

FURTHER STUDIES ON  
THE STRUCTURE AND FUNCTION OF  
THE CUCUMBER MOSAIC VIRUS GENOME



**FURTHER STUDIES ON  
THE STRUCTURE AND FUNCTION OF  
THE CUCUMBER MOSAIC VIRUS GENOME**

A thesis submitted to the University of Adelaide, South Australia,  
for the degree of Doctor of Philosophy,

by

**Rhys Harold Verdon George Williams, B.Sc. (Hons.)**

Adelaide University Centre for Gene Technology,  
Department of Biochemistry  
University of Adelaide,  
Adelaide, South Australia.

February, 1988.

12 July 88

## TABLE OF CONTENTS.

<b>SUMMARY</b>	[1]
<b>STATEMENT</b>	[3]
<b>ACKNOWLEDGEMENTS</b>	[4]
<b>ABBREVIATIONS</b>	[5]
<b><u>CHAPTER 1. General Introduction.</u></b>	
1. Plant RNA viruses.	2
2. Classification and nomenclature of plant viruses.	2
3. The Tricornaviridae.	3
Physical and biological properties.	4
Genome structure.	4
4. Cucumber mosaic virus.	5
Biological properties.	6
Virion structure.	6
RNA components.	6
Satellite RNAs.	7
Proteins encoded by viral RNAs.	7
Replication of viral RNA.	8
5. Aims and scope of this work.	8
<b><u>CHAPTER 2. Nucleotide Sequence of RNAs 1 and 2 of Cucumber Mosaic Virus (Q-strain).</u></b>	
<b>Introduction.</b>	11
<b>Materials and Methods.</b>	13
Materials.	13
Methods.	
1. Virus and viral RNA.	13
2. Purification of CMV RNAs by spermine precipitation.	14
3. Preparation of cDNA clones from CMV RNA 1 and 2 (Summary).	14
4. Preparation of single-stranded M13 DNA.	14
5. Dideoxy sequencing of M13 clones.	15
6. Preparation of DNA primers from M13 clones.	15
7. Annealing of primers to CMV RNA for sequencing.	16
8. Dideoxy sequencing of RNA	16
9. Dideoxy sequencing of M13 clones and RNA using dITP and reverse transcriptase.	17
10. Decapping of CMV RNAs.	17
11. Removal of ribonuclease contamination from calf intestinal phosphatase.	18

12. 5'-Labelling of CMV RNAs with $^{32}\text{P}$ .	18
13. Sequencing of 5'-terminus of RNA.	18
14. Computer manipulation and analysis of sequence data.	19

**Results** 20

1. Strategy for sequencing CMV RNAs 1 and 2.	20
2. Open reading frames in RNAs 1 and 2.	22
3. Translation initiation signals in CMV RNAs.	22
4. Codon usage in the P1 and P2 coding regions.	23
5. Homology between CMV RNAs.	23
6. The 3' Non-coding regions.	24
7. The 5' Non-coding regions.	25
8. Structure conserved between CMV and BMV RNAs and CMV satellite RNA.	26

**Discussion.** 27

The CMV-encoded proteins.	27
Translation of CMV and other plant viral RNAs.	28
Structure and function of the 3' non-coding regions	30
Structure and function of the 5' non-coding regions.	32
Interaction between the CMV genome and CMV sat-RNA.	33
Summary: Similarities and differences between CMV RNAs.	34

**CHAPTER 3. Nucleotide and amino acid sequence homology between CMV and other viruses.**

**Introduction.** 37

**Methods.** 39

1. Source and nomenclature of sequences.	39
2. Dot matrix comparison of nucleotide and amino acid sequences.	39
3. Alignment of amino acid sequences.	40
3. Preparation of hydrophobicity plots.	40

**Results.** 41

1. Dot matrix comparison CMV RNAs 1 and 2 and their protein products with those of other viruses.	41
2. Alignment of homologous viral proteins.	42
3. Similarities in the properties of the conserved proteins.	43

**Discussion.** 44

Further homologies with the Tricornaviridae and TMV.	45
Homologies between other plant viral proteins.	46
Functional significance of homologous proteins in Tricornaviridae.	47
Homology between RNA polymerases from plant, animal and bacterial viruses.	47
Implications for the function of CMV-encoded proteins.	48

**CHAPTER 4. Characterization of CMV-Induced RNA-Dependent RNA Polymerase.**

<b>Introduction.</b>	51
1. Replication of RNA virus genomes.	51
2. Bacteriophage Q $\beta$ RNA replicase	51
3. Animal viral RNA replicases	52
4. Plant viral RNA replicases	52
Brome mosaic virus RNA replicase	52
Alfalfa mosaic virus RNA replicase	53
Tobacco mosaic virus RNA replicase	53
Turnip yellow mosaic virus RNA replicase	54
Cowpea mosaic virus RNA replicase	54
5. CMV-induced RNA-dependent RNA polymerase.	55
The particulate replication complex	56
The role of viral- and host-coded proteins.	57
Re-investigation of the Mr 100,000 protein.	58
Problems with previous approaches.	60
6. Experimental approach.	60
<b>Materials and Methods.</b>	62
Materials.	62
Methods.	63
1. Virus and plants.	63
2. Extraction of CMV-induced RNA-dependent RNA polymerase.	
Method A	64
Method B	65
Method B: Large-scale preparations	66
3. Storage of PEG-Enzyme.	67
4. Assay of RNA-dependent RNA polymerase activity.	67
5. Chromatography trials of Method A PEG-Enzyme.	68
6. Preparation of Method B PEG-Enzyme for column chromatography.	69
7. Hydrophobic interaction chromatography of PEG-Enzyme.	69
8. Chromatography of PEG-Enzyme on FPLC ion-exchange media.	70
9. Chromatography of PEG-Enzyme on BLUE-TRISACRYL M.	70
10. Measurement of protein concentration.	71
<b>Results.</b>	72
1. Method for extraction of RdRPase from CMV-infected cucumber seedlings.	72
Extraction using Method A and attempted column purification of Method A AmSO <sub>4</sub> -SN.	72
Attempted column purification of Method A PEG-Enzyme.	73

Requirements for an alternative extraction method.	74
Time-course of RdRPase-induction in CMV-infected plants.	74
Comparison of Methods A and B.	75
2. Large-scale preparation of Method B PEG-Enzyme.	75
3. Stability of PEG-Enzyme.	76
Stability under storage.	76
Stability of undiluted PEG-Enzyme.	76
Stability of PEG-Enzyme diluted for chromatography.	77
4. Hydrophobic interaction chromatography of PEG-Enzyme.	77
5. FPLC ion-exchange chromatography of PEG-Enzyme.	78
Effect of temperature and buffer composition	79
Effect of salt gradient elution.	80
6. Chromatography of PEG-Enzyme on BLUE-TRISACRYL M.	81
<b>Discussion.</b>	82
Method for extraction of RdRPase.	82
Time-course of RdRPase activity.	82
Factors affecting enzyme stability.	83
Evidence for hydrophobic interactions.	86
Development of a chromatographic purification scheme.	87
Future prospects for purification of CMV-induced RdRPase.	90
 <b><u>CHAPTER 5. Concluding Discussion.</u></b>	 92
 <b><u>APPENDIX A. Groups and Families of Viruses Infecting Plants.</u></b>	 98
 <b><u>APPENDIX B. Publications.</u></b>	 101
 <b>REFERENCES.</b>	 103



## SUMMARY

The genome of cucumber mosaic virus (CMV) consists of three single-stranded (+) RNAs (RNAs 1, 2 and 3) which are all required for infection. A fourth sub-genomic RNA (RNA 4) encoding the viral coat protein, is produced during infection and encapsidated in virions.

In order to understand the structure of the CMV genome and the control of its expression, the complete nucleotide sequences of RNAs 1 and 2 of the Q strain of CMV were determined using dideoxy sequencing of RNA and cloned cDNA thus completing the primary structure of the CMV genome. RNA 1 is 3389 nucleotides long ( $M_r$   $1.15 \times 10^6$ ) and RNA 2 is 3035 nucleotides long ( $M_r$   $1.15 \times 10^6$ ). RNAs 1 and 2 each contain a single long open reading frame encoding proteins of  $M_r$  110,791 (P1) and  $M_r$  94,333 (P2), respectively.

The 3' non-coding regions of all three RNAs are extensively homologous for approximately 300 nucleotides from the 3' end. In addition to the tRNA-like structure proposed for the 3'-terminal 130 nucleotides of all CMV RNAs (Symons, *Nuc. Acids Res.* 7:825, 1979), the region of RNAs 1 and 2 130 to 277 nucleotides from the 3' end can potentially form two hairpin loops with high predicted stability. These structures cannot form in RNA 3, since they occur in that section of the 3' non-coding region which shows the greatest variation between RNAs 1, 2 and 3.

The 5' non-coding regions of RNAs 1 and 2 are also extensively homologous and contain a 12-nucleotide sequence which is conserved in a similar location in RNAs 1 and 2 of the closely related brome mosaic virus (BMV). This sequence was also found in the complementary strand of the satellite RNA (sat-RNA) associated with CMV. The region surrounding this conserved sequence in both CMV and BMV RNAs and in sat-RNA could be folded into a hairpin loop structure in which most of the conserved residues are located in the loop-out region. There are no apparent primary or secondary structural similarities between the 5' non-coding regions of all 3 CMV RNAs.

The implications of these 5'- and 3'-terminal conserved structural features for the function of the CMV genome and its interaction with satellite RNA are discussed.

Dot matrix analysis was used to reveal homologies between the nucleotide sequences of CMV RNAs 1 and 2, the corresponding RNAs of BMV and alfalfa mosaic virus (AMV) and the single genomic RNA of tobacco mosaic virus (TMV). Even stronger homologies were found when the translation products of these RNAs were compared using a similar analysis. CMV P1 was found to contain N- and C-terminal regions of approximately 300 amino acids which are homologous to the P1 proteins of BMV and AMV and the 126k protein of TMV. Similarly,

CMV P2 contains a central domain which is homologous to the P2 proteins of BMV and AMV and to the readthrough region of the TMV 183k protein. Homology is greatest between the proteins of CMV and BMV and less prominent, although still significant, with AMV and TMV. Hydrophobicity plots of P1 and P2 proteins were also very similar between CMV and BMV.

These results highlight similarities in the structure and probably also function of the P1 and P2 proteins from different viruses and provide evidence for an evolutionary relationship between the tripartite viruses and TMV. Furthermore, they concur with homologies between viral non-structural proteins reported by other authors, and with circumstantial evidence from studies of BMV and AMV, suggesting that the P2 protein of CMV and other tripartite viruses, is an RNA-dependent RNA polymerase (RdRPase), or the core subunit thereof, and that the P1 protein is also involved in replication (reviewed in Goldbach, *Ann. Rev. Phytopath.* 24:289, 1986).

Highly-purified preparations of CMV-induced RdRPase have been found to consist of a protein of  $M_r$  100,000 (100k) and trace amounts of 110k and 35k proteins (Kumarasamy and Symons, *Virology* 96:622, 1979; Gill *et al*, *Virology* 113:1, 1981; Gordon *et al*, *Virology* 123:284, 1982). Peptide mapping and *in vitro* translation studies have suggested that these proteins are not viral-coded (Gordon *et al*, *op. cit.*). However, in view of the strong implications of the amino acid sequence homology studies and the somewhat ambiguous nature of previous results (Gordon *et al*, *op. cit.*), further work was undertaken with the aim of purifying the 100k protein from CMV-infected cucumbers in sufficient amounts to enable amino acid sequencing. This required re-investigation and optimization of the procedures used for extraction and storage of RdRPase and development of an efficient chromatographic purification scheme suitable for large-scale application.

High yields of RdRPase were obtained by extracting cotyledons and leaves at 6-9 days post-inoculation using the  $MgSO_4$ -solubilization method of Gill *et al* (*op. cit.*). The resultant PEG-precipitated enzyme (PEG-Enzyme) was stable under long term storage at  $-80^\circ C$  and for short periods at  $4^\circ C$  or room temperature even in the absence of glycerol or reducing agents. However, chromatography of PEG-Enzyme on a wide range of media resulted in loss of activity. Further investigation of the behaviour of PEG-Enzyme on FPLC ion-exchange columns, suggested that losses of activity during chromatography were at least partly due to unfavourable hydrophobic interactions between RdRPase and the column.

Numerous problems encountered during these studies did not permit the large-scale purification of 100k protein for amino acid sequence analysis. However, initial problems of loss of activity and lack of reproducibility were largely overcome and suitable large-scale purification procedures were found which could make this project feasible in the future.



**STATEMENT.**

This thesis contains no material which has been accepted for the award of any other degree or diploma in any University. To the best of my knowledge and belief, it contains no material previously published or written by any other person, except where due reference is made in the text.

I hereby consent to this thesis being made available for photocopying and loan if accepted for the award of the degree.

Rhys H.V.G. Williams.

## ACKNOWLEDGEMENTS.

I would like to thank Professor W.H. Elliott for the permission to undertake my Ph.D. studies in the Department of Biochemistry, University of Adelaide and Professor R.H. Symons for his supervision and advice during the course of my work and in the preparation of the manuscript.

I also wish to thank the members of our research group who have contributed helpful discussions over the past years – Drs. Karl Gordon, Jim Haseloff, Paul Keese, Rhett Swanson and Chris Davies, and Professor E.M.J. Jaspars. I am especially grateful to Dr Ali Rezaian, my collaborator in the CMV sequencing work, for his advice, encouragement and friendship, and for his help in preparing this manuscript.

I am indebted to Dr Alex Reisner, CSIRO Division of Molecular Biology, North Ryde for performing dot matrix and hydrophobicity analysis and for his advice on computing. I would also like to thank Drs Ian Dodd and A.V. Sivaprasad for supplying valuable programs and for answering a constant stream of questions on computing matters.

I am grateful to Dr R.I.B. Francki, Department of Plant Pathology, WAITE Institute, University of Adelaide for the use of glasshouse facilities and to Dennis Talfourd for his advice and help in the glasshouse.

I would like to express my thanks to the following people within our department who have generously given their help and advice:

Sharon Freund, Jenny Cassady and Tammy Edmonds for technical assistance, photography and artwork.

Dr Steve Wilton and Mr Peter Gilhaus from BRESATEC for their continual advice and generous supply of chromatographic media and other 'goodies';

Jenny Brazier for allowing me to use the BRESATEC FPLC system;

Stan Bastiras and Dr Rod Nicholls for advice on FPLC;

Drs. Bill Kalionis, Barry Powell and Fil Lim for valuable discussions.

