aa were found on RNA 5.3 reading frame 2 or 3. A summary of these results is presented in Table 3.2. The size of the three corresponding ORFs in the two RNA 5.2 clones are different due to a nucleotide deletion in clone b69. None of the RNA 5.2 ORFs are the same as the RNA 5.3 ORF.
Table 3.2: ORFs present in RNA 5.2 (clones a21 and b69) and RNA 5.3 (clone a22).

<table>
<thead>
<tr>
<th>cDNA clone</th>
<th>Reading Frame&lt;sup&gt;a&lt;/sup&gt;</th>
<th>ORF position&lt;sup&gt;b&lt;/sup&gt;</th>
<th>ORF size&lt;sup&gt;c&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>a21 (RNA 5.2)</td>
<td>1</td>
<td>127-186</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>137-199</td>
<td>21</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>150-248</td>
<td>33</td>
</tr>
<tr>
<td>b69 (RNA 5.2)</td>
<td>1</td>
<td>127-198</td>
<td>24</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>137-247</td>
<td>37</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>150-185</td>
<td>12</td>
</tr>
<tr>
<td>a22 (RNA 5.3)</td>
<td>1</td>
<td>127-183</td>
<td>19</td>
</tr>
</tbody>
</table>

<sup>a</sup> Reading frame number  
<sup>b</sup> Position in the nucleotide sequence of the ORF  
<sup>c</sup> Size of the open reading frames (number of amino acids)

3.3. Discussion

Conflicting reports on the nature of the RNA 5 population has led to confusion as to what genomic RNA RNA 5 is derived from. This has made it difficult to determine how RNA 5 is produced. Characterisation of Q-CMV RNA 5 in this chapter has resolved this conflict in results by demonstrating that RNA 5 is derived from both RNA 2 and 3. This has resulted in further studies which suggest the most likely method of RNA synthesis is via transcription by the viral replicase directed by a subgenomic promoter on the RNAs 2 and 3.

The RNA 5 sequences determined here matched the revised Q-CMV genomic RNA sequences (Davies and Symons 1988; Ding et al. 1995b) and the sequences of the genomic cDNA clones constructed in this project. The variation between RNA 5 sequences and the original genomic sequences could be related to errors in the original sequences (possibly made by enzymes such as taq polymerase or reverse transcriptase which have no proof reading function, Krawczak et al. 1989). An alternative explanation would be that the variation between the sequences is due to natural sequence drift which is enhanced by the absence of a proof reading function in the RNA-dependent RNA polymerase of RNA viruses (Eigen and Biebricher 1988).

There are numerous points of sequence variation among the various RNA 5.2 cDNA clones. This variation probably reflects the natural variation of the molecules from which they are derived. A similar variation was present in the RNA 4A population (Ding et al. 1994). The high sequence variation in these populations indicate absolute sequence conservation is not required for the functioning of these molecules (if, in fact, they have a function at all). One of the most notable variations in the RNA 5.2 sequence is the deletion found in the RNA 5.2 cDNA clone b69. This deletion truncates
the three largest ORFs found in RNA 5.2 and indicates that RNA 5.2 is probably not translated. The variations seen in the RNA 5 cDNA clones are summarised in Table 3.3 below.

Table 3.3: Summary of RNA 5 sequence variation

<table>
<thead>
<tr>
<th>cDNA clone&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Nucleotide&lt;sup&gt;b&lt;/sup&gt; position</th>
<th>Change&lt;sup&gt;c&lt;/sup&gt;</th>
<th>Nucleotide&lt;sup&gt;d&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>b37 (5.1)</td>
<td>292</td>
<td>Substitution</td>
<td>T to G</td>
</tr>
<tr>
<td>a21 (5.2)</td>
<td>193</td>
<td>Substitution</td>
<td>A to T</td>
</tr>
<tr>
<td></td>
<td>283</td>
<td>Substitution</td>
<td>C to A</td>
</tr>
<tr>
<td>b69 (5.2)</td>
<td>184</td>
<td>Deletion</td>
<td>C</td>
</tr>
<tr>
<td></td>
<td>292</td>
<td>Substitution</td>
<td>T to G</td>
</tr>
<tr>
<td>b40 (5.2)</td>
<td>193</td>
<td>Substitution</td>
<td>A to T</td>
</tr>
<tr>
<td>b14 (5.2)</td>
<td>17</td>
<td>Substitution</td>
<td>C to A</td>
</tr>
<tr>
<td></td>
<td>161</td>
<td>Substitution</td>
<td>C to T</td>
</tr>
<tr>
<td></td>
<td>163</td>
<td>Substitution</td>
<td>C to T</td>
</tr>
<tr>
<td></td>
<td>168</td>
<td>Substitution</td>
<td>G to A</td>
</tr>
<tr>
<td>b36 (5.2)</td>
<td>194</td>
<td>Substitution</td>
<td>T to C</td>
</tr>
<tr>
<td></td>
<td>283</td>
<td>Substitution</td>
<td>C to T</td>
</tr>
<tr>
<td>a6 (5.3)</td>
<td>187</td>
<td>Substitution</td>
<td>G to T</td>
</tr>
</tbody>
</table>

<sup>a</sup> Name and type of RNA 5 cDNA clone.
<sup>b</sup> Position with respect to the aligned sequences
<sup>c</sup> Nature of variation
<sup>d</sup> Nucleotide changes

No variations were found at internal positions of stems in the predicted CMV 3’ terminal, tRNA-like, pseudoknot secondary structures (Joshi et al.1983). Five of the eleven variations (a21-283; b14-161, 163, 168; b36-283) observed in the RNA 5.2 clones were identical to the nucleotide variations seen in the corresponding positions of the RNA 1 sequence (Table 3.3). This indicates that the 3’ ends of RNA 1 and 2 are more closely related than previous studies have shown (Symons 1979). Sequence variants of RNA 1 and 2 which contain identical ends may exist or may have been produced during the evolution of CMV. The presence of sequence variants is thought to be a normal occurrence in viral RNA populations and is important for the evolution of the virus (Eigen and Biebricher 1988). The lower amount of sequence variation in the RNA 5.3 clones may reflect a different functional role for the region encompassed by RNA 5 sequence in RNA 3, however a larger number of sequences need to be analysed to confirm this lower variation. Mutational analysis has shown that the 3’ terminal 275 nt are essential for efficient accumulation of RNA 3 (Boccard and Baulcombe 1993) which may explain the almost 100% similarity of the RNA 5.3 clones in this region. The remaining 29nt upstream is conserved between the CMV RNAs which may mean that it may have an essential role in the life cycle of CMV. However, this role does not
appear to be involved with RNA accumulation as deletion of this region has no effect on RNA accumulation. However, if this region had no function at all the sequence would not be expected to be conserved at all.

The optimisation of primer extension reactions for the differentiation of viral RNA molecules with high sequence similarity was achieved by testing a variety of primer sizes for their ability to initiate transcription in primer extension reactions. The presence of extension products in all tracks corresponding to primer extension reactions using primers which are 18 and 17 nt in length (see Fig. 3.5) indicates that these primers are too long to retain binding specificity as they are able to initiate transcription with one or two nucleotide mismatches, respectively. The 10 nt primers were only capable of initiating synthesis on templates to which they were designed to bind. This indicates that a length of 10 nt is a suitable size to retain specificity during primer extension when templates differ by only one or two nucleotides. The 8 nt primer also showed binding specificity to the template to which it was designed to bind, although the exposure time used in these experiments did not produce a sufficiently strong signal during autoradiography. This indicates that an optimum primer size for differentiating these viral RNA sequences is 10 nucleotides.

A disadvantage of reducing the primer size was the increased probability that the primer would bind to other regions on the template. This was evident with the P2-10 primer as a lower molecular weight extension product was often produced by binding to a region with some sequence complementarity closer to the 5' end of the RNA 5 molecule (see Fig. 3.7).

The use of primers 10 nt in size was effective in differentiating RNA 5.1 and RNA 5.2. A primer of this size was not required for the differentiation of RNA 5.3 as there was sufficient sequence variation to enable differentiation with a larger primer. As the primer extension reactions on all RNA transcribed from cDNA clones only produced extension products of the predicted sizes with the templates to which they were designed to bind, the primers were shown to be specific for those templates. When these specific primers were used in primer extension reactions with purified RNA 5, the presence of the different types of RNA 5 molecules were detected. Distinct bands corresponding to RNA 5.2 and RNA 5.3 were visible in lanes corresponding to purified RNA 5 indicating the presence of these RNAs in the natural RNA 5 population. The absence of an extension product with P1-10 on purified RNA 5 indicated that RNA 5.1 may not exist in the natural RNA 5 population. This result was reproducible even with increased amounts of purified RNA 5 (data not shown). The most likely explanation for the occurrence of the RNA 5.1 cDNA clones is as a result of the accidental cDNA
cloning of degraded genomic RNA 1. This is supported by the fact that only 1 out of the 20 cDNA clones sequenced was identical to RNA 1.

Primers used to differentiate between RNA 5.1 and 5.2 were not suitable for primer extension reactions on total RNA prepared from virus-infected plants. Efficient primer extension was not possible with primers less than 18 nt long on such preparations. This was probably due to the low concentration of viral RNA in total RNA preparations. The inability of the smaller primers to bind viral RNA from total RNA preparations indicates the optimum size of primers would have to be determined again if RNA molecules from total RNA preparations with high similarity were to be differentiated.

There are a number of possible ways in which RNA 5 could be produced. The first is that RNA 5 could be a specific degradation product of the larger RNAs. Similar molecules have been produced by specific cleavage at a point near the 3' end of BMV RNAs to form molecules which are resistant to further degradation due to stabilisation by the formation of extensive secondary structure (Dasgupta and Kaesberg 1977). If RNA 5 was produced by specific cleavage, an RNA 1 3' fragment should have been produced as RNA 5.1 has the same 5' sequence as RNA 5.2 and RNA 5.3. Also, if RNA 5 was produced by specific cleavage 5' fragments should also be present. It is improbable that the 5' fragment would be completely degraded as this would indicate no stabilising secondary structure is present upstream of the RNA 5 region. This is not likely to be the case as secondary structure has been proposed for the 5' end of CMV RNAs (Rezaei et al. 1985) and it is likely that CMV RNAs contain secondary structures to inhibit binding of (+) and (-) strands (Eigen and Biebricher 1988). No such 5' fragments were detected by Gordon and Symons (1985).

A second explanation for the production of RNA 5 is by premature termination of (-) strand synthesis followed by (+) strand synthesis. The premature termination may be caused by the presence of secondary structures upstream from the RNA 5 start sites. This method of RNA synthesis would require the 5' end of RNA 5 (-) strand to act as a promoter for (+) strand synthesis. It was once thought that subgenomic RNAs were produced in this way but it is likely that they are produced as outlined below.

Finally, RNA 5 molecules may be synthesised in the same way as subgenomic RNAs. This involves initiation of (+) strand synthesis from an internal promoter on the (-) strand of the RNA from which it is derived. This idea is supported by the sequence similarity to the alphavirus subgenomic promoter consensus and the regions around the RNA 5 initiation points on the larger RNAs. The regions around the start of RNA 5.1, 5.2 and 5.3 all display some sequence similarity with the alphavirus sequence.
However, if this sequence is acting as a subgenomic RNA promoter it is surprising that RNA 1 contains this sequence as RNA 5.1 does not seem to form part of the RNA 5 population. Also, similarity may be expected between the region upstream of the RNA 5.2 or 5.3 start site and the regions upstream of either RNA 4 of RNA 4A on the genomic RNAs due to the presence of common subgenomic promoter sequences. This absence of sequence similarity may indicate that the two types of RNA 5 molecules may be synthesised from different promoters or may be produced in two different ways.

The third possible way in which RNA 5 is synthesis (i.e., the synthesis of RNA 5 is directed by a subgenomic promoter on RNAs 2 and 3) seems to be the most likely explanation for the origin of RNA 5. This is based on the similarity of regions upstream and around the start point of RNA 5 on genomic RNAs with other subgenomic promoter element and because it is unlikely that RNA 5 is produced by specific degradation due to the absence of a 5' cleavage fragment and the absence of a product corresponding to RNA 5 from RNA 1.