I am especially grateful to Tammy Edmonds for drawing and preparing the diagrams so well and so quickly, and to Litsa and John Karageorgos for use of their ● Macintosh.

I would like to express my gratitude to my parents for their support over so many years of study, and to those people whose friendship and support have helped "keep me going", especially, Pat, the guys from "Glass", Adrienne, Lisa and Kees.

Finally, I owe a very special thankyou to my wife, Diana, whose patience understanding, encouragement and constant support have made it possible for me to finish this work.

**ABBREVIATIONS.**

AmSO <sub>4</sub> -SN	ammonium sulphate supernatant
BSA	bovine serum albumin
cDNA	complementary DNA
Ci	Curies (1 Ci = 37 Gigabecquerel)
CP	coat protein
DEP	diethyl pyrocarbonate
DTT	dithiothreitol
EDTA	ethylenediaminetetraacetic acid
FPLC	Fast Protein Liquid Chromatography
2-ME	2-mercaptoethanol
PEG	polyethylene glycol
PF	particulate fraction
p.i.	post-inoculation
POPOP	1,4-bis[2-(5-phenyloxazolyl)]benzene
PPO	2,5-diphenyloxazole
RdRPase	RNA-dependent RNA polymerase
sat-RNA	satellite-RNA
SDS	sodium dodecyl sulphate
Tris	tris (hydroxymethyl) aminomethane
NTP	nucleoside-5'-triphosphate
AMV	alfalfa mosaic virus
BBMV	broad bean mottle virus
BMV	brome mosaic virus
BNYVV	beet necrotic yellow vein virus
BSMV	barley stripe mosaic virus
CarMV	carnation mottle virus
CMV	cucumber mosaic virus
CCMV	cowpea chlorotic mottle virus
CPMV	cowpea mosaic virus
PSV	peanut stunt virus
TAV	tomato aspermy virus
TEV	tobacco etch virus
TRV	tobacco rattle virus
TSV	tobacco streak virus
TMV	tobacco mosaic virus
TVMV	tobacco vein mottling virus
TYMV	turnip yellow mosaic virus

**CHAPTER ONE.**

**GENERAL INTRODUCTION.**

## **1. Plant RNA viruses.**

Viruses provide a valuable system for the study of cellular function at the molecular level. They have a relatively simple structure consisting of nucleic acid surrounded by a protective coat of protein and sometimes also lipid. Upon infection of a host cell, they are able to use their limited genetic information to adapt the cell's biological machinery for the synthesis of viral nucleic acid and protein. The study of viral infections therefore provides information not only on the organization and replication of the virus, but also on molecular processes occurring in the healthy cell.

The exact replication strategy employed by a virus depends primarily on the nature and structure of its nucleic acid, which may be single-stranded (ss) or double-stranded (ds), and either DNA or RNA. Most viruses infecting plants have ss-RNA genomes of the (+) or messenger sense. Furthermore, many of these viruses are multipartite in that their genomes are split into two or three separately-encapsidated RNA fragments which are all required for viral infection. The study of multipartite viral genomes therefore provides additional information on the co-ordination of processes such as transcription, translation and encapsidation of viral RNA, since signals controlling these processes must be conserved on all genome segments.

The work presented in this thesis concerns the genome structure and function of cucumber mosaic virus (CMV) which is a well-characterized example of a multipartite (+)RNA virus.

## **2. Classification and nomenclature of plant viruses.**

An understanding of the system for classification and nomenclature of viruses is essential not only for diagnosis, but also for research, especially when the results, like those presented in this thesis, bear upon the relationships between viruses.

The numerous early attempts at virus classification (reviewed by Francki, 1981 and Matthews, 1983) were hampered by the lack of reliable and useful parameters by which viruses could be classified. The past two decades have seen

the acquisition of vast amounts of information on the physical, chemical and biological properties of viruses and the evolution of a widely accepted taxonomy under the guidance of the International Committee on Taxonomy of Viruses (ICTV).

ICTV classification of vertebrate, invertebrate and bacterial viruses follows the classical system of families and genera. Plant viruses, however, are divided into 24 loosely-defined groups and 2 families—the Reoviridae and Rhabdoviridae—which also contain members infecting organisms other than plants (Matthews, 1982). A list of these plant virus groups and families together with some basic information about their members appears in Appendix A.

### **3. The Tricornaviridae.**

Within the 24 groups of plant viruses are four groups—Bromovirus, Cucumovirus, Ilarvirus and Alfalfa mosaic virus (Appendix A and Table 1.1)—which share several important features. They all contain tripartite ss(+)RNA genomes encapsidated in small polyhedral or bacilliform particles. These characteristics distinguish them from all other plant viruses and it has therefore been proposed that they be grouped as a virus family with the name Tricornaviridae (van Vloten-Doting *et al*, 1981). This name summarizes the features of these viruses— "tri" indicating the tripartite nature of the genome, "co" referring to the necessity of cooperation between the components to initiate infection, and "rna" indicating that the genome is composed of RNA. The ending "-viridae" is that approved by the ICTV for virus families (Matthews, 1982).

The current state of knowledge about the Tricornaviridae—their protein and nucleic acid structure, genetics, symptomatology, cytopathology, serology and epidemiology—has been reviewed extensively in a recent volume (Francki, 1985a) and the reader is referred to this publication and to Francki *et al*, (1985) for detailed information. It will suffice, in the following sections, to summarize the important general features of the Tricornaviridae and to specifically describe the properties of CMV.

**TABLE 1.1 : Properties of the type members of the groups of Tricornaviridae.**

The major properties of cucumber mosaic virus (CMV), brome mosaic virus (BMV), tobacco streak virus (TSV) and alfalfa mosaic virus (AMV) are presented as representative of the virus group for which they are the type members. Data are from Francki (1985a).

	<b>Group:</b> Cucumovirus	Bromovirus	Iarvirus	Alfalfa mosaic
	<b>Type member:</b> cucumber mosaic (CMV)	brome mosaic (BMV)	tobacco streak (TSV)	alfalfa mosaic (AMV)
<b>Virus Particles morphology</b>	icosahedral	icosahedral	spheroidal	bacilliform
approximate dimensions (nm)	29	26	27-35	30-56 x 18
<b>Host range</b>	very broad	narrow	broad	very broad
<b>Vectors</b>	aphids	beetles	seed & pollen	aphids
<b>Genome RNAs 3' end</b>				
tRNA-like structure	Yes	Yes	No	No
Aminoacylatable	Yes	Yes	No*	No
<b>Coat protein dependence</b>	independent	independent	dependent	dependent

\* It is not known for certain that the RNAs of Iarviruses cannot be aminoacylated, since they have not been tested. It is assumed that they would behave like the AMV RNAs since they too lack tRNA-like structures and are dependent on coat protein for initiation of replication.

Table 1.1 summarizes some of the properties of the type members of the four groups of Tricornaviridae and highlights the similarities and differences between these groups.

### **Physical and biological properties.**

Cucumovirus and Bromovirus particles are of similar size and are both composed of 180 identical protein molecules arranged with icosahedral symmetry. In contrast, preparations of TSV (Iilarviruses) and AMV contain particles of different sizes depending on their nucleic acid content. Those of AMV are all 18 nm wide but vary in length, producing characteristic bacilliform particles. The Iilarviruses have not been studied as extensively but are known to have roughly spheroidal particles of about 30 nm diameter.

All of the Tricornaviridae can be transmitted mechanically, but they differ with respect to their mode of vector transmission and their host ranges (Table 1.1). Both CMV and AMV are transmitted non-persistently by aphids and can infect a very wide range of hosts (see also below). The host range of TSV is also wide, although Iilarviruses are probably transmitted by seed and pollen rather than by vectors. In contrast, the host range of the Bromoviruses is very restricted and they are transmitted mainly by beetles.

### **Genome structure.**

The genome structure of the Tricornaviridae (reviewed by Symons, 1985) is presented in generalized form in Figure 1.1. In addition to the three genomic RNAs, designated RNAs 1 to 3 in order of decreasing molecular weight, these viruses contain a fourth small RNA (RNA 4) which is a subgenomic mRNA for the viral coat protein. A substantially homologous region is found at the 3'-terminal end of each of the RNAs of each virus (Figure 1.1) and all RNAs have m<sup>7</sup>G caps.

*In vitro* translation and sequencing studies of several Tricornaviridae have revealed that RNAs 1 and 2 are monocistronic, encoding the P1 and P2 proteins respectively and that RNA 3 is dicistronic. The 5' cistron encodes the 3a (or P3) protein and the 3' cistron encodes the coat protein (CP) which is expressed only from RNA 4 (Figure 1.1).



**FIGURE 1.1: Generalized genome structure of the Tricornaviridae.**

RNAs 1 to 4 each contain a single cistron (open boxes) encoding the P1, P2, and 3a (or P3) proteins and coat protein (CP), respectively. The CP cistron on RNA 3 is not translated *in vivo*. The numbers inside each box refer to the approximate molecular weight (Mr) of each protein. The closed boxes represent regions of sequence homology (140 to 320 residues) which are conserved between the four RNAs of each virus.



The four groups of Tricornaviridae differ in the structure of the 3' termini of their RNAs (Table 1.1). All Cucumo- and Bromoviral RNAs so far sequenced have characteristic 3'-terminal tRNA-like structures which are strongly conserved, not only within, but between viruses, suggesting that they have a significant biological role (see Symons, 1985). In addition these RNAs can serve as templates for aminoacyl tRNA synthetases. In contrast, the 3' termini of Ilarvirus and AMV RNAs do not possess tRNA-like structures and are not aminoacylatable.

Infection by Ilarviruses or AMV requires, in addition to the three genomic RNAs, a small amount of the viral coat protein or its messenger, RNA 4. These viruses are therefore said to be "coat protein-dependent". Furthermore, AMV and TSV RNAs have specific sites near their 3'-termini with high affinity for binding of viral coat protein (reviewed in Jaspars, 1985), which has been proposed as an important step in the initiation of viral RNA replication. This suggests differences in the mechanism of initiation between these viruses and Bromo- and Cucumoviruses which have no coat protein binding requirement and a totally different 3'-terminal RNA structure.

The determination of the nucleotide sequences of RNAs of viruses from all four Tricornaviridae groups has enabled extensive comparative analysis of the viral RNAs and their putative protein products. These studies are discussed in detail in Chapters 2 and 3.

#### **4. Cucumber mosaic virus.**

CMV is the type member, and by far the best characterized, of the Cucumovirus group. The other two members, tomato aspermy virus (TAV) and peanut stunt virus (PSV), have been studied less but have provided interesting comparisons with CMV (Kaper and Waterworth, 1981; Habili and Francki, 1974a,b,c; Rao and Francki, 1981; Lot and Kaper, 1976a).

A great deal of information has been accumulated on the physico-chemical, biological and serological characteristics of CMV (see Francki *et al*, 1979, 1985 and references therein) and more recently, on its molecular biology (Symons *et al*, 1982; Symons, 1985; this work).

**Biological properties.**

CMV is an extremely ubiquitous virus which occurs throughout the world, producing a wide variety of symptoms on a very large number of hosts. It can infect at least 775 species in 52 families of flowering plants (Francki *et al*, 1985 and references therein) and can be transmitted by at least 60 species of aphids (Francki *et al*, 1979). A large number of strains of CMV have been described (Kaper and Waterworth, 1981) although only a few have been well characterized. The experimental results presented in this thesis, and all discussions concerning CMV, refer to the Q-strain (Francki *et al*, 1966), unless otherwise specified.

**Virion structure.**

The CMV capsid is approximately 29 nm in diameter and is composed of 180 molecules of coat protein ( $M_r$  approx. 24,000; Davies and Symons, 1988) arranged with icosahedral (T=3) symmetry (Finch *et al*, 1967). The viral RNA forms a layer inside the protein shell leaving a central hole approximately 12 nm in diameter (Jacrot *et al*, 1977). CMV particles are held together by RNA-protein interactions, as evidenced by their susceptibility to disruption by SDS or high salt concentrations or by treatment with ribonuclease ( Francki *et al*, 1966; Kaper and Geelen, 1971; Habili and Francki, 1974b). The factors controlling the structure and stability of particles of CMV and other Tricornaviridae have been reviewed by Johnson and Argos (1985).

**RNA components.**

The molecular weights ( $M_r$ ) of CMV RNAs 1 to 4 have been estimated by polyacrylamide gel electrophoresis to be  $1.35 \times 10^6$ ,  $1.16 \times 10^6$ ,  $0.85 \times 10^6$  and  $0.35 \times 10^6$  respectively (Peden and Symons, 1973). Infectivity studies with combinations of purified RNAs revealed that RNAs 1, 2 and 3 were necessary and sufficient for infection (Peden and Symons, 1973; Lot *et al*, 1974). The RNAs are encapsidated to produce three types of particles which have indistinguishable morphological and sedimentation properties – one containing a single copy of RNA 1, one containing a single copy of RNA 2 and another containing one copy each of RNAs 3 and 4 (Lot and Kaper, 1976b).

The nucleotide sequence of the entire CMV genome has now been determined (Davies and Symons, 1988; Chapter 2 of this work) confirming earlier conclusions from liquid hybridization experiments (Gould and Symons, 1977) that CMV RNAs 1 to 3 are unique species except for a common sequence at their 3'-termini. A detailed discussion of the structure of the CMV genomic RNAs appears in Chapter 2.

### **Satellite RNAs.**

Certain preparations of CMV also contain satellite RNA which is a small RNA molecule (about 350 nucleotides) unable to replicate without the assistance of CMV as a "helper" virus, but showing no appreciable sequence homology with the CMV genome (reviewed in Francki, 1985b). Several satellite RNAs have been isolated from different strains of CMV and each one is able to modify the expression of CMV symptoms. A satellite has also been found associated with PSV (Kaper *et al.* 1978) but not with any other members of the Tricornaviridae. The biological effects of sat-RNAs appear to be the result of their interaction with the host plant and the helper virus. Interactions between satellite RNA and the CMV genomic RNAs are discussed in detail in Chapter 2.

### **Proteins encoded by viral RNAs.**

Translation of CMV RNAs 1 to 4 *in vitro* has revealed that each acts as a monocistronic mRNA producing proteins of Mr approximately 95,000 (P1), 110,000 (P2), 35,000 (3a) and 24,500 (CP) respectively (Schwinghamer and Symons, 1975, 1977; Gordon *et al.* 1982). However, in CMV-infected protoplasts and leaves, the coat protein is the only detectable virus-encoded protein (Gonda and Symons, 1978; Roberts and Wood, 1981). The functions of the P1, P2 and 3a proteins are unknown, although computer-assisted comparison of their amino acid sequences with those of other viral proteins have revealed homologies with proteins of known or implied function. These comparisons, and their implications for the possible *in vivo* role of CMV-encoded proteins, are discussed in Chapter 3.

## **Replication of viral RNA.**

CMV RNA is replicated by a virus-induced RNA replicase which has been extensively purified from both the particulate and soluble fractions of CMV-infected cells and found to consist of a major polypeptide of Mr 100,000 (Gill *et al*, 1981; Gordon *et al*, 1982; Kumarasamy and Symons, 1979a). Similar enzymes are induced by other plant viruses (reviewed in Hall *et al*, 1982) although their molecular composition differs between viruses and is not yet completely understood. In many plants, study of viral replicases is complicated by the presence of a host-coded RNA-dependent RNA polymerase (reviewed in Fraenkel-Conrat, 1986) which may be involved in the formation of the viral replicase complex. Information on the structure of these enzymes, particularly the CMV-induced RNA replicase, appears in Chapter 4.

## **5. Aims and scope of this work.**

The overall aim of studies on CMV is to understand its structure and function at the molecular level. The work described here was undertaken to increase our understanding of the genome structure of CMV and those sequence and structural features which are important for its function or expression.

Chapter 2 describes the determination of the complete nucleotide sequences of CMV RNAs 1 and 2. This completed the primary structure of the CMV genome and enabled an analysis of the structural features conserved between the RNAs of CMV and between CMV and other viruses.

Chapter 3 describes the comparison of the nucleotide sequences of CMV RNAs 1 and 2, and the amino acid sequences of the P1 and P2 proteins with sequences of other viral RNAs and proteins. The results provided evidence for evolutionary relationships between the Tricornaviridae and other viruses with vastly different physical and biological properties, and suggested a role for the CMV P1 and P2 proteins in RNA replication. This prompted an investigation of the host- or virus-encoded nature of the protein components of the CMV-induced RNA replicase.

Chapter 4 therefore describes work aimed at purifying the major Mr 100,000 protein component of CMV-induced RNA replicase in sufficient amounts to allow amino acid sequencing. This involved investigation of different extraction, storage and chromatographic purification procedures suitable for large-scale preparation of RNA replicase from CMV-infected plants.

The concluding discussion in Chapter 5 considers the approaches which could be taken in the future to further increase our understanding of the function of the CMV genome.

CHAPTER TWO.

THE NUCLEOTIDE SEQUENCE OF RNAs 1 AND 2 OF  
CUCUMBER MOSAIC VIRUS (O STRAIN).



## INTRODUCTION.

One of the primary pre-requisites for study of a virus at the molecular level is a knowledge of the primary structure of its genome. The development of rapid sequencing techniques has enabled the determination of the nucleotide sequences of many viral genomes and allowed a detailed study of their structure.

The genome structures and replication strategies of plant viruses, deduced from sequencing, *in vitro* translation and other studies, have been the subject of several reviews (Davies and Hull, 1982; Joshi and Haenni, 1984; Dougherty and Hiebert, 1985; Symons, 1985), although even the most recent are now outdated due to the continual increase in sequence information and its analysis.

This chapter describes the nucleotide sequencing of Q-CMV RNAs 1 and 2, completing the sequence of the CMV genome. The originally published sequence of RNA 3 (Gould and Symons, 1982) has been recently re-investigated and a corrected version of the sequence, together with partial amino acid sequence data for the coat protein, has been reported by Davies and Symons (1988).

The nucleotide sequence information available for the Tricornaviridae is summarized in Tables 2.1 and 2.2. The complete sequences of the CMV, BMV and AMV genomes have been determined, together with RNA 3 of TSV and RNA 4 of CCMV (Table 2.1). Several other RNAs have been partially sequenced, mainly at their 3'-termini (Table 2.2). This sequence information has allowed detailed, comparative studies of each viral genome and of those features which are similar between different genomes. The structural features of the CMV genome, and their biological implications, are discussed in this chapter. Similarities between the proteins encoded by CMV and other viruses are discussed in Chapter 3.

Sequencing of CMV RNAs 1 and 2 was done in collaboration with Dr. M.A. Rezaian and published in Rezaian *et al* (1985, 1984), respectively. Initial sequencing studies utilized cDNA clones of CMV RNAs 1 and 2 which had been prepared in M13 vectors by Dr. A.R. Gould. Since this work was done by Dr. Gould, the methods used have not been included here. Further M13 clones were

constructed by Dr. M.A. Rezaian using DNA fragments from the original clones as internal primers for cDNA synthesis. The methods used to produce these clones have been included in this chapter in a summarized form only.

All of the clones thus prepared were sequenced by the dideoxy chain termination method of Sanger *et al* (1980). In addition, both RNAs were sequenced by the dideoxy technique (Zimmern and Kaesberg, 1978) using synthetic oligonucleotides or fragments prepared from selected M13 clones as primers. The extreme 5'-terminal sequence of each RNA was determined by partial enzymatic cleavage of decapped, 5'-<sup>32</sup>P-labelled RNA. All nucleotide sequencing work, and analysis of sequence data, was a collaboration between the author and Dr. M.A. Rezaian with roughly equal practical and intellectual contributions.

**TABLE 2.1: Complete<sup>†</sup> nucleotide sequences of RNAs of Tricornaviridae.**

GROUP	VIRUS	STRAIN	RNA SEGMENT	LENGTH (nucleotides)	REFERENCE
Cucumovirus	CMV	Q	1	3389	this work
			2	3035	this work
			3	2197	Davies and Symons (1988)
			4	1031	Davies and Symons (1988)
Bromovirus	BMV	Russian	1	3234	Ahlquist <i>et al</i> (1984)
			2	2865	Ahlquist <i>et al</i> (1984)
			3	2111 <sup>a</sup>	Ahlquist <i>et al</i> (1981b)
			4	876	Dasgupta and Kaesberg (1982)
	CCMV	4	824	Dasgupta and Kaesberg (1982)	
Alfalfa mosaic virus group	AMV	425-L <sup>b</sup>	1	3644	Cornelissen <i>et al</i> (1983a)
			2	2593	Cornelissen <i>et al</i> (1983b)
			3	2142	Langereis <i>et al</i> (1986)
			4	881	Brederode <i>et al</i> (1980)
		425-M <sup>b</sup>	3	2037	Barker <i>et al</i> (1983)
		S	3	2111 <sup>c</sup>	Ravelonandro <i>et al</i> (1984)
Ilarvirus	TSV	WC	3	2205	Cornelissen <i>et al</i> (1984)

<sup>†</sup> Only those RNAs for which a complete nucleotide sequence is available have been listed, together with the major reference.

a. BMV RNA 3 contains an internal poly(A) tract of heterogeneous length (16-22 nucleotides). The total length of RNA 3 therefore varies from 2111-2117 nucleotides.

b. The 425 strain of AMV exists as two isolates, 425-L and 425-M, which have been propagated separately in Leiden and Madison, respectively.

c. Ravelonandro *et al* (1984) reported the 5'-untranslated region of this RNA as 257 nucleotides (ntd) long and the total length as 2055 ntd. Langereis *et al* (1986) found that a 56-ntd segment of the 5' untranslated region was repeated taking the length of this region to 313 ntd and the total length of the RNA to 2111 ntd.

**TABLE 2.2: Partial<sup>†</sup> nucleotide sequences of RNAs of Tricornaviridae.**

GROUP	VIRUS	STRAIN	RNA SEGMENT	REGION	LENGTH (nucleotides)	REFERENCE
Cucumovirus	CMV	Y	4	5'	106	Hidaka <i>et al</i> (1985)
			1	3'	189	Wilson and Symons (1981)
	TAV	N	3	3'	189	Wilson and Symons (1981)
			1	3'	189	Wilson and Symons (1981)
			3/4	3'	188	Wilson and Symons (1981)
Bromovirus	CCMV		1	3'	208	Ahlquist <i>et al</i> (1981a)
			2	3'	234	Ahlquist <i>et al</i> (1981a)
			3/4	3'	266	Ahlquist <i>et al</i> (1981a)
	BBMV		1	3'	200	Ahlquist <i>et al</i> (1981a)
			2	3'	214	Ahlquist <i>et al</i> (1981a)
			3/4	3'	266	Ahlquist <i>et al</i> (1981a)
Alfalfa mosaic virus group	AMV	Q	1	3'	226	Gunn and Symons(1980)
			2	3'	228	Gunn and Symons(1980)
			3/4	3'	230	Gunn and Symons(1980)
Ilarvirus	TSV		2	3'	180	Koper-Zwarthoff and Bol(1980)

<sup>†</sup>Only those RNAs of Tricornaviridae for which partial nucleotide sequences have been obtained are tabulated together with the source reference.

## MATERIALS AND METHODS

### MATERIALS.

#### **1. Enzymes.**

Avian myeloblastosis virus reverse transcriptase was obtained from Molecular Genetic Resources (Florida). Klenow fragment of *E. coli*. DNA polymerase I was from BRESATEC (Adelaide, South Australia) and restriction endonucleases were from New England Biolabs. Bacteriophage T4 DNA ligase and polynucleotide kinase were both obtained from Boehringer. Alkaline phosphatase (Type VIII-L from bovine intestine) was from Sigma.

#### **2. Sequencing reagents.**

The bacteriophage M13-specific oligonucleotide sequencing primer, 5'-d(GTAAAACGACGGCCAG), was obtained from New England Biolabs and deoxy- and dideoxy-nucleotide triphosphates were from Sigma. CMV RNA-specific oligonucleotides used for sequencing were kindly synthesized by Dr. Derek Skingle and Stephen Rogers.

#### **3. Radiochemicals.**

[ $\alpha$ -<sup>32</sup>P]-dATP, specific activity 2000 Ci/mmol, and [ $\gamma$ -<sup>32</sup>P]-ATP, specific activity approx. 2000 Ci/mmol, were obtained from BRESATEC (Adelaide, South Australia).

#### **4. General Materials.**

Stock solutions of acrylamide and N,N'-methylene-bis-acrylamide (Sigma) in water were deionised with Amberlite MB-1 mixed bed resin (Sigma) and stored at 4°C. All chemicals and solvents were of analytical reagent grade.

### METHODS.

#### **1. Virus and viral RNA.**

Cucumber mosaic virus, strain Q (Francki *et al*, 1966) was grown in *Cucumis sativus* (cucumber) cv. Supermarket or in *Nicotiana clevelandii* under glasshouse conditions. Virus was extracted from infected leaves by the method of

Peden and Symons (1973). RNA was isolated from purified virus by phenol extraction in the presence of 0.3% (w/v) SDS, also as described by Peden and Symons (1973). Total viral RNA was fractionated into individual RNAs by the two-step polyacrylamide gel electrophoresis and electro-elution method of Symons (1978).

## **2. Purification of CMV RNAs by spermine precipitation.**

Individual CMV RNAs were further purified by precipitation with spermine (Hoopes and McLure, 1981) to remove contaminating acrylamide. A sample of RNA (50-100  $\mu$ l) in an Eppendorf tube was made 5 mM with spermine and incubated on ice for 15 minutes. After centrifugation for 15 minutes at 4°C, the pellet was washed with 1 ml of 0.3 M sodium acetate, 0.01 M magnesium acetate, 75% (v/v) ethanol on ice for 1 hour and then pelleted again by centrifugation for 10 minutes. The pellet was washed briefly with 1 ml of cold 75% (v/v) ethanol, dried under vacuum and resuspended in 0.1 mM EDTA. RNA purified in this way was found to give good results during preparation of cDNA and in RNA dideoxy sequencing reactions.

## **3. Preparation of cDNA clones from CMV RNA 1 and 2 (Summary).**

DNA primers prepared from selected M13 clones (prepared by Dr A.R. Gould) were hybridized to purified CMV RNAs 1 or 2 and cDNA synthesized essentially as described by Gould and Symons (1982). The fragments produced by Sau3A digestion of ds-cDNA were separated by electrophoresis on polyacrylamide gels containing TBE buffer (90 mM Tris-borate, pH 8.3, 1 mM EDTA) and 2.5 M urea. Individual fragments were eluted from gel slices, ethanol precipitated and ligated into BamH1-cleaved, phosphatase-treated M13mp9 vector (Messing and Vieira, 1979). These recombinants were used to transform *E. coli* strain JM101 (Messing, 1979) and white recombinant plaques were selected for preparation of M13 DNA and nucleotide sequencing (see below).

## **4. Preparation of single-stranded M13 DNA.**

Single-stranded DNA was prepared from selected recombinant phage essentially as described by Winter and Fields (1980). The final DNA pellet, after

ethanol precipitation, was resuspended in 30-50  $\mu$ l TE buffer (10 mM Tris-HCl, pH 8.0, 0.1 mM EDTA).

### 5. Dideoxy sequencing of M13 clones.

Single-stranded M13 DNA was sequenced by the dideoxy chain-termination method (Sanger *et al*, 1980) using synthetic universal 17-mer primer, [ $\alpha$ - $^{32}$ P]-dATP and Klenow fragment of *E. coli* DNA polymerase I. Improved sequencing results were obtained by using a "chase solution" containing 250  $\mu$ M dATP and 75  $\mu$ M each of dCTP, dGTP and dTTP. Sequencing reactions were stopped by the addition of an equal volume of formamide-dye mix (95% v/v deionized formamide, 10 mM EDTA, 0.1% w/v each of xylene cyanol FF and bromophenol blue, 50 mM NaOH) and then heated at 100°C for 2 minutes before fractionation on 6% polyacrylamide gels (40 x 20 x 0.025 cm) containing TBE buffer and 7M urea. Gels were fixed for 10 minutes in a solution of 10% acetic acid, 20% ethanol and washed with 20% ethanol before being dried onto the glass plate for autoradiography.

### 6. Preparation of DNA primers from M13 clones.

Selected M13 clones containing sequences of the same polarity as viral RNA were used to prepare primers for RNA dideoxy sequencing. M13-specific 17-mer primer was hybridized to 5-20  $\mu$ g of single-stranded recombinant DNA and extended using Klenow fragment of *E. coli* DNA polymerase I in a reaction containing 10 mM Tris-HCl, pH 7.5, 10 mM MgCl<sub>2</sub>, 10 mM DTT, 0.15-0.2 mM each of dCTP, dGTP and dTTP, 2.5  $\mu$ M unlabelled dATP and 50-100  $\mu$ Ci [ $\alpha$ - $^{32}$ P]-dATP. Incubation was for 50 minutes at 37°C, followed by a chase with 0.1 mM dATP for a further 10 minutes.