It is not known why RNA 5 molecules only appear to be produced from RNAs 2 and 3 and not from RNA 1. Both RNA 2 and 3 act as templates for subgenomic RNAs, however, no subgenomic RNAs have been reported from RNA 1. The absence of any reported subgenomic RNAs being synthesised from RNA 1 may indicate that RNA 1 is not capable of acting as a template for subgenomic RNA synthesis. Alternatively, RNA 5 molecules themselves may have a requirement to be synthesised from other subgenomic RNAs, i.e., RNA 5.2 may be derived from RNA 4A and RNA 5.3 may be derived from RNA 4. Therefore, if RNA 1 does not act as a template for other subgenomic RNAs, RNA 5.1 may not be synthesised due to the absence of a subgenomic template.

The function of RNA 5 remains unknown. RNA 5 was previously thought to occur only in subgroup II strains of CMV which were considered to produce less severe symptoms than subgroup I strains. This indicated that RNA 5 may have a role in the attenuation of symptoms. Results outlined in Chapter 4 indicate RNA 5 is not limited to subgroup II strains, however, there still may be a correlation between the presence of RNA 5 and a reduction in the severity of symptoms. Symptom attenuation may be facilitated in two ways. Firstly, the RNA molecules themselves may directly attenuate symptoms or secondly, the RNA molecules may encode a protein which may have a symptom attenuation role. CMV may produce molecules to attenuate severity of infection in order to induce chronic infections. This would allow time for the virus to be transmitted to a new host as the infected host would still be able to survive and possibly reach maturity which could facilitate seed transmission to the next generation.
There are a number of possible ways 3' coterminal molecules could attenuate symptoms. As the characterisation of RNA 5 has shown, it contains all the sequences required for recognition by the viral replicase. This would allow the RNA 5 molecules to compete with the genomic RNAs for replication factors thereby reducing virus replication and attenuating symptoms. This process is similar to that proposed for symptom attenuation by DI particles (Holland 1990). The RNA 5 molecules may also inhibit RNA synthesis via co-suppression. This phenomenon is often observed after over expression of transgenes and is thought to function by an RNA-dependent RNA polymerase copying small segments of RNA molecules that accumulate to unacceptably high levels, resulting in the production of a double stranded complex which is targeted for degradation (Lindbo et al. 1993). Such a mechanism may result in the complete absence of infection (Smith et al. 1994). Another way that RNA 5 molecules could act to inhibit viral RNA synthesis is to bind to the 5' ends of newly-synthesised (-) strand genomic RNAs to reduce access to promoters by the replicase.

The function of RNA 5 may be to encode a protein, however this possibility seems unlikely. This was investigated by calculating the size of putative ORFs. The largest ORF in RNA 5.3 is very small and although it is possible that RNA 5.3 encodes a functional protein or peptide, it is unlikely. The variation in the ORFs in the different RNA 5.2 cDNA clones also indicates that RNA 5.2 is unlikely to encode a protein. However, a mutation in the RNA 5 sequence may affect the function of RNA 5 but not affect the viability of the whole virus therefore such a variant may survive natural selection and form part of the CMV RNA population. The only example of a 3' coterminal molecule found in association with a plant virus is a small subgenomic RNA from BYDV (Kelly 1994). It also does not encode a protein and RNA 5 in CMV may have a role similar to this RNA which is thought to be synthesised to perform some regulatory role rather than act as mRNA.

The sequencing of a number of RNA 5 cDNA clones and the direct analysis of purified RNA 5 using primer extension analysis has shown what genomic RNAs RNA 5 is derived from, but the way in which it is derived and what function it plays is still unclear. The most likely way in which RNA 5 is produced is as a subgenomic RNA but it does not appear to encode a protein like most subgenomic RNAs. A possible function is the attenuation of symptoms to maintain chronic infection. A survey of other cucumovirus strain will provide further evidence that RNA 5 is a subgenomic RNA which may be responsible for symptom attenuation.
Chapter 4: Survey for RNAs 4A and 5 in V-TAV and a Number of CMV Subgroup I and II strains

4.1. Introduction

The genomic RNAs 1, 2 and 3 and the subgenomic RNA 4 are all essential to the life cycle of CMV and are always present in virus preparations. However, it is not known if the RNA 2-derived RNA 4A and the RNA 2- and 3-derived RNA 5 are always present. RNA 4A is identical in sequence to the 3’ terminal 682 nt of RNA 2. An ORF which overlaps the 2a gene and encodes a small protein (2b) has been shown to be expressed \textit{in vivo} (Ding et al. 1994). The 2b protein is thought to have a host-specific long-distance movement function (Ding et al. 1995c). Alignment of sequences surrounding the first nucleotide of RNA 4A with RNA 2 sequences from various cucumoviruses showed the presence of a highly conserved region which is thought to form part of a putative subgenomic promoter. In addition to this conserved nucleotide sequence, alignment of the 2b gene ORF shows that it is conserved among all cucumoviruses but is not found in other Bromoviruses. Despite the presence of the conserved nucleotide sequences and ORF, to date RNA 4A molecules have been found in other cucumoviruses. RNA 4A is readily detected in encapsidated preparations of subgroup II CMV RNAs stained with ethidium bromide whereas RNA 4A could not be detected in encapsidated subgroup I CMV RNA preparations (Anderson and Blanchard, unpublished data). Studies to determine whether other cucumovirus strains contain RNA 4A will determine if this RNA is a generalised feature of cucumoviruses or if it is only found in Q-CMV.

It is not known if RNA 5 is common to all cucumoviruses or if it is only present in some CMV strains. RNA 5 was first reported in the subgroup II strain Q-CMV (Peden and Symons 1973). Two further reports indicated this population contained molecules derived from either RNA 2 (Blanchard 1991) or RNA 3 (Palukaitis et al. 1992). The characterisation of Q-CMV RNA 5 in Chapter 4 indicated that it consists of both sequence variants. Other reports have indicated that RNA 5 may be present in other subgroup II strains of CMV i.e., Kin and S (Boccard and Baulcombe 1993; Symons 1978; Gould et al. 1978; Avila-Rincon et al. 1986). It has been suggested that RNA 5 is restricted to subgroup II strains of CMV (Palukaitis et al. 1992).

Analyses of cucumovirus sequences indicated the presence of RNA 5 may not be restricted to strains (Bernal et al. 1991) but may not be present in all CMV strains (Symons 1985). Nucleotide sequence comparisons between the 3’ ends of CMV and
another member of the cucumovirus group, TAV, have revealed the presence of regions which are conserved in TAV and CMV subgroup II sequences but do not occur in CMV subgroup I sequences. Regions upstream of the 3’ terminal 149 nt of the 3’ non-coding region have a higher degree of sequence similarity between TAV and Q-CMV (71%) than between Q-CMV and Fny-CMV (37%, Bernal et al. 1991). One of the conserved regions corresponds to the first 23 nt of QCMV RNA 5 indicating that if RNA 5 is present in TAV it may contain an identical 5’ terminal nucleotide sequence.

In this chapter, the presence of RNAs 4A and 5 in TAV and subgroups I and II CMV strains is investigated. The detection of these RNAs was carried out using the primer extension method developed in Chapter 3.

4.2. Results

4.2.1. RNA 4A is Present in Both Subgroup I and II Strains of CMV

Primer extension reactions were performed on encapsidated RNA from a number of subgroup II strains of CMV using a subgroup II RNA 4A-specific primer. This primer (P4AII) has the sequence 5’-CTTCTTAACGGTAGTGC-3’ which is complementary to a region of Q-CMV RNA 4A beginning 49 nt from the 5’ end (Fig. 4.1). The extension products from primer P4AII on Q-CMV RNA 4A are consequently predicted to be 49 nt long.

Primer extension reactions were performed on six subgroup II CMV strains (Q, Pri, P, B14, Wc and Sn) to assay for the presence of RNA 4A. A dideoxy sequencing reaction with 32P end-labelled P4AII was used to show that the extension products were of the predicted size in all subgroup II strains examined (Fig 4.2). Two extension products which differ slightly in electrophoretic mobility (a doublet) were visible. This doublet may be due to the presence of two types of molecule; one which contains a 5’ cap structure and one without a cap. Primer extension of molecules with a cap produce extension products which have an extra nucleotide that alters their electrophoretic mobility resulting in a doublet when electrophoresed with extension products from non-capped molecules.
Chapter 4: Survey for RNAs 4A and 5 in V-TAV and a Number of CMV Subgroup I and II strains

Fig. 4.2: Detection of RNA 4A in subgroup II CMV strains by primer extension analysis. Lanes are labelled with the respective CMV strains used in the primer extension reactions. Lanes corresponding to the products from dideoxy sequencing with the P4AII primer and a Q-CMV RNA 4A cDNA clone are labelled A, C, G and T. Reaction products were electrophoresed on an 8% polyacrylamide gel containing 7M urea and autoradiographed.

A subgroup I-specific primer was designed to assay for RNA 4A in subgroup I CMV strains. This primer (P4AI) has the sequence 5’-GCACCTACGTTCAATTCC-3’. Its design was based on a putative conserved region in subgroup I RNA 2 sequences (obtained from the Genbank sequence database). The binding region of P4AI to putative RNA 4A molecules from CMV subgroup I strains O, Y and Fny is illustrated in Fig. 4.3. The sequence of L-, and Ny-CMV RNA 2 could not be included in Fig. 4.3 as they have not been determined. The putative start point of the RNA 4A molecules presented in Fig 4.3 is as proposed by Ding et al. (1994).

```
<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>50</th>
</tr>
</thead>
<tbody>
<tr>
<td>O</td>
<td>GTTTTGTAGT ACAGAGTTCA GGGTGAGCG TGTAAATTCC AATAAACAGC</td>
<td></td>
</tr>
<tr>
<td>Y</td>
<td>GTTTTGTAGT ACAGAGTTCA GGGTGAGCG TGTAAATTCC AATAAACAGC</td>
<td></td>
</tr>
<tr>
<td>Fny</td>
<td>GTTTTGTAGT ACAGAGTTCA GGGTGAGCG TGTAAATTCC AAcAAACAGC</td>
<td></td>
</tr>
</tbody>
</table>
```

Fig. 4.3: Binding site of P4AI to putative subgroup I RNA 4A sequences. The RNA 4A sequences from O-, Y- and Fny-CMV were aligned using the Pileup, and Pretty programs from the GCG package. The proposed binding site for P4AI is indicated (---). Only the first 100 nt of RNA 4A are presented.

Primer extension analysis was undertaken on encapsidated RNA from the two subgroup I CMV strains L- and Ny-CMV (Fig. 4.4). Fig. 4.3 indicates that extension products 79 nt in length will be produced by primer extension with P4AI on RNA 4A in subgroup I CMV strains if all subgroup I strains contain the binding site sequence and do not vary.
in length in this area. As no clone containing a subgroup I RNA 4A cDNA was available to create a sequencing ladder, the size of extension products was estimated using 5' end labelled oligonucleotides of known sizes (90 and 76 nt). Bands of the predicted size were produced from primer extension on RNA from both L- and Ny-CMV, indicating that RNA 4A is present in both of these subgroup I CMV strains (Fig 4.4).

**Fig. 4.4:** Detection of RNA 4A in subgroup I CMV strains by primer extension analysis. Lanes are labelled with the respective strains used in the primer extension reactions. The sizes of the labelled oligonucleotides are indicated. Labelled oligonucleotides and primer extension products were electrophoresed on a 6% polyacrylamide gel containing 7M urea and autoradiographed.

### 4.2.2. RNA 4A is Present in Another Cucumovirus, V-TAV

To investigate the possibility that RNA 4A is present in other cucumoviruses, a primer designed to bind to sequences corresponding to a putative RNA 4A molecule in V-TAV was constructed. This primer (P4A-TAV) had the sequence 5'-TTACTCTCAAACGCTTTTCTC-3'. The position at which P4A-TAV was designed to bind to the putative RNA 4A molecules is shown in Fig. 4.5.