Following incubation at 65°C for 10 minutes to destroy Klenow fragment, appropriate restriction endonucleases were used (according to manufacturer's instructions) to excise the required  $^{32}$ P-labelled fragments. Digestion was terminated by the addition of an equal volume of formamide-dye mix and heating at 100°C for 2 minutes before fractionation on a 6% polyacrylamide gel (40 x 20 x 0.05 cm) containing TBE buffer and 7M urea. In some cases digestion products

were ethanol precipitated and resuspended in formamide-dye mix prior to gel electrophoresis.

Primer bands, located by autoradiography, were excised from the gel and eluted by soaking the gel slice overnight at 37°C in 50 mM Tris-HCl, pH 7.0, 2 mM EDTA, 0.1% SDS. Eluted primers were recovered by ethanol precipitation in the presence of 0.3 M sodium acetate, washed twice with cold 70% ethanol, and dried under vacuum.

The elution buffer of Maxam and Gilbert (1980) was avoided since it contained high levels of ammonium acetate which was found to interfere with subsequent sequencing reactions.

### **7. Annealing of primers to CMV RNA for sequencing.**

Purified DNA primers were hybridized to total CMV RNA in 20-50  $\mu$ l of 10 mM Tris-HCl, pH 7.0, 0.18 M NaCl, 1 mM EDTA, 0.05% SDS, by incubation at 90°C for 2 minutes and then at 65°C for 60 minutes. The DNA-RNA hybrid was recovered by ethanol precipitation, washed twice with cold 70% ethanol, dried under vacuum and finally resuspended for sequencing reactions.

In the case of synthetic oligonucleotide primers, 1  $\mu$ g of total CMV RNA was mixed with a 20-fold molar excess of primer in 0.1 mM EDTA, heated at 75°C for 1 minute and allowed to equilibrate slowly to room temperature. The hybrid in this solution was then added directly to the sequencing reactions.

### **8. Dideoxy sequencing of RNA.**

Aliquots of 200 ng of CMV RNA with hybridized primer were added to 5  $\mu$ l dideoxy sequencing reactions containing 50 mM Tris-HCl, pH 8.3 (at 42°C), 70 mM KCl, 10 mM MgCl<sub>2</sub>, 10 mM DTT, 2 Units reverse transcriptase, 10  $\mu$ M dATP containing 15  $\mu$ Ci [ $\alpha$ -<sup>32</sup>P]-dATP, 50  $\mu$ M of dNTP corresponding to the ddNTP in the reaction, 250  $\mu$ M each of the other two dNTPs and one of the ddNTPs at the following concentrations: 2  $\mu$ M ddATP, 15  $\mu$ M ddCTP, 15  $\mu$ M ddGTP or 25  $\mu$ M ddTTP. Reactions were incubated at 42°C for 15 minutes. After addition of 1  $\mu$ l of 1 mM dATP (unlabelled) as a chase, and incubation for a further 30 minutes, the reactions were stopped with an equal volume of formamide-dye mix and



fractionated on 6% polyacrylamide gels as described for M13 dideoxy sequencing (see above).

Some ambiguities in the sequencing results could be resolved by increasing the ddNTP concentrations 2-3 fold, although strong compressions required the use of dITP in place of dGTP (see below).

### **9. Dideoxy sequencing of M13 clones and RNA using dITP and reverse transcriptase.**

Band compressions in both M13 and RNA dideoxy sequencing could be resolved by modifying the RNA dideoxy reactions (see above) so that dGTP was replaced with dITP (Mills and Kramer, 1979) at a concentration of 0.5 mM and the ddGTP concentration was reduced to 4  $\mu$ M. Reverse transcriptase was then used for dideoxy sequencing on both DNA and RNA templates.

### **10. Decapping of CMV RNAs.**

The m<sup>7</sup>G cap structure was removed from purified CMV RNA 1 or RNA 2 by treatment with periodate and aniline essentially as described by Kemper (1976). Ten  $\mu$ g of RNA was incubated in 0.1 M sodium acetate, pH 5.3, 1 mM EDTA, 1 mM sodium periodate (freshly prepared) for 2 hr at room temperature in the dark. The RNA was precipitated with ethanol, washed twice with cold 70% ethanol, dried under vacuum and then resuspended in 50  $\mu$ l of freshly prepared 0.33 M aniline-HCl. After a further incubation for 2 hours in the dark, the RNA was ethanol precipitated, washed and dried as before.

To remove 5'-phosphate groups, the decapped RNA (10  $\mu$ g) was resuspended in 50 mM Tris-HCl, pH 8.0 and incubated at room temperature for 30 minutes in the presence of 0.05 U calf intestinal phosphatase, which had been previously treated to remove ribonuclease contamination (see below). The phosphatase was inactivated by adding nitrilotriacetic acid, pH 7.6 to 6 mM and incubating at room temperature for 20 minutes and then at 100°C for 1 minute (Silberklang *et al*, 1979).

### **11. Removal of ribonuclease contamination from calf intestinal phosphatase.**

Prior to use on CMV RNA, calf intestinal phosphatase (Sigma, Type VII-L) was treated with diethyl pyrocarbonate (DEP) essentially as described by Efstratiadis *et al* (1977). Phosphatase, diluted in 10 mM Tris-HCl, pH 8.0, 50 mM KCl, 50% (v/v) glycerol, was made 10 mM with triethanolamine, pH 7.8 and a half volume of DEP added. The mixture was left on ice for 1 hour. DEP was then removed by washing 5 times with diethyl ether and finally aspirating with air to remove traces of ether. Phosphatase activity was assayed by measuring the increase in absorbance at 410 nm due to the hydrolysis of p-nitrophenyl phosphate to p-nitrophenol (Worthington enzyme manual, 1972, p 150). DEP treatment reduced the phosphatase activity from 0.05 U/ $\mu$ l to approximately 0.04 U/ $\mu$ l, where 1 Unit is defined as hydrolysing 1  $\mu$ mole of p-nitrophenyl phosphate per minute at 25°C, pH 8.0.

Absence of nuclease activity was confirmed by incubating total CMV RNA in the presence of DEP-treated phosphatase and then analyzing the intactness of the RNA by agarose gel electrophoresis.

### **12. 5'-Labelling of CMV RNAs with $^{32}\text{P}$ .**

Decapped and phosphatase-treated RNA was labelled at the 5'-terminus with [ $\gamma$ - $^{32}\text{P}$ ]-ATP and polynucleotide kinase. Reaction mixtures (20  $\mu$ l) containing 50 mM Tris-HCl, pH 8.0, 10 mM  $\text{MgCl}_2$ , 0.2 mM spermidine-HCl, 10 mM DTT, 100  $\mu$ Ci [ $\gamma$ - $^{32}\text{P}$ ]-ATP, 2 Units T4 polynucleotide kinase and approximately 1  $\mu$ g of decapped RNA, were incubated at 37°C for 1 hr. The mixture was then diluted to 100  $\mu$ l with  $\text{H}_2\text{O}$  and the labelled RNA removed from unincorporated [ $\gamma$ - $^{32}\text{P}$ ]-ATP by precipitation with 6 mM spermine (Hoopes and McClure, 1981).

### **13. Sequencing of 5'-terminus of RNA.**

RNA labelled at the 5'-terminus with  $^{32}\text{P}$  was sequenced using the partial enzymatic cleavage method, as described by Haseloff and Symons (1981).

#### **14. Computer manipulation of sequence data.**

Sequence data were stored and manipulated on a PDP-11/23 computer using a package of programs based on those of Staden (1980, 1982a). Completed sequences were transferred to VAX 11/750 and 11/785 computers for detailed analysis using an extensive package of programs generously supplied by Dr. A.H. Reisner, Commonwealth Scientific and Industrial Research Organization, Division of Molecular Biology, North Ryde, NSW.

Homologies between CMV RNAs 1, 2 and 3 were detected using the dot matrix comparison program DIAGON (Staden, 1982b). The program depicts homologies as dots on the terminal screen if a measure of the proportion of similar residues (the 'score') within a specified length (the 'span') reaches a preset level. The significance of homologies is assessed by calculating the 'double-matching probability' (McLachlan, 1971). This is the probability of obtaining a particular score given two infinitely long sequences of the same composition as those being compared, and assuming the the same span length.

The program SCAN (written by I.B. Dodd, Department of Biochemistry, University of Adelaide), based on the weight matrix search method of Staden (1984), was used to search the nucleotide sequences of tripartite and other viruses for a small conserved sequence in the 5'-non-coding region of CMV RNAs 1 and 2 (see Results). A weight matrix table was constructed from the aligned conserved sequences. The program calculates a mathematical score for the degree of match between the table and each position of the sequence being scanned. Homologous regions, which need not be exact matches, are detected by their high scores.

## RESULTS

### 1. Strategy for sequencing CMV RNAs 1 and 2.

The nucleotide sequences of CMV RNAs 1 and 2 were determined using both cDNA clones and RNA sequencing methods. The approach taken can be summarized as follows:

#### **i) Clones of CMV RNA 3.**

Gould and Symons (1982) prepared clones of CMV RNA 3 by ligating TaqI restriction fragments of ds-RNA 3 cDNA into the unique ClaI site of pBR325. One of the clones obtained by this procedure, p5T28, contained sequences conserved at the 3' termini of all three CMV RNAs and was therefore useful to prepare a primer which could anneal to all CMV RNAs.

#### **ii) Clones of CMVs RNA 1 and 2.**

A 41-nucleotide fragment of clone p5T28, prepared by digestion with AluI and TaqI was used as a primer to initiate cDNA synthesis on purified CMV RNAs 1 and 2. Double-stranded cDNA thus obtained was digested with TaqI and the resultant fragments purified and individually cloned into the AccI site of M13mp7 (Messing *et al*, 1981). This work was done by Dr. A.R. Gould. Further clones of the 5' region of RNAs 1 and 2 were prepared in a similar way by Dr. M.A. Rezaian using the original TaqI clones as internal primers to initiate cDNA synthesis. In this case, fragments of Sau3A- digested cDNA were cloned in BamHI-cleaved M13mp9 (Messing and Vieira, 1982).

#### **iii) Dideoxy sequencing of M13 clones.**

All of these clones were sequenced by the dideoxy chain termination technique (Sanger *et al*, 1980) but failed to yield the complete sequences of CMV RNAs 1 or 2.

#### **iv) Dideoxy sequencing of RNA.**

To complete the sequence of CMV RNAs and to confirm data obtained from sequencing of clones, synthetic oligonucleotides or fragments prepared from selected M13 clones were used, together with reverse transcriptase, for

dideoxy sequencing of CMV RNA (Zimmern and Kaesberg, 1978).

**v) Direct sequencing of RNA 5'-terminus.**





The extreme 5'-terminal sequence of each RNA was determined by partial enzymatic cleavage of decapped, 5'-<sup>32</sup>P-labelled RNA.

The relative positions of the M13 clones and those regions of RNA sequenced by each of these methods are summarized in Figure 2.1. During sequencing of RNA 2, the RNA dideoxy sequencing technique proved to be rapid and reliable, especially when synthetic oligonucleotides were used as primers. This technique was therefore used more extensively during the subsequent sequencing of RNA 1. Furthermore, dideoxy sequencing directly on purified viral RNA ensured that the sequence was free from errors caused by cloning artefacts, rearrangements or by cloning of a minor species present in the RNA preparation.

Band compressions were sometimes encountered during both M13 and RNA dideoxy sequencing, but in all cases they could be effectively resolved by substituting dGTP in sequencing reactions with its weakly base-pairing analogue, dITP (Mills and Kramer, 1979; Symons, 1979; see Methods). In addition it was found that M13 DNA could be used as a template by reverse transcriptase, and that sequencing of M13 clones using the dideoxy sequencing procedure normally used for RNA yielded very clear, non-compressed results, albeit with lower radioactive incorporation than that obtained using the usual M13 sequencing reaction.

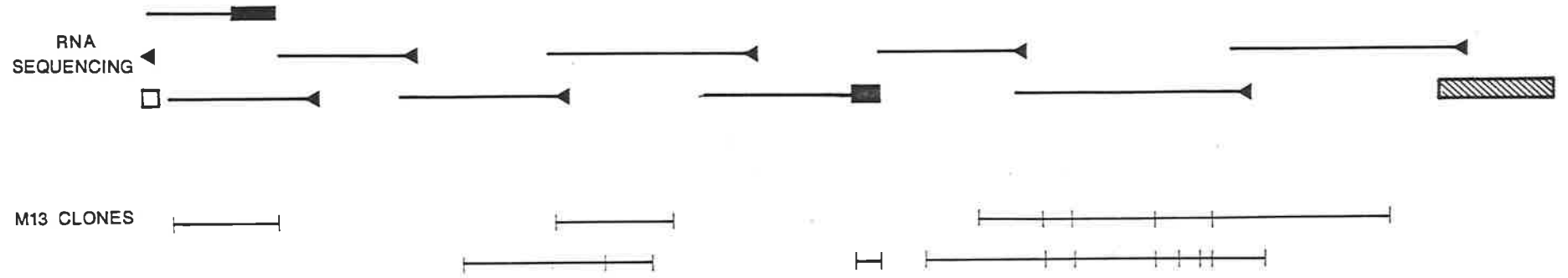
The complete nucleotide sequences of Q-CMV RNAs 1 and 2 are presented in Figures 2.2 and 2.3, respectively, together with the predicted amino acid sequences of proteins encoded by their long open reading frames (see below). RNA 1 is 3389 nucleotides long ( $M_r 1.15 \times 10^6$ ) and RNA 2 is 3035 nucleotides long ( $M_r 1.03 \times 10^6$ ), not including the m<sup>7</sup>G cap present at their 5'-termini (Symons, 1975).

**FIGURE 2.1: Strategy used for sequencing CMV RNA 1 and RNA 2.**

In addition to sequencing M13 clones of CMV cDNA, dideoxy RNA sequencing was performed using cloned DNA fragments (  ) or synthetic oligonucleotides (  ) as primers. The 5' terminus of each RNA was determined by partial enzymatic cleavage (  ). The sequence of the 3'-terminal 270 nucleotides (  ) has been published by Symons (1979).