**Fig. 4.5:** Binding site for P4A-TAV on putative RNA 4A from TAV. The first 100 nt of the region on V-TAV RNA 2 corresponding to the 5' end of the putative RNA 4A sequence are shown. The binding site of P4A-TAV is underlined.
Primer extension with P4A-TAV on total RNA isolated from V-TAV-infected *N. clevelandii* was performed to assay for the presence of RNA 4A in TAV. The size of the extension product was determined by comparison with the products from a dideoxy sequencing reaction on Q-CMV RNA 5 cDNA clone a21 with the RNA 5.2-specific primer P2-18. The sequencing product corresponding to the first nucleotide in the RNA 5 sequence was calculated to be the same size as the extension product expected from the primer P4A-TAV on TAV 4A (Fig. 4.6).

![Fig. 4.6: Detection of RNA 4A in V-TAV by primer extension analysis.](image)

Fig. 4.6: Detection of RNA 4A in V-TAV by primer extension analysis. The lanes corresponding to products of the primer extension reaction is labelled TAV. Lanes containing dideoxy sequencing reactions of the RNA 5 cDNA clone are labelled A, C, G and T. The band corresponding to the 5' terminal nucleotide of RNA 5.2 is labelled with an arrow. Reaction products were electrophoresed on a 6% polyacrylamide gel containing 7M urea and autoradiographed.

Fig. 4.6 shows the presence of an extension product with the approximate size predicted to be produced from primer extension with P4A-TAV on V-TAV RNA 4A. This extension product is approximately 1 to 2 nucleotides smaller than the predicted size based on comparison with the sequencing reaction. This may either indicate an error in the original sequencing of V-TAV RNA 2 or that the putative start point of TAV RNA 4A as predicted in Ding *et al.* (1994) is incorrect.

**4.2.3. Survey of Subgroup I and II Strains of CMV for RNA 5**

The presence of a highly-conserved region in the sequence corresponding to putative RNA 5 molecules in all published cucumoviral sequences provided the opportunity to survey all cucumoviruses for the presence of RNA 5 using a single primer. The region to which the primer binds in all sequenced TAV and CMV strains is shown in Fig. 4.7. The primer designed to bind to the conserved region (PCMV5) has the sequence...
Chapter 4: Survey for RNAs 4A and 5 in V-TAV and a Number of CMV Subgroup I and II strains

5'-CCATTTTAGCCGTAAGCTGG-3'. The position at which the primer binds to the aligned RNA 5 sequences (Chapter 3) is shown in Fig. 4.7.

<table>
<thead>
<tr>
<th>Consensus</th>
<th>CCAGCTTAG GCTAAAATGG</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fny- and Y-CMV RNA 1</td>
<td>CCAGCTaACG GCTAAAATGG</td>
</tr>
<tr>
<td>Try-CMV RNA 3</td>
<td>CCAGCTTAG aCTAAAATGG</td>
</tr>
</tbody>
</table>

Fig. 4.7: Conserved region present in all TAV and CMV strains that was used to detect RNA 5 in cucumoviruses. The 3’ end of all published TAV and CMV RNA sequences were aligned using the Pileup and Pretty programs of the GCG package. Only the conserved region is shown. The consensus sequence of this region is identical to the conserved sequence in most RNAs analysed (Q-CMV RNA 1, Fny-, Y-, O-, Q-CMV RNA 2, As-, Japanese Y-, Y-, L-, Fny-, Ny-, D-, M-, C-, O-, i17f-, nt9-, Kin-, Q-, Try-, Sn-, WI-CMV RNA 3, V-TAV RNA 1, V-TAV RNA 2, V-, C-, B-, P-TAV RNA 3) except for Fny- and Y-CMV RNA 1 which contained a substitution of A instead of T at position 7 on the consensus sequence and Try-CMV RNA 3 which contained a substitution of A instead of G at position 11 of the consensus sequence. The non-conserved nucleotides are shown in lower case.

| R5.2 1 GTCCGAAGACGTTAAACTACGCTCTCTTTATTGCGAGTGCTGAGTTGGTA 50 |
| R5.3 1 GTCCGAAGACGTTAAACTACACTCTC..AATCGCGAGTGCTGAGTTGGTA 48 |

Fig. 4.8: Binding site of the primer PCMV5 on Q-CMV RNA 5 molecules. The 5' 148 and 146 nt of Q-CMV RNAs 5.2 and 5.3, respectively, were aligned using the Bestfit program from the GCG package. Gaps (.) have been introduced to allow optimal alignment of the sequences. Vertical lines indicate conserved nucleotides. The position at which the primer, PCMV5, binds is indicated under the sequence (----).

| R5.2 51 GT.TTGCCTCTAATACTAATCTGAAGTCGTAAATCCATTACTTG.GTTCGCA 98 |
| R5.2 49 GTGCTGCTCATAAATCTGCTGAGTGCTAAATCCATTACTTG.GTTCGCA 98 |
| R5.2 99 CGGGTTGTGCTCATTCCAGCTCAGCTCAGCTCAAAATGGTCAGTCGTGCTCCTTAC 148 |
| R5.2 99 CGGGTTGTGCTCATTCCAGCTCAGCTCAGCTCAAAATGGTCAGTCGTGCTCCTTAC 146 |

PCMV5

Fig. 4.8 shows that primer extension on both Q-CMV RNA 5.2 and 5.3 with the primer PCMV5 should yield extension products 130 nt in length. Primer extension on encapsidated viral RNA preparations from CMV subgroup II (Q, Pri, P, B14, Wc and Sn) and subgroup I (L, and Ny) strains was carried out to assay for the presence of RNA 5 in these strains (Fig. 4.9). The size of the extension products was determined by comparison with dideoxy sequencing products with PCMV5 as the primer and the RNA 5.2 cDNA clone a21 as the template. Extension products of the approximate size
predicted from RNA 5 were detected in all subgroup II CMV strains examined indicating that all CMV strains may contain RNA 5. One of the two subgroup I strains (L) contained a band which is approximately 3-4 nt smaller than the subgroup II RNA 5 extension products. This is the first report of a subgroup I RNA 5 molecule. No extension product was seen in the primer extension reaction on Ny-CMV RNA.

Fig. 4.9: Detection of RNA 5 in subgroup I (Sub I) and II (Sub II) CMV strains. Lanes corresponding to products from dideoxy sequencing of the RNA 5.2 cDNA clone are labelled A, C, G and T. The band corresponding to the 5' terminal nucleotide of RNA 5.2 is labelled with an arrow. Lanes corresponding to each CMV strain are indicated. Reaction products were analysed on a 6% polyacrylamide gel containing 7M urea and autoradiographed.

The absence of an extension product corresponding to RNA 5 in the Ny-CMV encapsidated RNA preparation indicates that either (i) RNA 5 does not exist in this strain, (ii) RNA 5 synthesised but not encapsidated, (iii) RNA 5 synthesised but not accumulated, or (iv) the P4AI does not bind to Ny-CMV sequences. To investigate the possibility that RNA 5 is synthesised but not encapsidated, a primer extension reaction with PCMV5 was carried out on total RNA isolated from L- and Ny-CMV-infected plants. A sequencing reaction on Q-CMV RNA 5 cDNA clone a21 using PCMV5 as a sequencing primer was used as a size reference (see Fig. 4.10). An extension product indicative of the presence of RNA 5 was detected in total RNA isolated from the L-CMV infected plant but not in the Ny-CMV RNA preparation. The extension product from the L-CMV total RNA preparation was the same size as the product obtained with encapsidated L-CMV. To ensure that PCMV5 binds to Ny-CMV RNA, a sequencing reaction with this primer was carried out on an Ny-CMV RNA 4 cDNA clone. Clear sequence data was obtained indicating that P4AI is able to bind to the Ny-CMV sequence (data not presented). This indicates that RNA 5 is present in some but not all subgroup I strains of CMV.
Chapter 4: Survey for RNAs 4A and 5 in V-TAV and a Number of CMV Subgroup I and II strains

4.2.4. There are Two Sequence Variants in the RNA 5 Population in Subgroup II CMV Strains

As demonstrated in Chapter 3, RNA 5 in Q-CMV consists of two sequence variants derived from RNAs 2 and 3. To determine whether the RNA 5 population in other subgroup II CMV strains consists of a mixture of sequence variants, primer extension analysis was undertaken. A primer was designed which would bind to both Q-CMV RNA 5.2 and 5.3 giving differently sized extension products. The sequence of this primer (P5-doublet) is 5'-TGATTTCAAGGGTACCTCGAC-3' and it binds to Q-CMV RNA 5.2 and 5.3 at the region shown on the aligned sequences in Fig. 4.11. Fig. 4.11 predicts that the size of the extension product using P5-doublet produced from RNA 5.2 should be 191 nt and that from RNA 5.3 it should be 188 nt.

Fig. 4.11: Binding site of P5-doublet on Q-CMV RNAs 5.2 and 5.3. The 5' ends of Q-CMV RNA 5.2 and 5.3 were aligned using the Bestfit program from the GCG package. Gaps (.) have been introduced to allow optimal alignment of the sequences. Vertical lines indicate conserved nucleotides. The position at which P5-doublet binds is...
indicated under the sequence (---).

When the two extension products were electrophoresed on a 6% polyacrylamide denaturing gel, they appeared as two distinct bands (Fig. 4.12). The size of the extension products was determined by comparison with the products of dideoxy sequencing reactions of the RNA 5.3 cDNA clone a22 using P5-doublet as a sequencing primer as primer.

All subgroup II strains examined (Wc, B14, P, Pri, Sn, and Q) contained two populations of RNA 5 molecules with a size variation in the 3' end extended region (Fig. 4.12). Variation in intensity existed between the higher and lower molecular weight bands among the different strains. For example, the lower molecular weight band in the Wc-CMV lane is less intense than the higher molecular weight band. However, the opposite is the case with Pri-CMV and the two bands were of equal intensity in other strains. This indicates that RNA exists at different relative concentrations in different samples. The lower band corresponds to RNA 5.3 and the higher band corresponds to either RNA 5.1 or RNA 5.2. The exact origin of the higher band could not be determined as the predicted sizes for the extension products with RNA 1 and 2 are identical. However, as Q-CMV was not shown to contain RNA 5 molecules derived from RNA 1 (see Chapter 3), it is unlikely that other subgroups II strains contain RNA 1-derived RNA 5.

![Fig. 4.12: Detection of two sequence variants in the RNA 5 population of subgroup II CMV strains. Lanes corresponding to the products from dideoxy sequencing reactions of the RNA 5.3 cDNA clone are labelled A, C, G and T. The band corresponding to the 5' terminal nucleotide of RNA 5.3 is labelled with an arrow. Lanes corresponding to the subgroup II CMV strains are indicated by the strain names above each lane. Reaction products were electrophoresed on a 6% polyacrylamide gel containing 7M urea and autoradiographed.](image-url)
4.2.5. RNA 5 is Present in Another Cucumovirus, V-TAV

To investigate the possibility that RNA 5 exists in other cucumoviruses, primer extension analysis was carried out on V-TAV RNA using the primer, PCMV5. As shown in Fig. 4.13, V-TAV also contains the conserved region complementary to the primer PCMV5. The binding site of PCMV5 is illustrated on putative TAV RNA 5 sequences aligned in Fig. 4.13. The start point of the putative V-TAV RNA 5 sequence was calculated by comparison with the Q-CMV RNA 5 sequence. The 5' end of the putative V-TAV RNA 5 sequence is exactly the same as the corresponding region on Q-CMV RNA 5.

Fig. 4.13: Binding site of PCMV5 on putative V-TAV RNA 5 molecules. The 5' ends of putative V-TAV RNA 5 sequences were aligned using the Pileup and Pretty programs from the GCG package. Non-conserved molecules are shown in lower case. The position at which PCMV5 binds is indicated under the sequence (---).

Fig. 4.13 predicts that primer extension analysis with the primer, PCMV5, on the putative V-TAV RNA 5 should yield an extension product of 130 nt which is identical in size to the extension products produced with Q-CMV RNA 5. The identification of the 5' end of V-TAV RNA 5 sequence is based on the sequence similarity with Q-CMV RNA 5 which was identified by Bernal et al. (1991). Primer extension reactions were carried out on V-TAV RNA obtained by total RNA extraction from virus-infected N. clevelandii. The size of the extension products was determined by comparison with the products from dideoxy sequencing reactions of RNA 5.2 cDNA clone a21 as template with PCMV5 as a sequencing primer. An extension product of the size predicted to be synthesised from TAV RNA 5 is present in TAV (lane TAV of Fig. 4.14). This indicates that RNA 5 occurs in TAV and is the first demonstration of RNA 5 in another cucumovirus.
Chapter 4: Survey for RNAs 4A and 5 in V-TAV and a Number of CMV Subgroup I and II strains

4.2.6. The CMV Strain Which Does Not Contain RNA 5 Produces More Severe Symptoms Than Those That Do Contain RNA 5

RNA 5 was originally thought to be associated only with CMV subgroup II strains (Palukaitis et al. 1992). It has also been suggested that symptoms induced by strains in this subgroup were less severe (Owen and Palukaitis 1988). This may indicate that RNA 5 could play a role in symptom attenuation. However, comparison of symptoms from a number of CMV strains, from both CMV subgroups, have indicated that this is not necessarily the case (Blanchard, unpublished data). Also, it has been demonstrated here that RNA 5 is not limited to CMV subgroup II strains. Nevertheless, the possibility still exists that the presence of RNA 5 may correlate with a less severe phenotype. This possibility was supported by a comparison of the CMV strains which contained RNA 5 (Wc, Sn, Q, B14, P, and L) to the only strain which did not contain RNA 5 (Ny). Symptoms on *N. clevelandii* were much more severe on the only CMV strain found not to contain RNA 5 (NY-CMV) than the strains which did contain RNA 5 (Fig. 4.15). This observation does not confirm that RNA 5 actually functions in symptom attenuation as modulation of symptoms may be controlled by other viral or host factors. To accurately determine if such a role exists, comparisons would have to be made between the infectivity of a CMV strain with and without RNA 5. This could be achieved by deleting the region corresponding to the 5' end of RNA 5 from the genomic RNAs of a CMV strain which contains RNA 5.
Chapter 4: Survey for RNAs 4A and 5 in V-TAV and a Number of CMV Subgroup I and II strains

Fig. 4.15: Symptoms displayed by *N. clevelandii* infected with various strains of CMV. The severity of CMV infected plants are compared with a non-infected control plant (Control). Plants are labelled with the strain with which they were infected.

### 4.3. Discussion

Primer extension analyses performed here indicated that RNA 4A is present in both subgroups of CMV as well as in TAV. This supports the prediction by Ding *et al.* (1994) that RNA 4A is present in all cucumoviruses. The presence of RNA 4A in all cucumoviruses may indicate that this RNA plays an important role which may be related to the unique properties of cucumoviruses. For example, as compared with other members of the bromoviridae family, cucumoviruses have a wide host range.

RNA 4A may have arisen by "overprinting" (Keese and Gibbs 1992) on the 3' terminal coding sequence of the 2a gene. Overprinting refers to the translation of previously unused reading frames and in the case of RNA 4A, the nucleotide sequence used includes the 3' end of the 2a ORF as well as sequence downstream from the 2a stop codon. Similar overlapping genes have been found to correspond with previously-defined taxonomic groups. An example of this are the luteoviruses which contain a Mr 17,000 protein which is not present in other virus taxonomic groups. This supports the idea that the appearance of such genes may play a role in the generation of a new virus group and indicates that the emergence of RNA 4A may have played a role in the evolution of cucumoviruses (Ding *et al.* 1994).

The presence of a conserved open reading frame indicates that RNA 4A may contain a cap structure as caps are thought to play a role in translational. The proposal that RNA 4A contains a cap is supported by the electrophoretic pattern of the extension products produced from CMV subgroup II RNA 4A molecules. The electrophoretic pattern of TAV or CMV subgroup I RNA 4A extension products were not typical of those produced from a capped molecule, however, this may have been due to the lower
resolution of 6% polyacrylamide gels compared with the 8% polyacrylamide gel used with CMV subgroup II primer extension analysis.

The primer extension analyses presented here show that the previous report that RNA 5 is only present in CMV subgroup II strains (Palukaitis et al. 1992) is incorrect. RNA 5 was also shown to be present in TAV and a CMV subgroup I strain. The presence of RNA 5 in V-TAV was not unexpected in view of the nucleotide sequence similarities between the 3' ends of Q-CMV and V-TAV. Like CMV subgroup II strains, V-TAV contains a conserved region approximately 300 nt in length at the 3' end of all three genomic RNAs. When the consensus of these conserved regions of V-TAV are aligned with a consensus of the conserved regions at the 3' end of Q-CMV, identical sequences in the region corresponding to the 5' end of RNA 5 can be seen between the two virus sequences (Fig. 4.16). The 5' terminal conserved region is absent in an alignment of the Q-CMV RNA 3' end consensus sequence to that of the CMV subgroup I strain Fny (Fig 4.17). The absence of this conserved region may account for the absence of RNA 5 in some subgroup I strains. These conserved sequences may have a function in the synthesis of RNA 5. The variation around the 5' terminal sequence of the region encompassed by RNA 5 and between subgroup I and II CMV strains may also be related to the size variation between the extension products produced by the subgroup II strains and L-CMV.

Q-CMV 1 GTCCGAAGACGTTAAACTACGCT..CTCTTTATTGCGAGTGCTGAGTTGG 48
V-TAV 1 GTCCGAAGCATGTTAAACTACGCTTTGAACCGTGTTCGAGTGTCTGAGTTGG 50
49 TAGTTTTGCTCTAAACTATCTGAAGTCGCTAAATCCATTATTGXGTTGCG 98
51 TAGTATTCCTAAACTATCTGAAGTCACTAAACGC....TTGTGCGGTG 96
99 AACGGGTTGTCCATCCAGCTTACGGCTAAAATGGTCAGTCATGCCCAA 148
97 AACGGGTTGTCCATCCAGCTAACGGCTAAAATGGTCAGTCATGTCGT 145
146 GACATGCCGTC..GGCTTTGTAGGAGGAGCTCTTTGGAAATCTCGTAA 197
134 GACATGCCGTCT..GGCTTTGTACGAGGAGCTCTTTGGAAATCTCGTAA 193
198 CTAGATTTCTTCGGAAGGGCTTTCGTGAGAAGC..............TCGT 233
194 CCGGGTTGTCTCCGGAAGGAGCTCTTTGGAAATCTCGTAA 243
234 GCACGGTATACACTGATATTACCAAGAGTGCGGGTATCGCCTGTGGT 281
244 GCACGGTATACACTGATAGCTTCAGACATGCTTCCGAGGCTCTACGG 293
282 TTCCCAACGGTTCTCCATAGGAGACCA 309
294 TTCCGATAGGCTTCCCGCTAGGGGTCTCCA 321

Fig.4.16: Alignment of Q-CMV and V-TAV 3' terminal consensus sequences. The 3' terminal conserved regions of Q-CMV and V-TAV were aligned using the Pileup program. Consensus sequences were then generated using the Pretty and Ugly programs.
and aligned using the Bestfit program. Gaps (.) have been introduced to allow optimal alignment of the sequences and conserved nucleotides are joined by a vertical line. X indicates the inability of the Pileup program to generate a consensus sequence.

Q-CMV 1 ...........GTCCGAAGACGTAAAACCTTTATTGCGAGTG 39
Fny-CMV 1 ATTCGAAACCTTXXCTCTXGCTXGXXCTXGXXCTXGTGTTGGCGGAX 50

Fig. 4.17: Alignment of Q-CMV and Fny-CMV 3' terminal consensus sequences. The 3' terminal conserved regions of Q-CMV and the 3' terminal 310 nt of Fny-CMV were aligned using the Pileup program. Consensus sequences were then generated using the Pretty and Ugly programs and aligned using the Bestfit program. Gaps (.) have been introduced to allow optimal alignment of the sequences and conserved nucleotides are joined by a vertical line. X indicates the inability of the Pileup program to generate a consensus sequence.

To investigate whether the 5' terminal sequences of RNA 5, which are conserved in Q-CMV and V-TAV, are found in other viruses, the first 23 nt of Q-CMV (and V-TAV) RNA 5 were used to search the nucleotide sequence database (Genbank) for similar sequences. A search with the Fasta and Blastn programs from the GCG package revealed the presence of sequences almost identical to the 5' terminal 23 nt of Q-CMV RNA 5 in a proposed member of the furovirus group, beet necrotic yellow vein virus (BNYVV). BNYVV generally contains four RNA species termed RNA 1 (6.8 kb), RNA 2 (4.7 kb) RNA 3 (1.8 kb) and RNA 4 (1.8 kb) (Balmori et al. 1993). Three subgenomic RNA species are produced from RNA 2 and one subgenomic RNA (RNA 3sub) is produced from RNA 3. The subgenomic RNA promoter responsible for the synthesis of RNA 3sub has been mapped to a region beginning 16 nt upstream of the start site of RNA 3sub on RNA 3 and extends to somewhere between 100 and 208 nt
downstream of the start site of RNA 3sub. The region identical to the 5' end of Q-CMV RNA 5 begins four nucleotides downstream of the start site of RNA 3sub and therefore forms part of the subgenomic promoter (Fig 4.18). This indicates that the 5' end sequence of Q-CMV RNA 5 may also form part of a subgenomic promoter involved in the synthesis of Q-CMV RNA 5.

\[
\begin{align*}
\# & \quad \ast \\
GGGCAAAUUAAUUAAUUUCUUUGUGAAUCGUCCGAAGACGUUAAACUACACGUGA
\end{align*}
\]

Fig. 4.18: Location of Q-CMV RNA 5 5' terminal sequence on BNYVV RNA 3. The region identical to Q-CMV RNA 5 is underlined. The start point of RNA 3sub (\(\ast\)) and the start of the RNA 3sub subgenomic promoter (\(#\)) are indicated.

In addition to the RNA species mentioned above, two more RNA species are sometimes present in BNYVV (RNA 5, 1.4 kb and RNA 6, 1 kb). It is thought that RNA 6 is derived from RNA 5 (Brunt and Richards 1989). RNA 5 from four isolates has been cloned and sequenced (Kiguchi et al. unpublished data) and a sequence identical to the 5' terminal 20 nt of Q-CMV RNA 5 is present on all four BMYVV RNA 5 sequences. In each case, the identical sequence begins 201 nt from the 3' end of the RNA. It is not known what the role of this sequence is in BNYVV RNA 5, however, based on the position it has in the RNA 3sub subgenomic promoter, it is possible that it may be involved with the production of small 3' coterminal RNAs similar to RNA 5 in Q-CMV. Thus far there have been no reports of the presence of such an RNA population in BNYVV.

Comparison of RNA 3 sequences from various TAV strains revealed a novel insertion in the 3' UTR of V-TAV RNA 3 (F. García-Arenal, personal communication). The first 164 nt of the conserved region, corresponding to the putative RNA 5 sequence, is directly repeated. If the 5' terminal sequence that is conserved between Q-CMV and TAV RNA 5 is responsible for the synthesis of RNA 5, it is possible that an additional RNA is produced in V-TAV which is 164 nt larger than RNA 5. If this is the case, primer extension with the PCMV5 primer on V-TAV may yield three possible extension products; one from the proposed RNA 5 sequence corresponding to the 3' terminal conserved region, another due to the primer binding site being repeated upstream which would produce an extension product the same size as the previously mentioned product, and another which is 164 nt larger due to the addition of the repeated sequence. It is not known whether the extension product in Fig. 4.14 corresponds to the first or second extension product mentioned above. A primer extension product which may correlate with the larger molecule was produced by primer extension with the PCMV5 and V-TAV primers. The size of this band was estimated by comparison with the sequencing
reaction used to determined the size of the smaller product. However, difficulty with resolving bands in the sequencing ladder in this area of the gel did not allow confirmation of the exact size (data not presented).

Although the characterisation of Q-CMV RNA 5 in Chapter 3 indicated that this population consisted of molecules derived from RNA 2 and 3, the exact nature of the population in other cucumovirus is not known. Based on primer extension analysis with the primer P5-doublet, all the subgroup II strains appear to contain two sequence variants like that seen in Q-CMV. These are probably derived from RNAs 2 and 3, but the possibility of an RNA 1-derived RNA 5 molecule cannot be eliminated. The difference in the relative amounts of the RNA 5 sequence variants in individual strains may indicate that either (i) the different variants are produced by two different mechanisms, (ii) the variants are differentially expressed in different strains, or (iii) variants are differentially expressed throughout different stages of the virus or plant life cycle. It would be interesting to study the effect of encapsidation and various host have on the differential levels of RNA 5 variants in RNA preparations.