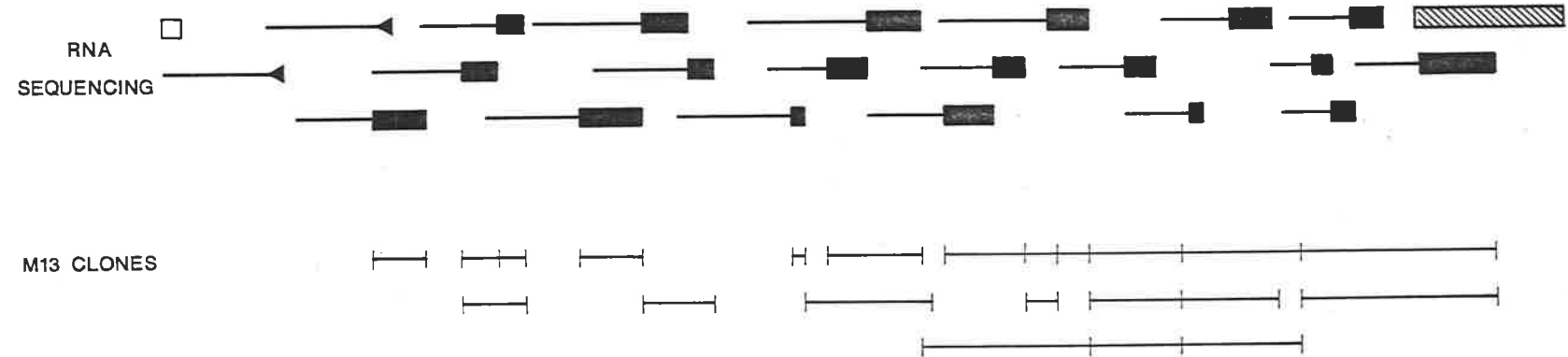
# RNA 1

RESIDUE NUMBER FROM 5'-END 1 200 400 600 800 1000 1200 1400 1600 1800 2000 2200 2400 2600 2800 3000 3200 3389



# RNA 2

RESIDUE NUMBER FROM 5'-END 1 200 400 600 800 1000 1200 1400 1600 1800 2000 2200 2400 2600 2800 3035



**FIGURE 2.2: Nucleotide sequence of Q-CMV RNA 1 and deduced amino acid sequence of the P1 protein.**

The stop codon of the P1 reading frame is identified as \*\*\*.



met ala thr ser ser phe asn ile asn glu leu val ala ser his gly asp lys gly leu leu ala thr ala leu val asp lys thr ala  
98 AUG GCA ACG UCC UCA UUC AAC AUC AAU GAA CUG GUA GCC UCC CAC GGC GAU AAA GGA CUA CUU GCG ACA GCC CUC GUU GAU AAG ACA GCU  
his glu gln leu glu glu gln leu gln his gln arg arg gly leu lys val tyr ile arg asn val leu asp val lys asp ser glu val  
188 CAU GAA CAG CUG GAG GAG CAA CUU CAA CAA CGU AGA GGC CUU AAG GUC UAC AUC CGU AAU GUU UUG GAA GUA AAG GAC UCC GAA GUC  
ile arg thr arg tyr gly gly lys tyr asp leu his leu ala gln gln glu leu ala pro his gly leu ala gly ala leu arg leu cys  
278 AUU CGG ACU CGG UAU GGU GGU AAG UAC GAC CUC CAU CUC GCC CAG CAG GAA UUA GCU CCC CAU GGC CUC GCU GGU GCC CUC CGC UUG UGU  
glu thr leu asp cys leu asp phe phe pro arg ser gly leu arg gln asp leu val leu asp phe gly gly ser trp val thr his tyr  
368 GAA ACU CUC GAU UGU CUA GAC UUU UUC CCU GAC UCA GGU CUC GCG CAG CUC GUC UUA GAU UUC GGA GAA AGU UGG UCC ACA CAC UAU  
leu arg gly his asn val his cys cys ser pro cys leu gly ile arg asp lys met arg his thr glu arg leu met ser met arg lys  
458 CUU CGC GGA CAC AAC GUA CAU UGC UGC UCC CCA UGU UUG GGU AUU CGU GAC AAG AUG CGU CAC ACU GAA AGG UUG AUG ACC AUG CGC AAG  
val ile leu asn asp pro gln gln phe asp gly arg gln pro asp phe cys thr lys ser ala ala glu cys lys val gln ala his phe  
548 GUC AUU UUA AAC GAU CCA CAA CAG UUU UUC GGC CAG CUC GCG CAG CCG GAU AAU UUA UGC ACU AAG UCU GCG GCU AAA GGU CAA GCC CAC UUU  
ala ile ser ile his gly gly tyr asp met gly phe arg gly leu cys glu ala met asn ala his gly thr thr ile leu lys gly thr  
638 GCU AUU UCU UUA CAC GGA GGA UAU GAU AUG GGC UUU AGA GGU UUA UGU GAG GCA AUG AAC GCU CAC GGA ACU ACG AUU UUG AAA GGG ACG  
met met phe asp gly ala met met phe asp asp gln gly phe ile pro glu leu lys cys gln trp arg lys ile lys ser ala phe ser  
728 AUG AUG UGU CUC GGU GGU AUG AUG UUU GAC AAA CAA CAG UUU UUC GAG CUC UUU AUU AAA UGU CAG UGU AAA UGC AAA ACC GCU UUU UCU  
glu glu glu asp ala thr cys ser ala ala lys leu asn ser ser val phe ser arg val arg asn gly lys thr leu ile ala phe asp  
818 GAG GAG GAG GAU GCC ACU UGU UCA GCA GCU AAA CUC AAU UCU AGU GUU UUU UCA CGC GUG CGA AAU GGG AAA ACC UUA AUC GCA UUU GAC  
phe val glu glu ser thr met ser tyr val his asp trp asp asn ile lys ser phe met thr asp gln thr tyr ser phe asn gly met  
908 GUC AUU UGU CUC GAU UGC UCU UAU GUU CAC GAU UGU CCG GAU AAA UCG UUU AUG ACA GAU GAU ACG GAA GCU AAA ACC GCU UUU GGU AUG  
thr tyr gly ile glu arg cys val ile tyr ala gly val met thr tyr lys ile val gly val pro gly met cys pro pro glu leu ile  
998 ACC UAU GGA AUU GAG CGU UGU GUU AUU UAC GCU GGC GUG AUG ACU UAC AAG AUU GUU GGC GUG CCU GGU AUG UGU CCG CCC GAA CUC AUU  
arg his cys ile trp phe pro ser met lys asp tyr val gly leu lys ile pro ala ser asp asp leu val lys trp lys thr val arg  
1088 CGA CAU UGU AUC UGG UUC CCC UCU AUG AAG GUC AUU GUU GGU CUU AAA UUU CCC GCG UCU GAU CCC GCG UCU GAA UGC UUG AAA ACC GAC GUC CGU  
ile leu leu ser thr leu arg glu thr glu glu ile ala met arg cys tyr asn asp lys lys asn trp met asp leu phe lys ile ile  
1178 AUU UUA CUG UCA ACA UUA CGU GAG ACU GAA GAA AUA GCU AUG CGU UGU UAU AAC GAC AAG AAG AAU UGG AUG GAU CUA UUC AAG AUU AUU  
leu gly val leu ser ser lys ser ser thr ile val ile asn gly met ser met gln ser gly glu arg ile asp leu asn asp tyr his  
1268 CUC GGU GUU UUA UCA UCG AAG UCC UCC ACG AUC GUU AUC AAU GGU AUG UCU AUG CAA UCC GGU GAA CGU AUU GAU CUA AAU GAU UAU CAU  
tyr ile gly phe ala ile leu leu his thr lys leu lys tyr glu gln leu gly lys met tyr asp met trp asn ala ser phe ile trp  
1358 UAC AUC GGU UUU GCU AUU CUU CUC CAC ACG AAA UUA AAA UAC GAA CAA CUU GGA AAA AUG UAC GAU AUG UGG AAU GCU UCC UUC AUU UGG  
lys trp phe ala ser met ser arg pro phe arg val phe phe ser thr val val lys thr leu phe pro thr leu arg pro arg glu glu  
1448 AAG UGG UUU GCG UCU AUG UCU AGA CCA UUC CGU GUU UUC UUU UCC ACU GUU UUU AAG ACU UUG UUG CCG ACU UUG AGA CCG CGC GAG GAA  
lys glu phe leu val lys leu ser thr phe val thr phe asn glu glu cys ser phe asp gly gly lys glu trp asp val ile ser ser  
1538 AAG GAG UUU UUG GUC AAA CUU UCC ACU UUC GUC ACC UUU AAC GAG GAG UGC UCU UUU GAC GGA GGG AAA GAA UGG GAC GUG AUA UCA UCA  
ala ala phe val ala thr gln ala val ala asp gly thr ile leu ala glu glu lys ala lys lys leu ala asp arg leu ala val pro  
1628 GCG GCU UUC GUA GCA ACU CAG GCU GUU GCA GAU GGC ACU AUU CUG GCC GAG GAG AAA GCU AAG AAA UUA GCU GAU GCU CUG GCC GUG CCU  
val glu glu val thr ala ile pro glu val ser pro thr pro val asp gln gly thr ala cys gly thr glu thr glu thr ser glu leu  
1718 GUU GAA GAA GUU ACU GCU AUU CCU GAG GUG UCU CCU ACA CCU GUU ASP CAG GGC ACU GCU UGU GGA CUG GAA ACA GAA ACA UCG GAA CUG  
asp ser leu ser ala gln thr arg ser pro ile ala arg ile ala glu arg ala thr ala met leu glu tyr ser ala tyr glu lys gln  
1808 GAU UCU CUG UCU GCC CAA ACA CGU UCC CCC AUC GCA CGG AUC GCU GAA AGA GCG ACC GCU AUG CUU GAA UAU UCA GCU UAU GAG AAA CAA  
leu his asp thr thr val ser asn leu gln arg ile trp cys met ala gly gly asp asn lys arg asn ser leu glu ser asn leu lys  
1898 UUG CAC GAU ACC ACC GUU UCA AAU CUU CAA CGA AUU UGG UGC AUG GCG GAC AAC AAG AGA AAC UCU UUA GAG AGU AAU UUG AAA  
phe val phe asp thr tyr phe ser val asp ala leu val asn val his phe pro thr gly arg trp met his pro val pro glu gly val  
1988 UUU GUG UUU GAC ACU UAU UUU UCU GUU GAC GCC CUA GUG AAU GUU CAC UUU CCU ACU GGG AGA UGG AUG CAC CCU GUG CCU GAG GGC GUG  
val tyr ser val gly tyr asn glu lys gly leu gly pro lys leu asp ser glu leu tyr ile val asn gly asp cys val ile ser asn  
2078 GUA UAC UCU GUU GGU UAU AAU GAG AAG GGU CUU GGU CCU AAA CUU GAU AGU GAG UUG UAC AUA GUU AAU GGU GAU UGU UUC AUU UCG AAC  
ser his asp leu phe ser ile thr lys ser leu leu ala pro thr gly thr ile ser gln val asp gly val ala gly cys gly lys thr  
2168 AGU CAU GAU UUG UUU UCA AUC ACU AAA UCU UUG UUA GCU CCC ACC GGA ACC AUC AGC CAA GUC GAU GGU GUA GCU GGG UGC GGG AAA ACC  
thr ala ile lys ser met phe asn pro ser thr asp ile ile val thr ala asn lys lys ser ala gln asp val arg tyr ala leu phe  
2258 ACU GCU AUA AAA UCC AUG AAU CCU UCC ACA GAU AUA AUU GUC ACR AAC AAG AAA UCU UCU GCU CAA GAA GUG CGU UAU GCG GCU UUU  
lys ser thr asp ser lys glu ala cys ala phe val arg thr ala asp ser ile leu leu asn asp cys pro thr val ser arg val leu  
2348 AAA UCU ACU GAC UCC AAA GAA GCU UGU GCU UUU GUU AGG ACC GCU GAU UCC AUA UUA CUC AAU GAU UGC CCU ACG GUG UCU CGA GUG CUU  
val asp glu val val leu leu his phe gly gln leu cys ala val met ser lys leu his ala val arg ala leu cys phe gly asp ser  
2438 GUG GAU GAA GUU GUU UUG UUG CAC UUU GGU CAG UUG UGC GCU UUG UGC UCG AAA CUU CAU GCU UGC AGA GCU UUG UUC GGA GAC UCC  
glu gln ile ala phe ser ser arg asp ala ser phe asp met arg phe ser lys leu ile pro asp glu thr ser asp ala asp thr thr  
2528 GAA CAG AUA GCC UUC UCC UCU CGU GAC GCU UCG UUC GAC AUG CGU UUC UCU AAG CUU AUU CCG GAU GAG ACC AGU GAU GCG GAC ACA ACU  
phe arg ser pro gln asp val val pro leu val arg leu met ala thr lys ala leu pro lys gly thr arg thr lys tyr ser asp gly  
2618 UCU CGU AGC CCA CAA GAU GUA GUA CCA CUU GUG CGU UUG AUG GCU ACG AAG GCU CUA CCG AAA GGG ACC CGU ACG AAA UAC UCA GAU GGU  
ala gln ser lys val arg lys ser val thr ser arg ala val ala ser val ser leu val glu leu asp pro thr arg phe tyr ile thr  
2708 GCC CAA UCU AAA GUG AGG AAG UCU GUU ACU UCG CGU GCU GUU GCU AGU GUA UCU CUA GUU GAA CUG GAC CCC ACC AGA UUC UAU AUU ACG  
met thr gln ala asp lys ala ser leu ile thr arg ala lys glu leu asn leu pro lys ala phe tyr thr asp arg ile lys thr val  
2798 AUG ACG CAA GCU GAU AAA GCC UCA CUA AUC ACU AGA GCU AAG GAA CUG AAC CUA CCC AAG GCA UUC UAC ACU GAU AGG AUA AAG ACU GUC  
his glu ser gln gly ile ser glu asp his thr leu val arg leu lys ser thr lys cys asp leu phe lys phe ser tyr cys  
2888 CAU GAA UCU CAA GGU AUU UCA GAA GAU CAU GUG ACC CUG GUU AGG UUG AAG AGU ACU AAA UGU GAC CUG UUC AAG AAA UUU UCU UAC UGC  
leu val ala val thr arg his lys val thr phe arg tyr glu tyr cys gly val leu gly gly asp leu ile ala asn cys ile pro leu  
2978 UUA GUU GCA GUU ACU CGA CAC AAG GUC ACC UUC CGC UAU GAG UAC UGU GGU GUG UUA GGU GGG GAC CUA AUC GCU AAU UGU AUU CCG UUA  
val \*\*\*  
3068 GUC UGA UCGCGGA AGGUCCGAAG ACGUAAAACU ACGCUCUCUU UAUUGCGAGU CGUGAGUUGG UAGUUUUGCU UUAACUAUC UGAAGUCGU AAAUCCAGUA UUGG  
3175 UUGCGA ACGGGUUGUC CAUCCAGCUU ACGGCUAAAA UGGUCAGUAV GCCCCAAAGG CAUGCCGACA UCCUACAAGG UUGUCGAGCU ACCCUUGAAA UCAUCUCCUA GA  
3283 UUUUCUUG GAAGGGCUUC GUGAGAAGCU CGUCACGGU AAUACACUGA UAUUACCAAG AGUGCGGGUA UCGCCUGUGG UUUUCCACAG GUUCUCCUA AGGAGACCA  
3389

**FIGURE 2.3: Nucleotide sequence of Q-CMV RNA 2 and deduced amino acid sequence of the P2 protein.**

The stop codon of the P2 reading frame is identified as \*\*\*. The first AUG, upstream of the initiation codon, and the in-phase stop codon which follows it, are underlined.