Primer extension analysis indicates that RNA 4A is likely to be produced by all cucumoviruses and that RNA 5 is produced in all TAV strains, all CMV subgroup II strains, but may be only produced in one or some CMV subgroup I strains. This indicates the role of RNA 4A may be more important than that of RNA 5 which is not as widely distributed. As CMV subgroup II RNA 4A is easily detectable on ethidium bromide stained agarose gels and CMV subgroup I RNA 4A is not, the general level of expression between the two subgroups may be different or that RNA 4A may not be encapsidated as efficiently in subgroup I strains. The presence of a sequence at the 5' end of RNA 5 that is identical to a sequence in a BNYVV subgenomic promoter provides further evidence that RNA 5 is produced as a subgenomic promoter. The observation that the CMV strain which does not contain RNA 5 (NY) produces the most severe symptoms of the strains examined may indicate that the natural role of RNA 5 is involved with symptom attenuation and that RNA 5 may be useful in the design of a strategy to engineer virus tolerant plants.
Chapter 5: Investigation of the Possible Use of Q-CMV RNA 5 to Genetically Engineer Virus-Resistant Plants

5.1. Introduction

Cucumber mosaic virus is a serious pathogen which is difficult to control by conventional means. This difficulty is due to the wide host range of CMV and the mechanism of infection. CMV is seed transmissible in a number of economically important crop species which makes obtaining virus-free seed difficult. Seed transmission can be controlled to a certain extent by certain agronomic management practices and seed certification systems but these methods are not effective in protecting plants from infection via aphid transmission from neighbouring infected plants. CMV infection may be overcome by the construction of virus-resistant plants using genetic engineering. A number of examples of genetically-engineered resistance to CMV and other viruses already exist. Most strategies for engineering virus resistance involve introducing into the genome of the plant virus sequences (PDR) or sequences which interrupt the life cycle of the virus. The various strategies for engineering virus-resistant plants used so far and some of their advantages and disadvantages are outlined below.

Coat protein-mediated resistance (CPMR) was the first reported example of pathogen-derived genetically engineered resistance (Powell-Abel et al. 1986) and has since been reported in more than 20 different viruses (Wilson 1993) including CMV (Cuozzo et al. 1988; Quemada et al. 1991; Namba et al. 1991). The resistance afforded by a coat protein may be due to the presence of the protein product or the RNA transcript, as resistance has been observed with constructs containing functional (Beachy et al. 1990) and non-functional (Lindbo and Dougherty 1992; Lindbo et al. 1993) ORFs. The level of CPMR ranges from a delay in symptoms (Powell-Abel et al. 1986) to immunity (Hemenway et al. 1988; Stark and Beachy 1989; Tumer et al. 1987). Early examples indicated that the coat protein was interfering with events occurring early in the infection process such as uncoating or gene expression. This was supported by the ability to overcome the CPMR by inoculation with viral RNA. However, other examples have shown that the coat protein may interfere with later events in the infection process such as spread of the virus (Beachy et al. 1990).

An advantage of CPMR is that all viruses contain a coat protein gene, and hence this strategy may be attempted with all viruses. Also, in rare cases, the introduction of a coat protein gene from one virus results in resistance to a number of viruses which are only distantly related (Stark and Beachy 1989; Ling et al. 1991). A disadvantage of CPMR
is the possibility of transencapsidation of other viral RNAs, however, this may be overcome by using constructs which do not contain functional ORFs. Plants which only provide a low level of resistance may result in an increased pressure for the evolution of resistance-breaking virus strains due to the ability of virus replication to occur in these plants. The resistance shown in some examples is limited to the strain from which the coat protein gene is derived or closely related strains (van Dun et al. 1988), which means that multiple gene constructs would have to be used to achieve multiple virus resistance.

Plants immune to viral inoculation have been engineered using gene constructs containing viral sequences derived from replicase genes (reviewed by Carr and Zaitlin 1993). The first example of this type of resistance was discovered accidentally when a sequence coding for the 54-kDa open reading frame corresponding to part of the replicase gene in tobacco mosaic virus was introduced into the genome of tobacco plants (Golemboski et al. 1990). Introduction of this gene resulted in extreme resistance to TMV or TMV RNA, even at the highest inoculum level of 500 µg/mL. The resistance, however, was only effective against the strain of TMV from which the gene was derived and some closely related strains. Similar results have been obtained with other viruses using either unmodified replicase genes or those which have been mutated to produce non-functional proteins (Wilson 1993).

Resistance to CMV has also been achieved through transformation of a host species with a truncated replicase gene which is missing the GDD motif thought to be essential for a functional replicase activity (Anderson et al. 1992). As with the TMV example, resistance was specific and only effective against strains of CMV in the same subgroup. An additional problem to the specificity of the resistance is the fact that this type of resistance may not be effective in all viruses. Transgenic plants expressing replicase genes of AMV and BMV were not resistant to infection with virus particles (Wilson 1993). However, the apparent lack of resistance with these viruses may be due to an inadequate number of transgenic lines being tested which may have only detected susceptible lines.

It is unclear whether the resistance is mediated by protein encoded by the replicase gene or by the RNA molecules themselves. It is likely that different mechanisms occur in different virus systems (Carr and Zaitlin 1993). For example, engineered resistance with TMV (Carr et al. 1992) and pea early browning virus (McFarlane and Davis 1992) appears to require expression of the gene at the protein level, however, resistance to tomato spotted wilt virus in tobacco is thought to be mediated by the gene rather than the protein it encodes. Also non-translatable, truncated BMV replicase genes were able
to inhibit replication in protoplasts (Marsh et al. 1991). The extreme resistance often associated with replicase-mediated resistance makes it a useful strategy, however it is not suitable when resistance to a number of different viruses is required.

Recent studies on the expression of movement proteins (MP) in transgenic plants have indicated that broad resistance may be achieved with these genes. Transgenic plants expressing a defective MP from TMV were shown to resist infection by a tobravirus, a nepovirus, an alfamovirus, a caulimovirus and a cucumovirus (Cooper et al. 1995). Similarly, transgenic plants expressing a mutated MP from the white clover mosaic virus species of the potexvirus group were resistant to infection by two other potexvirus species as well as the carlavirus potato virus S (Beck et al. 1994). The broad specificity of resistance with MP-mediated resistance indicate that this strategy may be useful when resistance to a number of different virus strains is required from a single gene. As CMV RNA 4A is thought to play a role in movement, it may be potentially useful in developing transgenic plants which are resistant to CMV as well as other related viruses.

Some DI particles (but not all) interfere with the replication of standard virus which often results in attenuation of viral symptoms. Due to the natural symptom attenuation properties of DI particles, attempts have been made to use them to construct virus-resistant transgenic plants. *Nicotiana benthamiana* plants were transformed with a naturally occurring DI particle associated with the ss DNA virus african cassava mosaic virus (ACMV, Stanley et al. 1990). These plants showed less severe symptoms when challenged with two different isolates of the standard virus. The reduction in symptoms was associated with mobilisation and amplification of the DI particle which caused a reduction in the amount of standard virus produced. Passaging of virus between transformants resulted in progressive reduction in the level of viral DNA and enhanced symptom amelioration. Naturally-occurring DI particles were also used to engineer virus resistance to cymbidium ringspot virus which contains a ss RNA genome. Plants expressing DI RNA were protected from apical necrosis and death normally caused by the virus (Kollar et al. 1993).

A possible limitation to the use of DI particles in engineering virus-resistant plants is the absence of naturally occurring DI particles associated with many plant viruses. This problem may be overcome by the construction of artificial DI particles. This has been achieved with BMV where artificial DI RNAs derived from RNA 2 were shown to dramatically reduce the accumulation of standard virus in barley protoplasts (Marsh et al. 1991). If DI RNAs can be generated which are replicated by a number of viruses using generalised replicase signals, these may be used as the basis of an effective
method of generating plants which are resistant to a number of different viruses.

Antisense RNAs are complementary strands of RNA with the ability to bind to RNA transcripts to prevent translation. Limited protection has been achieved with antisense RNA in a small number of viruses (Wilson 1993) including CMV (Rezain et al. 1988). Actually the resistant plant in the CMV study had very low mRNA levels indicating the resistance may have actually been due to sense suppression or homology-dependent gene silencing as proposed by Mueller et al. (1995). The problem with antisense resistance appears to be the high levels of antisense RNA required to provide protection (Rezain et al. 1988). This results in protection only at low inoculum levels. This method of engineering virus resistance is one of the least effective strategies.

Satellite RNAs have been found associated with a large number of plant virus groups including: the carmo- (Altenbach and Howell 1981), cucumo- (Richards et al. 1978), furo- (Bouzoubaa et al. 1985), luteo- (Miller et al. 1991) nepo- (Owens and Schneider 1978), sobemo- (Keese et al. 1983), tombus- (Gallitelli and Hull 1985) and umbraviruses (Murant et al. 1988). Thus, resistance strategies using satellite RNAs may be applied to a large number of virus groups. Satellite RNAs have been known to attenuate viral symptoms and it is for this reason they have been targeted for use in genetically engineered virus resistance. Harrison et al. (1987) were successful in obtaining virus-resistant transgenic plants which expressed a satellite RNA from CMV. Levels of satellite RNA dramatically increased upon infection of the plants with satellite-free CMV. This increase in satellite RNA levels corresponded to significantly lower virus levels in these plants as compared to the control plants. Similar results were obtained with transgenic tobacco plants expressing RNA transcripts of full length tobacco ringspot virus satellite (sTRSV) RNA or the complementary sequence (Gerlach et al. 1987). However, the way in which satellite RNA-mediated resistance works is not understood. It is thought that the satellite RNA competes with the helper virus for a limited amount of replicase resulting in inhibition of virus particle accumulation. Additional factors must also be involved because symptom attenuation has been observed without a corresponding reduction in virus particle accumulation (Baulcombe et al. 1989).

The expression of satellite RNAs to confer resistance has a number of potential disadvantages. Even though many satellite RNAs attenuate viral symptoms, there are some which actually enhance viral symptoms. It has been shown that the difference between symptom-attenuating and symptom-enhancing satellite RNAs may be as little as one nucleotide (Sleat and Palukaitis 1990). If such a mutation was to occur in the field, the viral disease would be intensified and this necrogenic satellite may be
transferred to neighbouring plants through aphid transmission. Another possible limitation on this strategy is the possibility that a satellite which attenuates symptoms associated with one helper virus may act synergistically with another virus in the same crop to cause more severe symptoms (Wilson 1989). A further disadvantage of this strategy would be the fact that the transgenic plants would not be completely immune, thereby allowing them to support low levels of virus replication. This would provide a reservoir of virus available for transmission by insect vectors.

RNA molecules capable of sequence-specific cleavage have been identified and are often termed ribozymes. Examples of these ribozymes include the bacterial RNase P which cleaves precursor tRNAs (Cech and Bass 1986, and references therein; Symons 1991), a number of small plant RNAs including the satellite RNA of tobacco ringspot virus (sTRSV) (Prody et al. 1986; Buzayan et al. 1986 and virusoid RNA of lucerne transient streak virus (vLTSV) (Forster and Symons 1987a 1987b), and hepatitis delta virus (HDV) RNA (Branch and Robertson 1991 and references therein). Ribozymes discovered so far can be divided into three different types: "hammerhead" (e.g. [+] sTRSV), "hairpin" (e.g. [-] sTRSV), "axehead" (e.g. HDV). The mechanisms by which ribozymes cleave RNA are not known, however, it has been established that Mg$^{2+}$ is required for cleavage (Zaug and Cech 1986). The RNA-processing abilities of ribozymes makes them particularly useful in strategies designed to engineer virus-resistant plants. Strategies for engineering virus resistance which use ribozymes usually involve the expression in transgenic plants of ribozymes which are designed to cleave viral sequences. Lamb and Hay (1990) have shown this to be possible in vitro with two ribozymes directed against sequences within the coat protein and polymerase genes of potato leafroll virus (PLRV). A number of examples of trans-acting ribozymes have been reported though these experiments either showed that the ribozymes were no better at inactivating the target gene than corresponding antisense constructs (Atkins et al. 1995; de Freyter et al. in press) or did not contain controls to address this possibility (Wegener et al. 1994; Mc Intyre et al. in press).