Met Ile Ser Pro Pro Pro Thr Phe Ser Phe Ala Asn Leu Leu  
AAU GGC UCC UAU GGU GAC ACU CCC GAG GAA GUG GAA CGC GUU AGA CGU GAA CAA CGC GAA GAU GCU GAG CGC GCU UUA CGU AAU UAU AAG CCU UUA CCC GCU GAU GUC AGU GAG

135  
Asn Gly Ser Tyr Gly Val Asp Thr Pro Glu Glu Val Glu Arg Val Arg Arg Glu Gln Arg Glu Asp Ala Glu Ala Ala Leu Arg Asn Tyr Lys Pro Leu Pro Ala Val Asp Val Ser Glu  
AAU GGC UCC UAU GGU GAC ACU CCC GAG GAA GUG GAA CGC GUU AGA CGU GAA CAA CGC GAA GAU GCU GAG CGC GCU UUA CGU AAU UAU AAG CCU UUA CCC GCU GAU GUC AGU GAG

254  
Ser Val Pro Arg Asp Glu Pro Ile Val Ser Gln Thr Val Thr Ala Ala Pro Val Thr Ser Val Asp Asp Ala Phe Val Ser Phe Gly Ala Glu Asp Tyr Leu Glu Met Ser Pro Ser Glu  
AGU GUC CCU AGA GAC GAA CCU AUU GUC UCG CAA ACC GUC ACU GCA GCU CCU GUU ACA UCA GUC GAU GAC GCG UUU GUV UCU UCC GGU GCU GAG GAC UAC CUU GAA AUG UCC CCA UCU GAG

374  
Leu Leu Ser Ala Phe Glu Leu Met Val Lys Pro Leu Arg Val Gly Glu Val Leu Cys Ser Ser Phe Asp Arg Ser Leu Phe Ile Ser Ser Val Ala Met Ala Arg Thr Leu Leu Leu Ala  
CUG CUU UCC GCU UUU GAG UUG AUG GUC AAA CCC UUG CGU GUC GGU GAA GUG UUG UCG UCG AGU UUU GAU CGU UCG CUA UUC UCC AGC GUC GCC AUG GCU AGG ACG UUG UUG GCA

494  
Pro Leu Thr Ser Thr Arg Thr Leu Lys Arg Phe Glu Asp Leu Val Ala Ala Ile Tyr Leu Lys Thr Asp Phe Phe Leu Glu Asp Asp Gly Pro Gln Thr Asp Val Ser Gln Ser Asp Val  
CCA CUC ACA UCC ACC CGA ACG UUG AAG CGU UUU GAA GAC CCU UUG GUG GGC GCG AUC UAU CUA AAA ACU GAU UUC UUU UUA GAA GAC GAU GGG CCC CAG ACU GAU GUC UCC CAA AGC GAU GUG

614  
Pro Gly Tyr Ile Phe Glu Pro Gly Gln His Ser Ser Gly Phe Glu Pro Pro Pro Ile Cys Ala Lys Cys Asp Leu Ile Leu Tyr Gln Cys Pro Cys Phe Asp Phe Asn Ala Leu Arg Glu  
CCC GGU UAU AUC UUC GAA CCA GGG CAA CAC UCA UCC GGU UUU GAA CCC CCC CCU AUU UGU GCU AAA UGU GAC UUG AUU UGU UAU CAA UGU CCG UGU UUC GAU UUU ACA GCA CUU CGU GAG

734  
Ser Cys Ala Glu Lys Thr Phe Ser His Asp Tyr Val Ile Glu Gly Leu Asp Gly Val Ile Asp Asn Ala Thr Leu Leu Ser Asn Leu Gly Pro Phe Leu Leu Pro Val His Cys Ser Tyr  
UCU UGC GCA GAG AAA ACA UUC UCU CAC GAC UAU GUU AUC GAG GGU CUC GAU GGU GUC AUU GAU AAC GCU ACG CUG UUA ACA AAU UUG GGA CCA UUU UUG UUA CUI GIIC CAU UGU UCU UAU

854  
Ser Lys Thr Glu Asp Pro Asp Phe Val Val Asp Pro Ser Leu Ala Arg Pro Thr Asp Arg Val Asp Val His Val Val Gln Ala Val Cys Asp Thr Thr Leu Pro Thr His Gly Asn Tyr  
UCA AAG ACC GAG GAC CCC GAU UUC GUC GUU GAU CCU AGU CUU BCG CGA CCU ACU GAU AGG GUU GAU GUA CAU GUA GUU CAA GCC GUU UGC GAU ACC ACU CUG CCC ACC CAU GGU AAC UAC

974  
Asp Asp Ser Phe His Gln Val Phe Val Asp Ser Ala Asp Tyr Ser Thr Asp Met Asp His Val Arg Leu Arg Gln Ser Asp Leu Val Ala Lys Ile Pro Asp Gly Gly His Met Leu Pro  
GAU GAU UCC UUU CAU CAA GUC UUU GUG GAC UCU GCU GAU UCC ACU GAC AUG GAA UUA CGU CGG UUA CAA UCC GAU UUA GUA GCA AAA AUU CCG GAU GGU GGG CAU UGU UUA CCG

1094  
Val Leu Asn Thr Gly Ser Gly His Gln Arg Val Gly Thr Thr Lys Glu Val Leu Thr Ala Ile Lys Lys Arg Asn Ala Asp Val Pro Glu Leu Gly Asp Ser Val Asn Leu Ser Arg Leu  
GUG UUA AAU ACC GGG AGU GGU CAC CAG AGA GUA GGU ACU ACG ARG GAG GUC UUA ACA GCA AUA AAG AAA AGA AAU GCU GAU GUC CCU GAG CUU GGU GAC UCU GUC AAU CUG ACA CGC CUG

1214  
Ser Lys Ala Val Ala Glu Arg Phe Arg Leu Ser Tyr Met Asn Val Asp Ala Leu Ala Lys Ser Asn Phe Val Asn Val Val Ser Asn Phe His Ala Tyr Met Gln Lys Trp Pro Ser Ser  
AGC AAA GCA GUA GCC GAA CGG UUU CGU CUC UCG UAU AUG AAU GUU GAC GCU UUG GCU AAA AGU AAC UUU GUU AAU GUU GUC AGU AAU UUU CAC GCU UAC AUG CAA AAA UGG CCA UCU UCU

1334  
Gly Leu Ser Tyr Asp Asp Leu Pro Asp Leu His Ala Glu Asn Leu Gln Phe Tyr Asp His Met Ile Lys Ser Asp Val Lys Pro Val Val Thr Asp Thr Leu Asn Val Asp Arg Pro Val  
GGG CUU UCU UAU GAU GAU CUU CCU GAU CUU CAC GCG GAG AAU UUA CAG UUU UAU GAU CAC AUG AUU ARG UCC GAU GUU AAG CCA GUU GUC ACU GAC ACG UUG AAC GUC GAC AGA CCU GUC

1454  
Pro Ala Thr Ile Thr Phe His Lys Lys Thr Ile Thr Ser Gln Phe Ser Pro Leu Phe Ile Ser Leu Phe Glu Arg Phe Gln Arg Cys Leu Arg Glu Arg Val Val Leu Pro Val Gly Lys  
CCA GCU ACU AUU ACA UUU CAC AAA AAG ACC AUA ACA UCC CAG UUC UCA CCG UUG UUU AUA UCU CUG UUU GAG AGA UUC CAG AGA UGC CUU CCG GAA CGU UUU CUG CCC GUU GGU AAG

1574  
Ile Ser Ser Leu Glu Met Thr Gly Phe Ser Val Leu Asn Lys His Cys Leu Glu Ile Asp Leu Ser Lys Phe Asp Lys Ser Gln Gly Glu Phe His Leu Met Ile Gln Glu His Ile Leu  
AUU UCC UUC CUU GAG AUG ACU GGU UUU UCA UUC AAC AAC CAU UGU CUU GAA AUU GAU UUA UCU AAA UUC GAC AAA UCU CAA GGU GAG UUC CAC CUU AUG AUU CAA GAC CAC AUU CUC

1694  
Asn Asp Leu Gly Cys Pro Ala Pro Ile Thr Lys Trp Trp Cys Asp Phe His Arg Phe Ser Tyr Ile Lys Asp Lys Arg Ala Gly Val Gly Met Pro Ile Ser Phe Gln Arg Arg Thr Gly  
AAC GAU CUU GGU UGC CCA GCA CCC AUC ACC AAA UGG UGG UGU GAC UUU CAU AGG UUU UCC UAU AUU AAA GAC AAA CGU GCU GGG GUU GGA AUG CCC AUC AGC UUU CAA CGC GCU ACU GGU

1814  
Asp Ala Phe Thr Tyr Phe Gly Asn Thr Ile Val Thr Met Ala Glu Phe Ala Trp Cys Tyr Asp Thr Asp Gln Phe Asp Arg Leu Leu Phe Ser Gly Asp Asp Ser Leu Ala Phe Ser Lys  
GAC GCU UUC ACC UAU UUU GGC AAU ACU AUU GUC ACG AUG GCU GAG UUU GCU UGG UGU UAU GAC ACU GAU CAA UUU GAC CGA UUG CUC UUU UCU GGU GAU GAU UCU CUG GCC UUU UCU AAG

1934  
Leu Pro Pro Val Gly Asp Pro Ser Lys Phe Thr Thr Leu Phe Asn Met Glu Ala Lys Val Met Glu Pro Ala Val Pro Tyr Ile Cys Ser Lys Phe Tyr Ser Leu Met Ser Leu Val Thr  
CUC CCA CCU GUU GSA GAU CCC AGU AAG UUU ACA ACU UUA UUC AAC AUG GAA GCU AAG GUG AUG GAA CCU GCA GUA CCU UAU AUC UGU UCG AAG UUC UCU CUG AUG AGU UUG GUA ACA

2054  
Arg Phe Gln Ser Pro Thr Ile Arg Glu Ile Gln Arg Leu Gly Thr Lys Lys Ile Pro Tyr Ser Asp Asn Asn Asp Phe Leu Phe Ala His Phe Met Ser Phe Val Asp Arg Leu Lys Phe  
CGU UUU CAG UCC CCG ACC AUA CGU GAA AUC CAG CGG UUA GGU ACC AAG AAG AUA CCG UAU UCG GAC AAC AAC GAU UUC UUG UUC GCU CAU UUU AUG AGU UUU GUU GAC AGG CUG ARG UUUU

2174  
Met Asp Arg Met Ser Gln Ser Cys Ile Asp Gln Leu Ser Ile Phe Phe Glu Leu Lys Tyr Lys Lys Ser Gly Asn Glu Ala Ala Leu Val Leu Gly Ala Phe Lys Lys Tyr Thr Ala Asn  
AUG GAC AGG AUG UCG CAG UCC UGU AUU GAU CAA UUG UCC AUU UUC UUC GAG UUG AAA UAC AAG AAG UCC GGG AAC GAA GCC GCC CUA GUU UUG GGC GCG UUU AAG AAG UAC ACC GCU AAU

2294  
Phe Asn Ala Tyr Lys Glu Leu Tyr Tyr Ser Asp Arg Gln Gln Cys Asp Leu Val Asn Thr Phe Cys Ile Ser Glu Phe Arg Val Ile Arg Arg Thr Thr Val Lys Lys Lys Lys Asn Gly  
UUU AAC GCC UAU AAG GAG CUU UAU UAC UCU GAC CGU CAA CAA UGU GAC UUG GUC AAU ACG UUU UGU AUA UCU GAG UUC CCG GUG AUU CGU CCG ACU ACC GUU AAG AAG AAG AAG AAU GGA

2414  
Cys Val Asp Ser Ser Gly Val Asp Arg Arg Pro Pro Leu Ser Gln Phe Ala Gly Gly Glu Thr Ser Lys Thr Lys Val Ser Arg Gln Lys Pro Ala Ser Glu Gly Leu Gln Lys Ser Gln  
UGU GUU GAC AGU AGU GGU GUC GAC CGC CGA CCU CCA CUU AGC CAA UUU GCA GGA GGU GAA ACG UCG AAG ACG AAG GUC UCA CGU CAG AAA CCG GCG ACG GAG GGG UUA CAA AAG UCC CAG

2534  
Arg Glu Ser Ala Ile Tyr Ser Glu Thr Phe Pro Asp Val Thr Ile Pro Arg Ser Arg Ser Arg Gly Leu Val Ser \*\*\*  
CGA GAG AGC GCG AUC UAU AGC GAG ACU UUU CCA GAU GUU ACC AUU CCA CGG AGU AGA UCC CGU GGA UUG GUU UCC UGAUGUGUU CGUCUCUGU CGCUGUACC CGUUGUACC CUIUGUUUCU UUAUGAUUCU UUGA

2663  
UGAUAC UGAUUGUUU GUGUGAACG AUUGGGCCG AGGGUCUUU UGAUUCCGA CCCUUCGUC UCCGAAGACG UUAACUACG CUCUCUUUUA USCGAGUGU GAGUUGUUA GUUUGUCUA AACUACUGA AGUCGUA

2808  
A UCCAUUACUG GUUGCGAACG GGUUGUCCAU CCAGCUUACG GCUAAAUGG UCAGUAUCC CCAAGGACG UGCCGACCC UACAGGGUUG UCGAGUACC CUUGAAUUA UCUCUAGAU UUCUUCGAA GAGCUUCUG AGA

2952  
AGCUCGU GCACGUAU ACACUGAUU UACCAAGAGU GCGGGUAGC CCUGUGUUC UCCACAGUU CUCCAUAAGG AGACCA

## **2. Open reading frames in RNAs 1 and 2.**

CMV RNAs 1 and 2 each contain only one long open reading frame consistent with their ability to function *in vitro* as monocistronic mRNAs (Schwinghamer and Symons, 1975, 1977).

The reading frame on RNA 1 (Figure 2.2) initiates at the first AUG codon from the 5' terminus and extends 2976 nucleotides to a UGA stop codon finishing at nucleotide 3073. This would produce a translation product (P1) of 991 amino acids with a predicted molecular weight of 110,791. This reading frame also contains 30 in-phase AUG codons, 22 of which can potentially encode translation products longer than 536 amino acids. The next largest reading frame on the positive strand is 210 nucleotides long, while the largest reading frame on the negative strand is 201 nucleotides long.