Ribozymes also have potential uses in the construction of plasmids used in genetic engineering (Alterschuler et al. 1992). A vector containing 5' and 3' processing ribozymes was designed by Taira et al. (1991) which would allow the production of RNA transcripts with defined 5' and 3' ends. When expressed in vivo these defined ends may allow viral RNA to fold into the natural form which may be advantageous in some strategies for engineered virus resistance such as that obtained with defective interfering particles.

Molecules derived from the 3' end of a viral genome have been shown to be effective in
blocking viral replication in vitro. Morch et al. (1987) produced cDNA clones of the 3' end of turnip yellow mosaic virus (TYMV) containing a tRNA-like structure which is thought to play a role in recognition by the replicase. RNA transcripts from these cDNA clones were shown to cause a 50% inhibition of replication in vitro when the transcript-derived molecules were added to the genomic RNA at an equivalent molar concentration. A 10-fold molar excess of the 3' molecules over the genomic RNA resulted in a reduction of genomic RNA replication of more than 90%. It was thought that this inhibition resulted from competition between the 3' molecules and the genomic RNA. This work was extended by Zaccomer et al. (1993) with the expression of the 3' terminal 100 nt of the non-coding region of TYMV in Brassica napus. Plants expressing high levels of the transgenes showed partial protection to infection with viral RNA or particles. Some of the plants did not become infected, indicating that the introduced sequences may have produced an RNA-mediated resistance similar to that observed with replicase or coat protein genes. This may also support the idea that any segment of a viral genome is capable of producing a highly resistant state by RNA-mediated resistance.

In addition to the strategies mentioned above, numerous other methods are being investigated to engineer virus resistance with the use of genes not derived from the host or the viral pathogen. Inactivation of the gene encoding s-adenosyl thomocysteine hydrolase (SAHH) in host plants was shown to confer resistance to a wide range of plant viruses. This was thought to occur through either inhibition of 5' capping of viral RNAs or because of an increased level of cytokinins (which have been found to induce acquired resistance, Masuta et al. 1995). Other novel resistance strategies include the expression of virus-specific antibodies in plant cells, expression of genes which encode toxic substances upon virus infection to quickly kill infected cells, and expression of defective virus transmission protein (a protein which is believed to recognise the viral coat protein and a surface receptor in the insect to facilitate transmission) to eliminate spread of virus from infected to healthy plants. (Wilson 1993). One or more of the strategies mentioned here may become the basis of a revolutionary approach to engineering virus resistance, however, further research is needed to establish the mechanisms by which they operate before they can be safely used in the field.

Q-CMV RNA 5 may be useful in genetically engineering virus-resistant plants for a number of reasons. RNA 5 is thought not to encode any proteins which may have detrimental effects on the host. It contains sequences or secondary structures thought to be recognised by replicase complexes for the initiation of (-) strand synthesis of the RNA (Boccard and Baulcombe 1993). These signals may be recognised by replicase complexes derived from other viruses which are only distantly related (as shown by the
ability of BMV to recognise the tRNA-like structure of CMV, Rao and Grantham 1994), indicating that it may be successful in affording resistance to a large number of viruses. As RNA 5 is abundant in encapsidated CMV preparations it is likely that RNA 5 is efficiently packaged into virus particles which may result in the symptom-attenuating molecules being transmitted with the virus, possibly reducing viral symptoms in subsequently infected hosts. Also, RNA 5 is small and easily manipulated for construction of cassettes used for virus resistance. Even if RNA 5 is not able to impart resistance by competition with the replicase, it is likely that a proportion of the plants would be resistant due to an RNA-mediated mechanism. It is for these reasons that the use of RNA 5 in engineered viral resistance was pursued. Engineered resistance to TYMV using molecules derived from the 3’ end of the viral genome is the only example which has been used to successfully engineer resistance with 3’ coterminal molecules. In this example, it was thought that TYMV-derived molecules were able to compete with the genomic RNAs for replication factors and reduce their replication (Zaccommor et al. 1993). Preliminary investigations of this method demonstrated that a large molar excess was required to provide a significant reduction in replication (Morch et al. 1987). This indicates that a high level of RNA 5 may be required to produce a significant level of resistance to CMV.

RNA-mediated resistance has been observed in a number of systems, however, the mechanism of action is not known. It is thought that the reduction of viral RNA may be mediated by a cellular pathway which is normally involved in the elimination or down regulation of aberrant or overexpressed mRNAs (Linbo et al. 1993; Dougherty et al. 1994). Three possible phenotypes have been described with this model which are related to the level of transgene expression (Smith et al. 1994). If transgene mRNA level reaches a certain threshold level, the cytoplasmically-based, cellular process which targets the RNA for elimination is activated and steady state levels of the transgene mRNA are reduced and the transgenic plants are highly resistant to infection with viral RNA containing the transgene sequence. The second phenotype is termed a recovery phenotype as viral RNA levels do not reach the activation threshold until the transgenic plants are inoculated with the virus. The RNA elimination process is activated upon virus inoculation resulting in the reduction of the transgene mRNA and producing a virus-resistant state. The final phenotype occurs when transgene expression is not high enough to activate the RNA elimination process and plants are susceptible to viral infection. To investigate the possibility of using RNA 5 to genetically engineer resistance via an RNA-mediated mechanism or by competition with replication factors, transgenic *N. tabacum* plants expressing RNA 5 were constructed and assessed for their ability to resist CMV infection.
5.2. Results

5.2.1. Subcloning of Q-CMV RNA5 cDNA into Binary Vectors

To facilitate expression of RNA 5 in transgenic plants, the RNA 5 cDNA from RNA 5.2 cDNA clone a21 was subcloned into the vector pWM38 which contains a cauliflower mosaic virus 35S transcription promoter (CaMV 35S), the 5' untranslated region of alfalfa mosaic virus (AMV 5' UT), which is thought to enhance translation and the 3' untranslated region from the tobacco ribulose bisphosphate carboxylase small subunit gene (Tob SSU 3') which functions as a transcription terminator. To confirm that RNA 5 cDNA was inserted in the vector and was in the correct orientation, plasmid DNA was isolated and subjected to restriction enzyme analysis (data not shown). The expression cassette of this newly constructed plasmid containing the promoter sequence, RNA 5 cDNA and terminator sequence was then subcloned into a binary plasmid pTAB10. This plasmid contains a CaMV35S promoter and the octopine synthase 3' termination signal regulating the transcription of a Basta resistance gene (bar) which encodes phosphinothricin acetyl transferase. This gene was used as a selectable marker in the transformation protocol. The binary plasmid also contains a tetracycline resistance gene used to select transformed bacteria. A graphical representation of the recombinant plasmid is illustrated in Fig. 5.1.

![Binary plasmid containing Q-CMV RNA 5 expression cassette](image)

Fig. 5.1: Binary plasmid containing Q-CMV RNA 5 expression cassette. The plasmid is labelled to indicate the location of the right border (BR), OCS 3' terminator signal
(OCR 3’), the bar gene (BAR), CaMV 35S transcription promoter (CaMV 35S), alfalfa mosaic virus 5' untranslated region (AMV 5' UT), RNA 5 cDNA (RNA 5), tobacco rubisco small subunit 3' untranslated region (Tob SSU 3' UT), the left border (BL) and the tetracycline resistance gene (Tet). The positions of restriction enzyme sites are also indicated.

5.2.2. Transformation of N. tabacum with the Binary Vector Containing Q-CMV RNA 5

The binary vector containing the RNA 5 cDNA was transferred into Agrobacterium tumefaciens (strain AGL 1) using triparental mating. To check that the A. tumefaciens contained the binary plasmid, DNA was isolated and subjected to restriction enzyme analysis. Restriction enzyme digestion patterns of plasmid DNA from A. tumefaciens after transfer were identical to digestion patterns of DNA isolated from E. coli containing the same binary plasmid construct (data not shown). The region between the left and right borders on the binary plasmid was transferred into the genome of N. tabacum (cultivar Wisconsin 38) using Agrobacterium-mediated DNA transfer. Transgenic plants were regenerated from transformed cells on media containing phosphinothricin for selection.

5.2.3. PAT Assay of Putative Transgenic Plants

To determine if putative transgenic plants contained the gene construct, an assay to detect the presence of the bar gene product, phosphinothricin acetyltransferase (PAT) was undertaken. Fig. 5.2 shows the results of PAT assays on the 16 putative transformed plants. Results on the putative transgenic plants are compared to assays on protein extracted from a transgenic plant known to contain the bar gene and a non-transgenic control plant.

![Fig. 5.2: PAT assay on putative transgenic plants. Autoradiograph of the products of PAT assays separated by thin layer chromatography. The reaction products from a transgenic plant known to contain the bar gene is labelled (+) and the products from a non-transgenic plant are labelled (-). Putative transgenic plants are labelled with their designated numbers. The band corresponding to acetylated phosphinothricin is indicated (*).]
Plants expressing the *bar* gene contain a product (acetyl phosphinothricin) in the bottom row of the autoradiograph shown in Fig. 5.2. Plants 1, 2, 5, 6, 7, 14, 15, and 16 contain bands corresponding to the band in the positive control track indicating the presence of an active *bar* gene in these lines.

### 5.2.4. Presence of the RNA 5 Transcript in Transgenic Plants

Although plants 1, 2, 5, 6, 7, 14, 15, and 16 were shown to express the *bar* gene, this does not mean that RNA 5 will also be expressed in these plants. To confirm that RNA 5 is expressed in these plants, the presence of RNA transcripts was determined by synthesising first-strand cDNA using the primer PQ-CMV3' (5'-CCGGATCCTGGTCTCCTTATGGAGAACC-3') and reverse transcriptase. Second strand cDNA synthesis and subsequent amplification of double-stranded cDNA was then carried out using PCR with the original 3' primer used to clone RNA 5 (PR55', 5'-GCGCTGCTATCGACTCACTATAGTCCGAAGACGTTAAAC-3'). The expected size of the PCR product which contained the RNA 5 cDNA as well as extra sequence on the primers included for cloning purposes was 341 bp. The sizes of PCR products were determined by comparison with DNA fragments from the *Eco RI*-digested bacteriophage SPP 1. Negative controls consisted of PCR reactions with no template or with total RNA extracted from non-transgenic plants. Products of the molecular weight predicted to be synthesised from RNA 5 are present in all the lanes corresponding to the transgenic lines. No PCR products were detected in the negative controls. This indicates that RNA 5 is probably expressed in all transgenic lines tested, however, there is a possibility that the PCR products were generated from genomic DNA if complete digestion with deoxyribonuclease before the transcription did not occur.

![Fig. 5.3: Determination of the presence of RNA 5 in transgenic plants by RT-PCR. PCR reactions are labelled with the respective transgenic plant numbers. RT-PCR](image-url)
products were analysed on a 1% agarose gel and made visible by ethidium bromide staining. Lanes containing PCR products from the various transgenic plants are indicated numerically. Negative controls containing PCR with total RNA from a non-transgenic plant (NegA) or with no RNA template (NegB) are indicated. The lanes containing the SPP1/Eco RI molecular weight marker is labelled M.

5.2.5. Virus Challenge of Transgenic Plants

To assess whether plants expressing RNA 5 could resist virus infection, transgenic plants were multiplied by vegetative propagation using tissue culture. Leaf discs from each transgenic line were incubated on Murashige and Skoog shoot-inducing medium (MS9, Murashige and Skoog 1962). These shoots were transferred to Murashige and Skoog root-inducing medium (MS0, Murashige and Skoog 1962) until they were ready for transplanting into soil in the glasshouse.

Virus particles used to inoculate transgenic lines were isolated from Q-CMV-infected *N. clevelandii* (Peden and Symons 1973). To assess the integrity of the virus particles and ensure virus preparations were free from contamination with other viruses such as tobacco mosaic virus, particles were examined under an electron microscope.

Transgenic lines were inoculated with Q-CMV virus particles at different inoculum concentrations as outlined in Chapter 2. Visual comparison between symptoms of inoculated transgenic plants and uninoculated plants indicated that the transgenic lines were as susceptible to Q-CMV as the non-transgenic lines. This was confirmed by ELISA (Clark and Adams 1977) on leaf samples taken three weeks after inoculation where no significant difference was detected between the level of virus in transgenic plants and non-transgenic plants.

5.3. Discussion

Transgenic plants expressing Q-CMV RNA 5 could have resisted infection by Q-CMV, using one of the following mechanisms: (a) Transgenic plants expressing RNA 5 molecules could compete with viral genomic RNAs for the viral replicase and reduce the synthesis of viral genomic RNAs in a similar way to the reduction of symptoms caused by TYMV 3’ coterminal molecules (Morch *et al.* 1987; Zaccomer *et al.* 1993) or the DI particle of ACMV (Stanley *et al.* 1990) when expressed *in vivo*. This may be possible as RNA 5 contains signals thought to be responsible for replicase recognition (Boccard and Baulcombe 1993). Further evidence to indicate that virus replication can be reduced by a competition mechanism has been provided by the recent discovery that
Chapter 5: Investigation of the Possible Use of Q-CMV RNA 5 to Genetically Engineer Virus-Resistant Plants

Satellite RNAs reduce viral RNA symptoms by competing with the genomic RNAs for replication factors (Wu and Kaper 1995). (b) The expression of RNA 5 in transgenic plants may have induced a resistance phenotype as seen with RNA-mediated resistance (Smith et al. 1994). This method would rely on engineering plants with a level of RNA 5 expression which was high enough to activate the RNA processing mechanism thought to cause the virus-resistant state (Smith et al. 1994). (c) Expression of RNA 5 in transgenic plants may have resulted in resistance in previously-uncharacterised ways such as the in vivo expression of an inhibitory protein encoded by RNA 5 which may have attenuated virus symptoms. (d) RNA 5 may have been able to act in trans to down regulate genomic RNA synthesis and result in attenuation of symptoms. The final two mechanisms may be possible if the role of RNA in CMV was to reduce viral symptoms. However, the last two examples are also the least likely to be mechanisms by which plants could resist virus as RNA 5 does not appear to encode a protein and the 3' end of CMV RNAs have not been shown to function in trans to affect the life cycle of CMV.

The simplest explanation for lack of resistance in these transgenic plants is that expressing RNA 5 does not protect plants from CMV. However, it may have been technical problems which resulted in the lack of resistance, for instance the level of RNA transcription in the plants may have been too low to compete with the genomic RNAs or to activate the putative RNA processing mechanism thought to be responsible for RNA-mediated resistance. This may be overcome by screening more lines to find transgenic lines which expressed RNA 5 at levels which are high enough to reach the threshold required to activate a putative RNA processing mechanism or high enough to allow introduced viral RNA to cause a recovery phenotype (Smith et al. 1994). These plants would not be identified by measuring steady state transgene mRNA levels due to the RNA processing mechanism causing a reduction in these and could only be identified by virus challenge. Plants with high steady state mRNA levels, however, may be able to inhibit viral RNA replication through the mechanism (a) and could be selected using northern blot hybridisation analysis. To determine which of the these mechanisms was causing virus resistance, in vivo RNA 5 levels could be measured. An RNA-mediated resistance would result in a low level of RNA 5 due to the RNA processing mechanism and the competition mechanism would result in an increased level of RNA 5 (-) strand due to replication by the viral replicase.

To optimise mechanism (a), alterations to the RNA 5 expression cassettes could be made to remove extraneous 3' end sequences which are added during transcription in vivo. The addition of a 3' ribozyme sequence may remove these sequences allowing the transcription of an RNA molecule which resembles naturally-occurring RNA 5 molecules more closely. The presence of extraneous sequences has been shown to
reduce the infectivity of transcripts from cDNA clones (Ahlquist et al. 1987) and ribozymes have been used to generate infectious viral transcripts with defined 3' termini (Dzianott and Burjarski 1989). This indicates that extra sequence may result in reduced ability of the transcript to be recognised by the replicase. The catalytic activity of a sTRSV hammerhead ribozyme on the 3' termini of Q-CMV RNA 4A was assessed by Bathgate (1991) who found that this ribozyme was able to remove extraneous 3' sequences from RNA 4A transcripts.

To investigate the possibility of expressing RNA 5 molecules in vivo which have defined 3' ends, a construct was designed which contained a 3' cis-acting ribozyme which should cleave extraneous sequence from the 3' end of the RNA transcripts. This construct was prepared from a cDNA clone of RNA 4A prepared by Bathgate (1992) which contained a 3' ribozyme sequence from sTRSV. As RNA 5.2 is identical to the 3' end of RNA 4A, the region corresponding to RNA 5 and the ribozyme sequence was amplified using PCR with the primers PR5-5' (5'-GCGCTGCAGTAATACGACTCACTATAGTCCGAACGTTAAAC-3') which contains a Pst I site, a T7 RNA polymerase promoter and PR5s/c-3' (5'-GTTTTCCCAGTCACGAC-3') which is complementary to a region downstream of the ribozyme sequence. Amplified cDNA was ligated to pUC 19 to produce the plasmid illustrated in Fig. 5.4.

Fig. 5.4: CMV RNA 5 construct containing a 3' end ribozyme. The position of the T7 RNA polymerase promoter (T7 RNA pol) and direction of transcription, the RNA 5 cDNA (RNA 5) and the ribozyme sequence (s/c seq) are indicated. The position of
some restriction enzyme sites are also shown.

RNA 5 containing a 3' cis-acting ribozyme was transcribed from Bam HI-digested plasmids using the T7 RNA polymerase promoter contained in the 5' end primer used in the cloning protocol. Radioactively labelled nucleotides were incorporated into the transcript to increase the sensitivity of transcript detection. RNA transcripts were heated and rapidly cooled in the presence of MgCl₂ to maximise the possibility of self-cleavage (Prody et al. 1986). These reactions were electrophoresed on a polyacrylamide gel to separate cleaved RNA fragments and visualised by autoradiography (Fig. 5.5). Fragments corresponding to cleaved and uncleaved RNA transcripts were detected in the lane containing the self-cleavage sequence. This indicated that the 3' end ribozyme sequence was able to self-cleave in vitro, however, it is not possible to predict if the sequence will self-cleave in vivo.

Fig. 5.5: Self-cleavage analysis of RNA 5 construct containing a 3' end cis-acting ribozyme. RNA transcripts were heated to 80°C for 5 min. followed by rapid cooling on ice for 2 min. prior to gel loading. Reactions were electrophoresed on a 4% polyacrylamide gel containing 7M urea and autoradiographed. The lane labelled s/c contains RNA 5 with the 3' end ribozyme. The lane labelled M contains radiolabelled Eco RI digested SPP 1 markers.

Another way to increase the ability of RNA 5 molecules to compete with the genomic RNAs would be to add a CMV-specific promoter to the 5' end to increase RNA 5 levels.
upon infection with the virus. The RNA 5 (-) strand synthesised by the viral replicase may not be able to act as a template in further cycles of RNA replication due to the absence of a signal to direct (+) strand synthesis such as those contained at the 5' end of the genomic RNAs. The addition of the 5' end signal for (+) strand synthesis (derived from RNA 2) to the 5' end of RNA 5 would result in the production of a molecule which may be able to autonomously replicate as shown in Fig. 5.6. This would be similar to the strategy used to reduce BMV RNA replication in barley protoplasts by artificial DI RNAs (Marsh et al. 1991). This may increase the amount of RNA 5 upon infection with CMV which would further limit the amount of replicase available for genomic RNA replication and may cause a reduction in symptoms.

Fig. 5.6: Amplification of RNA 5 using a 5' end promoter. With native RNA 5, the viral replicase would recognise the signal found at the 3' terminal region and produce (-) strand RNA 5. This (-) strand may not be able to act as a template in further rounds of RNA replication. However, the RNA 5 (-) strand which contained the RNA 2 5' end sequence would be able to act as a template for further (+) strand synthesis.

Boccard and Baulcombe (1993) showed that only 92 nt at the 5' end of CMV RNA 3 were required for efficient accumulation. It would seem likely that a region of similar
size is sufficient in RNA 2 to allow binding of the replicase. To ensure that all sequences required for replicase recognition were obtained, the length of the RNA 2 5' terminal sequence used as a 5' promoter would be 300 nt. The ability of the 5' end of RNA 2 to direct (+) strand transcription could be assessed in an \textit{in vivo} system by expressing the (-) strand of the region upstream of the (-) strand of a reporter gene (\textit{gus}) \textit{in vivo}. This would involve transforming CMV-infected plant tissue with a construct which contained an RNA transcription promoter upstream of the RNA 2 and \textit{gus} gene (-) strand sequences. This would result in the production of the first RNA transcript shown in Fig. 5.7 by the plant RNA transcription mechanisms. If the RNA 2 sequence is capable of directing (+) strand synthesis, the (+) strand of the reporter gene would be synthesised by the viral replicase in the CMV infected tissue and the gene product would be translated by the plant translation mechanisms (Fig.5.7). The gene product could then be detected by histochemical staining.

Fig. 5.7: (+) strand synthesis directed by RNA 2 5' terminal sequence. The viral replicase would be directed to synthesise (+) strand RNA by signals in the RNA 2 sequence resulting in the production of a (+) strand copy of the \textit{gus} gene. This gene would then be translated by the host cell.

Although no plants were found to be resistant to CMV in this study, it is still possible that RNA 5 may be useful in engineering virus resistance with the improvements
discussed in this section. If this is possible, then such a strategy could be used to reduce the serious damage CMV causes to the Australian lupin industry. The successful use of 3' coterminal RNA molecules in resistance strategies may also provide an approach to engineering virus-resistant plants which, in a number of ways, is better than many of the strategies currently available. These advantages would include a broad resistance due to the similarity between a number of virus 3' termini, the ability for the resistance gene to be transmitted with the virus and the ability to apply this technique to almost any virus.
Chapter 6: Conclusions and Future Directions

RNAs 4A and 5 are novel components of CMV and are the least characterised of the CMV RNAs. They are unlike any other plant virus RNAs characterised to date. A complete understanding of the structure and distribution of these RNAs is required to determined their function and method of generation. Further knowledge of the CMV life cycle gained by the study of these RNAs may aid in the design of transgenic plants that are resistant to CMV. This dissertation describes the characterisation of the Q-CMV RNA 5 population, surveys a number of cucumovirus strains for the presence of RNA 4A and 5 and explores the possibility of designing a strategy for using RNA 5 to genetically engineer virus resistance plants. The major findings of the project are summarised below.

1. The Q-CMV RNA 5 population consists of two sequence variants derived from the 307 nt and 304 nt 3' termini of RNAs 2 and 3, respectively. RNA 5 is probably generated as a subgenomic RNA but does not appear to encode a protein.

2. RNA 5 was shown to be present in all of the subgroup II CMV strains that were analysed, one of the two subgroup I strains and a TAV strain. All CMV subgroup II RNA 5 populations consisted of at least two different sequence variants. In contrast, RNA 4A was found in all CMV strains analysed as well as the TAV strain.

3. A number of transgenic plants were constructed to express CMV RNA 5 and tested for CMV resistance. Of the eight lines of transgenic plants that were analysed, none appeared to be resistant to Q-CMV. Two plausible reason why these transgenic plants did not resist infection would be: i) the inability of RNA 5 to interfere with the replication of CMV or ii) the expression of RNA 5 may have been too low to provide resistance via an RNA-mediated resistance mechanism.

6.1. Characterisation of Q-CMV RNA 5 Indicates RNA 5 is Derived from RNA 2 and 3 as a Subgenomic RNA

The characterisation of RNA 5 supports the proposal that RNA 5 is not a degradation product but is produced from subgenomic promoters on full length (-) strands of RNA 2 and 3. If RNA 5 was produced by degradation of genomic RNAs, then it would be expected that RNA 1-like RNA 5 molecules would have been present in the same proportion as other molecule types. The presence of products of precise size after primer extension reactions on encapsidated RNA preparations indicated that the RNA 5 molecules contained distinct 5' ends. It was demonstrated earlier (Blanchard 1991) that
the RNA molecules all contained identical 5' ends by direct RNA sequencing. If these molecules were derived by breakdown of genomic RNA, they would be expected to be of varying sizes with different 5' ends. If RNA 5 was produced by specific cleavage, of a larger molecule, products corresponding to the upstream sequence of parent molecules should be present. Gordon and Symons (1983) did not find any evidence for such products. The explanation of the absence of the 5' terminal cleavage product being due to lack of secondary structure is unlikely as regions upstream of the RNA 5 start point would form secondary structures to inhibit reannealing of (+) and (-) strands and, hence, be as resistant to degradation as the 3' end. The nature of the RNA 5 population indicates that the most likely explanation for the origin of RNA 5 is that it is synthesised as a subgenomic RNA.