The first AUG codon in RNA 2 occurs at nucleotide 19, but is followed by an in-phase stop codon at nucleotide 43 (Figure 2.3). The long open reading frame initiates at the second AUG at nucleotide 93 and extends 2520 nucleotides to a UGA stop codon finishing at nucleotide 2612. The predicted translation product (P2) is 839 amino acids long (Mr 94,333). The next largest open reading frame is 300 nucleotides long and the largest reading frame on the negative strand is 147 nucleotides.

The sizes of P1 and P2 deduced from the nucleotide sequences, differ from the estimates made by gel electrophoresis following *in vitro* translation – Mr 95,000 and 110,000 for P1 and P2 respectively. The reason for this discrepancy is not known, although similar observations have been made for the proteins encoded by BMV (Ahlquist *et al.*, 1984a). CMV P1 and P2 are very similar in size to the corresponding proteins of BMV, which have molecular weights of 109,000 and 94,000, respectively (Ahlquist *et al.*, 1984a).

## **3. Translation initiation signals in CMV RNAs.**

The "scanning model" for initiation of protein synthesis in eukaryotes (Kozak, 1980b, 1984, 1987a) proposes that initiation occurs at the first AUG from the 5' terminus provided that it is surrounded by a favourable consensus sequence. The

most recent comparison of 699 vertebrate mRNAs (Kozak, 1987a) revealed that this consensus sequence is  $^{-6}\text{GCC}(\text{A/G})\text{CCAUGG}^{+4}$ , with the most highly conserved features being a purine, usually an A, at position -3 (i.e., 3 nucleotides upstream of the AUG) and a G at position +4. Systematic site-directed mutagenesis experiments (Kozak, 1986a, 1987b) have confirmed the importance of nucleotides in positions -1 to -6 and shown that positions -3 and +4 have the strongest influence over the efficiency of translation.

Comparing the putative CMV RNA 1 and 2 initiation codons with this consensus sequence, only the AUG initiating the P2 reading frame has a preferred A at position -3, although it also has an unfavourable A at position +4. Both the upstream (presumably non-functional) AUG of RNA 2 and the P1 initiation codon are poor matches with the consensus and both have a highly unfavourable C at position -3. The 5' non-coding regions are also notably lacking in C residues, which occupy most upstream positions in Kozak's consensus sequence.

Recent studies of translation initiation in plant systems, together with deviation of many plant viral initiation sequences from Kozak's consensus (see Discussion) have led to the suggestion that the mechanism of translation initiation may differ in plants and animals.

#### **4. Codon Usage in the P1 and P2 coding regions.**

Codon usage in the P1 and P2 reading frames of CMV is summarized in Table 2.3A and Table 2.4A, respectively. The distribution of degenerate codons in both cistrons is distinctly non-random. In the third position of four-fold degenerate codons, U is strongly favoured and A and G appear to be avoided (Figure 2.3B and 2.4B), compared to the frequency expected from the nucleotide composition of the coding regions. A similar situation has been reported for many viral RNAs including those of BMV (Ahlgquist *et al.*, 1984a).

#### **5. Homology between CMV RNAs.**

Sequence homology between the genomic RNAs of CMV was initially examined using the computer program DIAGON (Staden, 1982b). DIAGON homology plots comparing CMV RNA 1 with each of RNAs 2 and 3 are shown in

Table 2.3

**TABLE 2.3: Codon usage and nucleotide composition in CMV RNA 1 and the P1 reading frame.**

**A.** Codon usage in the P1 reading frame of RNA 1. Stop codons are indicated as \*.

First Position	Second Position				Third Position				
	U	C	A	G					
U	Phe	28	Ser	29	Tyr	15	Cys	17	U
		22		19		14		10	C
	Leu	16		14	*	0	*	1	A
		20		7	*	0	Trp	12	G
C	Leu	17	Pro	13	His	11	Arg	20	U
		13		8		14		7	C
		9		5	Gln	18		6	A
		14		7		11		4	G
A	Ile	27	Thr	28	Asn	18	Ser	8	U
		13		15		11		3	C
		10		13	Lys	34	Arg	10	A
	Met	31		12		30		5	G
G	Val	33	Ala	39	Asp	38	Gly	25	U
		15		16		26		12	C
		9		9	Glu	30		13	A
		19		8		23		8	G

**B.** The percentage nucleotide composition of the non-coding and coding regions of RNA 1 and the proportion of each nucleotide occurring in the third position of four-fold degenerate codons (i.e., those encoding pro, thr, val, ala and gly).

Nucleotide Frequency (%) in	U	C	A	G
Whole RNA	29.9	21.7	25.3	23.1
5' non-coding	42.3	24.7	21.7	11.3
3' non-coding	27.8	23.1	24.4	24.7
P1 Coding region	29.8	21.4	25.5	23.3
Third codon positions	44.9	21.5	16.0	17.6

**TABLE 2.4: Codon usage and nucleotide composition in CMV RNA 2 and the P2 reading frame.****A.** Codon usage in the P2 reading frame of RNA 2. Stop codons are indicated as \*.

First Position	Second Position				Third Position				
	U	C	A	G					
U	Phe	36	Ser	19	Tyr	19	Cys	13	U
		21		22		7		5	C
	Leu	15		9	*	0	*	1	A
		24		8	*	0	Trp	4	G
C	Leu	16	Pro	16	His	8	Arg	17	U
		6		14		10		6	C
		3		13	Gln	18		5	A
		11		7		10		6	G
A	Ile	17	Thr	18	Asn	15	Ser	14	U
		9		14		12		8	C
		7		8	Lys	18	Arg	8	A
	Met	19		9		28		5	G
G	Val	26	Ala	22	Asp	37	Gly	21	U
		25		9		28		3	C
		7		10	Glu	19		7	A
		9		8		24		7	G

**B.** The percentage nucleotide composition of the non-coding and coding regions of RNA 2 and the proportion of each nucleotide occurring in the third position of four-fold degenerate codons (i.e., those encoding pro, thr, val, ala and gly).

Nucleotide Frequency (%) in	U	C	A	G
Whole RNA	30.8	22.9	23.7	22.6
5' non-coding	41.3	26.1	18.5	14.1
3' non-coding	30.0	23.9	22.0	24.1
2a Coding region	30.4	22.7	24.2	22.7
Third codon positions	40.7	25.7	17.8	15.8

Figure 2.4. Regions of homologous sequence, which appear as dots along the diagonal of the plot, are only apparent at the 3'-termini of all RNAs and the extreme 5'-termini of RNAs 1 and 2 but not RNA 3. The nature and possible implications of these homologies are discussed below. Apart from these regions, no other extended nucleotide sequence homology could be detected between CMV RNAs 1, 2 and 3 using DIAGON or other dot matrix comparative methods (data not shown).

### **6. The 3' Non-coding regions.**

The 3' non-coding regions of RNAs 1 and 2 are 316 and 423 nucleotides long, respectively (excluding the stop codons) and, as suggested by the DIAGON plot (Figure 2.4), are extensively conserved. Alignment of the 3'-terminal sequences (Figure 2.5) showed that homology extends 307 residues from the 3' end of RNA 1 and stops abruptly at that point. This confirms and extends the results of Symons (1979) who reported extensive homology in the 3'-terminal 270 nucleotides of all CMV RNAs. The sequences of RNAs 1 and 2 are extremely similar in this region with only 11 differences in approximately 300 nucleotides. The 3'-terminal 138 nucleotides of all 3 RNAs are almost identical showing only 1 or 2 differences. Thereafter, however, RNA 3 has two regions (nucleotides 139-171 and 212-257 from 3' end) which have very little homology with RNAs 1 and 2, separated by 40 nucleotides which are identical (Symons, 1979; Figure 2.5).

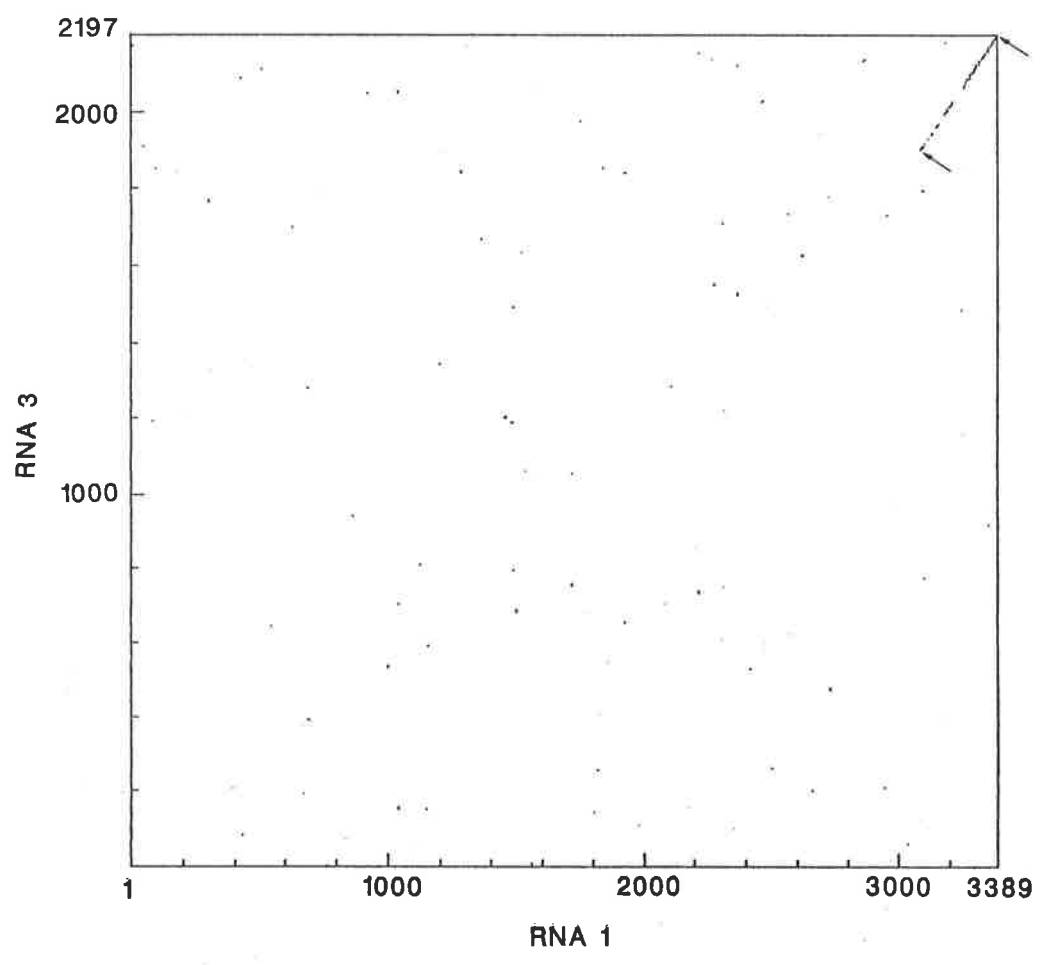
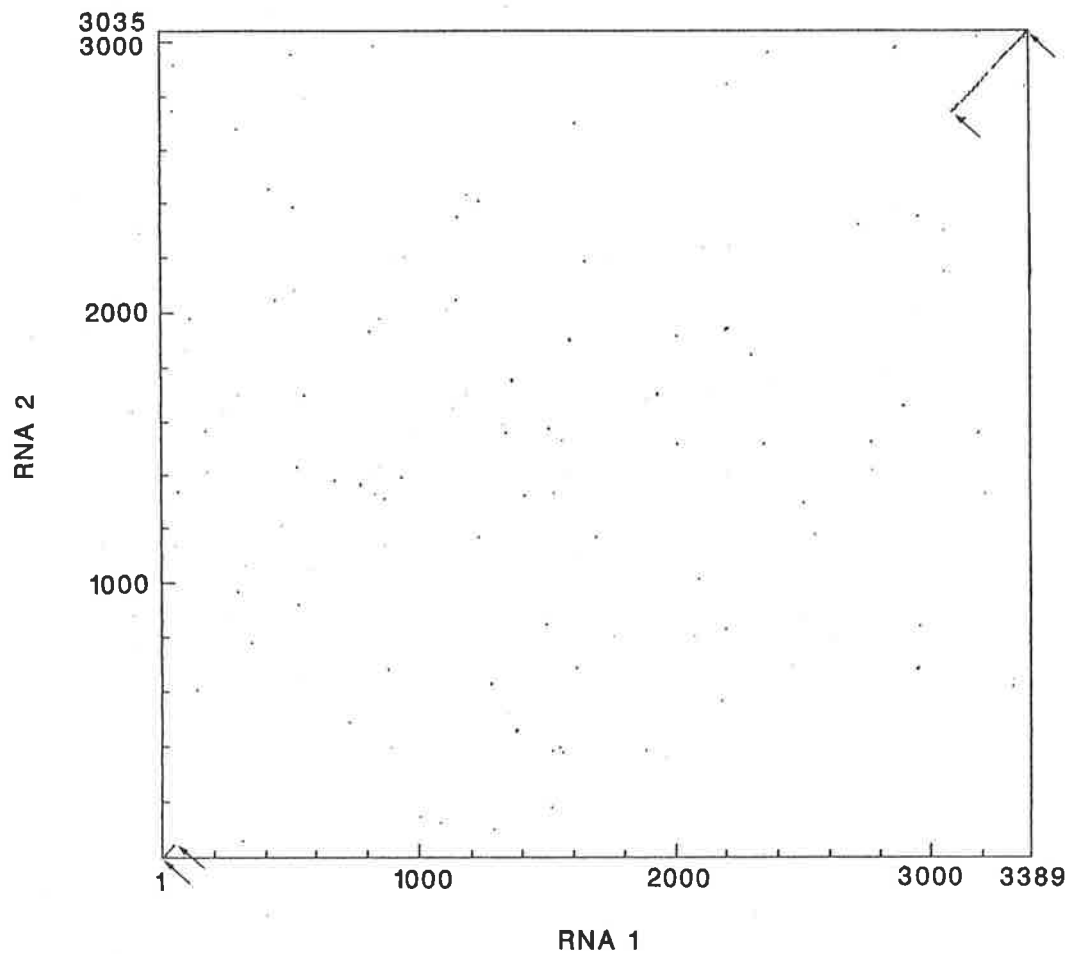
The 3'-terminal 130-190 nucleotides can be folded into a tRNA-like secondary structure which is conserved not only in all CMV RNAs, but also in those of the bromoviruses BMV, CCMV and BBMV (Symons, 1979; Ahlquist *et al* 1981a) and is consistent with the susceptibility of all these RNAs to aminoacylation with tyrosine (Kohl and Hall, 1974).