The characterisation of Q-CMV RNA 5 has raised a number of interesting questions. One of which is: what is the possible function of RNA 5 in the life cycle of cucumoviruses? Based on the fact that RNA 5 contains sequences and structures responsible for replicase recognition (Bocard and Baulcombe 1993) and that the only strain examined which did not contain RNA 5 was the most virulent, it is conceivable that RNA 5 may play a role in symptom attenuation and in maintaining chronic infections. For this reason it was proposed that RNA 5 may be able to function to attenuate viral symptoms when expressed in transgenic plants either by competing with the viral genomic RNAs for replication factors or by activating an RNA-mediated resistance mechanism which would result in viral RNAs being targeted for degradation (Smith et al. 1994). The fact that no resistance was observed does not discount the possibility that RNA 5 may function in symptom attenuation as the lack of resistance in transgenic plants expressing a gene for RNA 5 may have been due to technical problems. Although the function of symptom attenuation seems to be the most likely role for RNA 5, there is, as yet, no direct evidence to support this view.

Although it seems unlikely that RNA 5 encodes a protein, this could be assessed by expressing fusion proteins derived from RNA 5 in E. coli., raising antibodies to the proteins produced and using these antibodies to detect putative proteins in CMV-infected plants. If a protein was detected in vivo, the function of this protein could be determined by carrying out site directed mutagenesis to remove the ability of the RNA to translate the protein and comparing the mutated virus with wild-type virus.

Another interesting feature of RNA 5 that was established by its characterisation is that it is derived from RNAs 2 and 3 but not RNA 1. A possible explanation for this is that RNA 5 can only be synthesised from other subgenomic RNAs (that is, RNAs 5.2 and 5.3 are derived from RNAs 4A and 4, respectively). This could be assessed by
removing sequences corresponding to the subgenomic promoter on RNA 2 which is responsible for the synthesis of RNA 4A to stop the synthesis of this RNA. If the synthesis of RNA 5.2 ceased as well, it could be concluded that RNA 5.2 is only produced from RNA 4A and not RNA 2.

Further studies on the origin of RNA 5 should include the characterisation of RNA 5 from virus strains other than Q-CMV. The occurrence of RNA 5 in L-CMV indicates that there may be subgenomic promoters on L-CMV genomic RNAs. By characterising the L-CMV RNA 5 population and sequencing the 3' ends of the genomic RNAs encompassing the RNA 5 sequence, the genomic RNAs which contained these promoters would be identified as well as the approximate position of the putative subgenomic RNA promoters. Comparison of the putative promoter sequences to the region thought to form part of the subgenomic promoter responsible for RNA 5 synthesis in TAV and subgroup II CMV strains may identify common regions or secondary structures essential for the synthesis of RNA 5. Comparison of the sequence around the start point of L-CMV RNA 5 on the genomic RNAs to the corresponding region on Ny-CMV RNAs may reveal the presence of extra sequences on L-CMV RNAs which are responsible for RNA 5 synthesis.

The occurrence of RNA 5 in V-TAV demonstrated that RNA 5 is not limited to CMV. Further analysis of this population would determine if RNA 5 is produced in the same way in all cucumoviruses. This would be assessed by determining which genomic RNAs TAV RNA 5 is derived from by using the sequence specific primer extension analysis method developed in Chapter 3 or by cloning and sequencing these RNAs as was done with Q-CMV RNA 5. As PSV RNAs lack the sequence found in TAV and CMV subgroup II RNAs thought to be involved with RNA 5 synthesis, it seems unlikely that RNA 5 would occur in this virus. If this is correct, then primer extension on BDiMV RNA would yield RNA 5 extension products only from RNA 2-like RNA 5 as RNA 3 is derived from PSV. The absence of RNA 5 extension products from PSV and RNA 3 derived RNA 5 BDiMV would further support the proposed involvement of the 5' terminal sequence of RNA 5 in RNA 5 synthesis.

Primer extension analysis on the newly characterised cucumovirus, BDiMV, would provide an in

6.2. RNA 4A and 5 Occur in Most Cucumoviruses

The detection of RNA 4A and 5 in TAV and both subgroups of CMV indicates that the presence of these RNAs is probably a generalised feature of cucumoviruses and that
these RNAs may have a specific function and method of generation. This study has revealed a number of significant differences between RNAs 4A and RNA 5. The two most obvious differences are that RNA 5 consists of two types of molecules, whereas RNA 4A consists of a homogeneous population. While RNA 4A is probably accumulated by all cucumoviruses, RNA 5 does not appear to be produced by all subgroup I CMV strains. Also, RNA 4A is known to encode a protein in vivo (Ding et al. 1994), whereas RNA 5 is probably not translated.

Strong evidence already exists which indicates that RNA 4A is produced as a subgenomic RNA (Ding et al. 1994). This evidence includes the identification of a putative subgenomic RNA promoter which is conserved in cucumovirus strains, identification of an ORF which is conserved in all cucumoviruses and the detection of the predicted protein product in vivo. The prediction by Ding et al. (1994) that RNA 4A would be present in all cucumovirus strains, was shown directly in this dissertation. Based on the presence of RNA 4A in all cucumoviruses analysed and the presence of the putative RNA 4A subgenomic promoter element and 2b ORF on PSV RNA 2 (Ding et al. 1994) it is likely that RNA 4A is present in PSV. Further studies on the extent of RNA 4A in cucumoviruses would confirm the presence of RNA 4A in PSV by primer extension analysis.

The presence of RNA 4A in all CMV and TAV strains that have been examined indicates the function of RNA 4A is important in all cucumoviruses. However, recent studies (Ding et al. 1995b) have indicated the 2b protein is not essential for either systemic viral infection of N. glutinosa or replication of the satellite. This may mean that the presence of RNA 4A in all cucumoviruses may be related to its appearance at a point early in the evolution of cucumoviruses (and possibly may have played a role in cucumovirus evolution) rather than having an essential role in the life cycle.

Previous evidence in the literature indicated that RNA 5 may be produced after non-specific degradation of genomic RNAs. For example, Peden and Symons (1973) observed an increase in RNA 5 and a decrease in RNA 4 after freezing and thawing cycles. This indicated that RNA 5 was produced as a result of RNA 4 degradation. More recent studies, however, indicate that RNA 5 may be produced as a subgenomic RNA as promoter elements were identified upstream of the proposed initiation site of RNA 5 on RNA 3 (Peter Palukaitis, personal communication). The absence of RNA 5 in Ny-CMV is further evidence that RNA 5 is not a degradation or cleavage product which is only present due to stabilisation by secondary structure. If this was the case, then similar molecules would be expected in Ny-CMV as it is thought that the secondary structure at the 3’ end of all CMV RNAs is similar (Palukaitis et al. 1992).
The idea that RNA 5 is produced as a subgenomic RNA is further supported by the identification of sequence similarity between proposed subgenomic promoters and the sequence around the RNA 5 start points. Extensive sequence similarity at the 5' end of RNA 5 on subgroup II CMV and TAV RNAs and the absence of the conserved region in Ny-CMV is consistent with a role as subgenomic promoter for RNA 5. It is likely that L-CMV has a sequence in addition to Ny-CMV which has a similar role as the sequence conserved in TAV and CMV subgroup II strains that appears to be involved in RNA 5 production. There are two possible ways this sequence may be involved with the production of subgroup II RNA 5. Firstly, it may be involved with the formation of a structure which results in the specific cleavage on the 5' end of the sequence (which is unlikely, as the nature of the RNA 5 population suggested it was not produced in this manner). Alternatively, this sequence may form part of a subgenomic promoter element involved with the synthesis of RNA 5. The presence of the 5' terminal sequence of TAV and CMV RNA 5 in a putative subgenomic promoter of another (+) strand RNA virus (BNYVV) further supports the proposal that RNA 5 is synthesised as a subgenomic RNA.

While RNA 5 is not found in all subgroup I CMV strains, its presence in all subgroup II strains that were examined may mean that it has an important function in strains belonging to this subgroup. Infectivity studies may demonstrate what the function of RNA 5 is. This could be done by inoculating plants with a cucumovirus strain which does not contain RNA 5 (either due to a natural absence or by mutating sequences responsible for its synthesis) and comparing it with plants inoculated with the same strain which contains RNA 5 (either due to a natural occurrence or by coinoculation with RNA 5 transcribed a cDNA clone). Replication, virus movement and symptoms of virus could be compared to determine if they are affected by the presence of RNA 5.

The presence of RNA 5 in TAV and subgroup II CMV strains, and the absence of RNA 5 in a subgroup I strain may cast doubt on the current classification of cucumoviruses. Based on comparisons of the entire nucleotide and amino acid sequences, CMV subgroup I and II strains are more similar to each other than they are to TAV (McGarvey et al. 1995). However, it has been suggested that TAV, Q-CMV (subgroup II) and Fny-CMV (subgroup I) may be considered as representing three equivalent subgroups of a taxonomic entity based on comparisons between protein and RNA sequences (Bernal et al. 1991). The presence of RNA 5 in V-TAV and its absence in Ny-CMV suggests, in one way, that TAV is more closely related to Q-CMV than to Ny-CMV. This is also evident when the RNA 5 sequences of V-TAV and Q-CMV, and the corresponding Fny-CMV 3' terminal sequences are aligned. The percent similarity between the Q-CMV
and Fny-CMV sequences in this region is 69% whereas the percent similarity between Q-CMV and V-TAV sequences is 76%. A similar observation was made by Bernal et al. (1991) who reported that sequence similarities for the 3' non-coding regions in RNA 1 upstream of the 3' terminal 149 nt are much higher between TAV and Q-CMV (71%) than between Q-CMV and Fny-CMV (37%). The high sequence similarity between Q-CMV and V-TAV in this region is due to the presence of highly-conserved sequence motifs found at the 5' ends of the 3' conserved region which corresponds to RNA 5. Analysis of more TAV strains containing RNA 5 and more CMV subgroup I strains which do not contain RNA 5 would provide further evidence to indicate that CMV subgroup II strains are in some ways more similar to TAV than they are to CMV subgroup I strains.

Based on the presence of RNA 5 in L-CMV and its absence in Ny-CMV it is possible that further analysis of L-CMV may reveal that it is quite dissimilar to other CMV strains which have been characterised and may represent a third CMV subgroup strain. The presence of a third group is supported by the inability of Wahyuni et al. (1992) to assign the Ywa-CMV strain to either one of the two subgroups. As there is a distinct variation between the CMV subgroup I and II extension products produced with primer extension, this technique may be useful for accurately and easily determining the subgroup of a CMV strain.

6.3. Engineering Virus Resistance with RNA 5

In designing strategies to genetically engineer virus-resistant plants using a PDR strategy, it is important to have a complete knowledge of the structure and replication of the virus from which the gene is derived. Characterisation of Q-CMV RNA 5 and the identification of a possible function and mechanism of RNA 5 synthesis has provided the necessary knowledge to design a strategy for using RNA 5 to engineer virus resistance. The proposed mechanism of symptom attenuation (based on the presence of signals thought to be involved with replicase recognition on RNA 5 and the absence of the RNA 5 in the virulent Ny-CMV strain) indicates RNA 5 may be useful in a PDR strategy. However, as the plants constructed in Chapter 5 were not resistant to CMV infection, this strategy may require a design more complex than simply expressing RNA 5 in transgenic plants. A successful strategy would rely on the expression of RNA 5 at high levels and in a form that could be recognised by viral replicase. In doing so it is conceivable that a strategy using RNA 5 may be successful in reducing the losses to agricultural sectors such as the Australian lupin industry, due to damage caused to commercially important crop species.
Literature Cited


resistance against cymbidium ringspot tombusvirus in transgenic plants. *Virology*, 193, 313-318.


Virology, 172, 415-427.


Literature Cited

1777-1787.


*White* *Virology*, 207, 334-337.


Appendix 1: Publications Arising from Thesis