In addition, the region between nucleotides 130 and 277 from the 3' end of RNAs 1 and 2, immediately following the proposed tRNA-like structures, can form two hairpin loops with high predicted stability (Figure 2.6). Differences in the sequence of RNAs 1 and 2 do not significantly affect the stability of this structure. The two regions of sequence variation in RNA 3 occur in similar locations in each



**FIGURE 2.4: DIAGON homology plots comparing CMV RNAs 1, 2 and 3.**

The nucleotide sequences of RNAs 1 and 2 (upper panel) and of RNAs 1 and 3 (lower panel) were compared using the program DIAGON. The span length was 13 nucleotides and the score was set to detect only those matches with a double-matching probability (see Methods) of less than  $1 \times 10^{-5}$ . The arrows indicate the limits of the regions of homology.



**FIGURE 2.5: Alignment of the 3'-terminal sequences of CMV RNAs.**

The sequences of RNAs 2 and 3 have been aligned with the entire 3' non-coding region of RNA 1, from the UGA stop codon of the 1a reading frame to the 3' end. Numbering is from the 3' end of RNA 1. Homology with the RNA 1 sequence is indicated by a continuous line which is broken only by those nucleotides which are different. Gaps introduced to maximise alignment are indicated as ( $\nabla$ ). The two blocks of sequence enclosed in boxes are those regions where RNA 3 shows reduced homology with RNAs 1 and 2. The heavily underlined sequences are involved in formation of the hairpin loop structures shown in Figure 2.6.

**FIGURE 2.6: Proposed secondary structure in the 3' non-coding region of CMV RNAs 1 and 2.**

The sequence shown is that of RNA 1 between 130 and 277 nucleotides from the 3' end (underlined in Figure 2.5). Changes from this sequence in RNA 2 are indicated by arrows. The boxed sequences are those regions, also boxed in Figure 2.5, which show variability between RNAs 1, 2 and 3. The two residues enclosed in parentheses are found in only one of the RNAs. C(151) is present only in RNA 1 whereas G(159) occurs only in RNA 2.



of these hairpins (Figure 2.6). Because of the large number of sequence changes, these structures cannot form in RNA 3.

Similar hairpin structures do not exist in the RNAs of BMV or AMV which have 3'-terminal conserved regions of approximately 200 and 140-150 nucleotides respectively (Ahlquist *et al*, 1984a; Koper-Zwarthoff *et al*, 1979; Gunn and Symons, 1980).

### **7. The 5' Non-coding regions.**

The 5' non-coding regions of CMV RNAs 1 and 2 both start with a m<sup>7</sup>G cap structure (Symons, 1975) and are of similar length— 97 and 92 nucleotides, respectively. They are both particularly rich in U residues and lacking in G whereas the 3' non-coding regions and the RNAs as a whole do not show this biased base composition (Tables 2.3 and 2.4). A DIAGON comparison of RNAs 1 and 2 showed that the 5' regions have considerable nucleotide sequence homology (Figure 2.4). Alignment of the sequences (Figure 2.7) shows that extended homology is confined to the first 55 nucleotides and the residues surrounding the initiating AUG. The 5' non-coding region of RNA 3 is similar in length to that of RNAs 1 and 2 (95 nucleotides, Davies and Symons, 1988) but does not share extended sequence homology in this region.

Within this homologous 5' sequence is a block of 12 nucleotides (bracketed in Figure 2.7A) which was also found to be conserved at a similar location in BMV RNAs 1 and 2 (Figure 2.7B). A "weight matrix" (Staden, 1984) was constructed from the alignment of these four sequences (Figure 2.7B) and used in a computer program (SCAN, see Methods) to examine the occurrence of this conserved region in other sequences. The sequence was not found in a computer-generated random sequence of 180,000 nucleotides (equivalent to 12.5 times the CMV genome) with the same base composition as CMV RNA 1 and was not present in the genomes of AMV or TMV. It was found to occur only in those regions of CMV and BMV shown in Figure 2.7B.

The region surrounding this conserved sequence, in both CMV and BMV RNAs, can be folded into a hairpin loop structure in which most of the conserved

**FIGURE 2.7: Alignment of the 5' non-coding regions of CMV RNAs 1 and 2.**

**A:** The sequence at the 5' end of RNA 2 is aligned with the first 120 nucleotides of RNA 1. Homology between RNAs 1 and 2 is indicated by (\*). Gaps (-) have been included to maximize alignment. The initiating AUG codons are underlined. The sequence enclosed in brackets is conserved between BMV RNAs 1 and 2 also (see below).

**B:** Alignment of the conserved sequence, bracketed in **A**, in CMV and BMV RNAs 1 and 2 and the complementary strand of Q-CMV satellite RNA. The boxes enclose identical residues.

**FIGURE 2.8: Proposed secondary structure at the 5' end of CMV (A) and BMV (B) RNAs 1 and 2 and in Q-CMV satellite RNA (C).**

The structures shown in **A** and **B** are for RNA 1. Arrows indicate changes from this sequence in RNA 2. The residues in parentheses are present only in RNA 1. The boxed sequences are those conserved between CMV and BMV as shown in Figure 2.7. The residues in bold type in satellite RNA (**C**) can base pair with CMV RNA 1.



residues are located in the loop-out region (Figure 2.8).

The intercistronic regions of CMV and BMV RNA 3 (nucleotides 1101 and 1100, respectively) contain closely-related sequences which can be aligned to give 13 out of 14 matches with each other and 10 out of 12 with the sequence in RNAs 1 and 2 (Davies and Symons, 1988). The RNA 3 sequences, however, cannot form the hairpin loops found for RNAs 1 and 2 (Figure 2.8).

### **8. Structure conserved between CMV and BMV RNAs and CMV satellite-RNA.**

Comparison of the non-coding regions of the CMV genomic RNAs with the sequence of satellite RNA (sat-RNA) from Q-CMV (Gordon and Symons, 1983) revealed a block of 18 nucleotides (residues 27-43 of sat-RNA) which was complementary to the 5' conserved region of CMV RNAs 1 and 2. The region of complementarity included the 12 nucleotides conserved between CMV and BMV RNAs 1 and 2 (see Figure 2.7B). Furthermore, this region of sat-RNA was able to form a secondary structure similar to that found in the genomic RNAs (Figure 2.8C). Hence the 3' end of complementary sat-RNA (residues 297-310) contains a sequence which is conserved in the 5'-non-coding region of RNAs 1 and 2 of its helper virus and of BMV. This sequence, and its predicted secondary structure is almost completely conserved in the ten other CMV sat-RNAs for which nucleotide sequence data are available (Richards *et al*, 1978; Collmer *et al*, 1983; Hidaka *et al*, 1984; Avila-Rincon *et al*, 1986; Garcia-Arenal *et al*, 1987).



## DISCUSSION.

The sequencing of RNAs 1 and 2 described here completes the sequence of the entire Q-CMV genome. The structure of the genome, as determined from nucleotide sequencing studies (this work, Davies and Symons, 1988), is summarized diagrammatically in Figure 2.9. RNAs 1 to 4 each contain a single long open reading frame encoding the proteins P1 (Mr 110,790), P2 (Mr 94,333), 3a (30,353) and CP (Mr 24,247), respectively. Homology between CMV RNAs is restricted to the 5' and 3' non-coding regions. It is these regions which would be expected to contain primary and/or secondary structural features responsible for the control of viral processes.

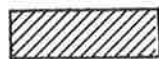


### **The CMV-encoded proteins.**

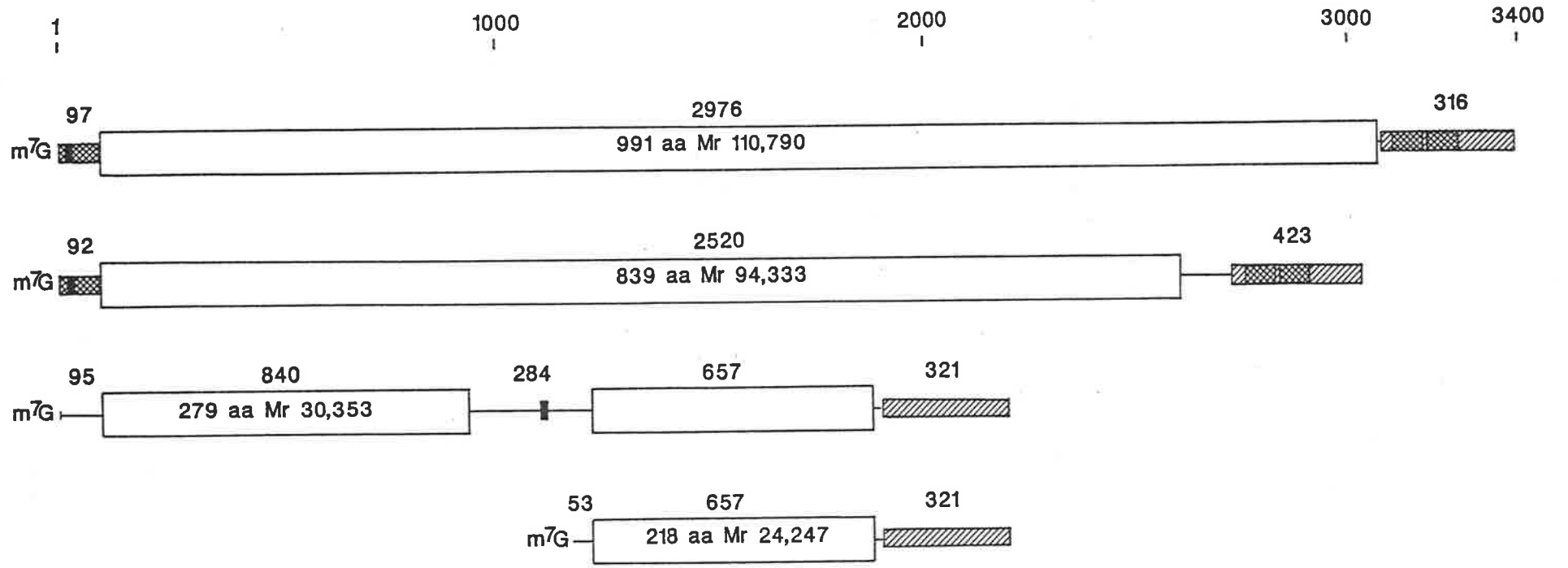
With the exception of the coat protein, encoded by subgenomic RNA 4, the functions of the proteins encoded by CMV are unknown. The completion of the nucleotide sequences of RNAs 1 and 2 has enabled comparisons of the predicted amino acid sequences of the P1 and P2 proteins with other viral proteins. These comparative studies and their implications are the subject of the next chapter.

As well as the long reading frames encoding the P1 and P2 proteins, RNAs 1 and 2 both contain AUG-initiated reading frames on both the genome (+) sense and complementary (-) sense strands, which could potentially encode smaller additional proteins. Similar reading frames have been found in most viral RNAs so far sequenced, although none have been shown to have any *in vivo* function. Computer analysis of the sequences of nuclear, phage and bacterial structural genes revealed that many of these also had the potential to encode polypeptides longer than 100 amino acids from their complementary strands (Casino *et al*, 1981). Van Vloten-Doting *et al* (1983) reported the presence of an open reading frame in a similar position in the (-) strand of both AMV and BMV RNA 4. However, using an *in vitro* system to produce (+) and (-) AMV RNA 4 transcripts from cloned cDNA, Nelson and Loesch-Fries (1987) showed that (-)RNA 4 had no messenger activity in several different translation systems. In the absence of any

**FIGURE 2.9: The genome structure of cucumber mosaic virus.**

The genome structure of Q-CMV has been summarized diagrammatically, based on the nucleotide sequence data for RNAs 1, 2 (this work) and 3 (Davies and Symons, 1988). Open boxes represent the putative proteins encoded by RNAs 1 (P1), 2 (P2), 3 (3a) and 4 (CP). Their predicted number of amino acids (aa) and calculated molecular weights (Mr) are included. The numbers above the lines refer to the number of nucleotides in each region. The scale is in nucleotides.

-  highly-conserved approximately 300-residue region at the 3'-terminus of all RNAs
-  regions of extensive homology between RNAs 1 and 2 which are absent from RNA 3, i.e., the entire 5' non-coding regions and the two blocks in the 3' non-coding region which form the hairpin loop structures shown in Figure 2.6.
-  block of 12 nucleotides conserved in the 5' non-coding regions of RNAs 1 and 2 (Figure 2.7) and also found (with lesser homology) in the RNA3 intercistronic region.



further evidence to the contrary, it therefore seems very unlikely that CMV or closely related viruses utilize translation from both strands as part of their genome strategy.

### **Translation of CMV and other plant viral RNAs.**

Non-random choice of degenerate codons, as observed in the P1 and P2 coding regions of CMV RNAs 1 and 2 (Table 2.3 and 2.4), is not peculiar to the genes of plant viruses but rather appears to be an important part of the translational strategy of most genomes. Analysis of synonymous codon choice in different organisms (Grantham *et al*, 1981; Bennetzen and Hall, 1982; Maruyama *et al*, 1986; Ikemura, 1985), has revealed that all genes of a particular organism have a characteristic pattern of codon choice which differs between taxonomically distinct organisms. In unicellular organisms the codon usage is determined primarily by the availability of different tRNAs and correlates with the level of expression of the gene (reviewed in Ikemura, 1985). In multicellular organisms, however, the tRNA populations are more diverse and vary between different cells or tissues (Ikemura, 1985). More information is therefore required on the translational strategies of plant cells, particularly their tRNA populations, to determine the importance of codon usage for the translation of plant viral genes.

The deviation of the sequences surrounding the initiation codons in CMV RNAs 1 and 2 from the vertebrate consensus sequence of Kozak is not unusual among plant viral RNAs. The AUG initiation codons of BMV RNA 3, RNA 3 of three strains of AMV (references in Table 1.1) and the mRNAs encoding the 30k proteins of 2 strains of TMV (Goelet *et al*, 1982; Ohno *et al*, 1984) all contain a pyrimidine in position -3. In vertebrates only 3% of the 699 initiation sites surveyed had this characteristic (Kozak, 1987a). Even more surprisingly, the efficiently-translated coat protein mRNAs of CMV (Davies and Symons, 1988) and TSV (Cornelissen *et al*, 1984) also have "unfavourable" C residues at position -3.

These observations may be in some way explained by the results of a recent study by Lutcke *et al* (1987) which suggest that the factors selecting AUG initiation codons are different in plants and animals. Comparison of the

translation efficiency, in both animal (reticulocyte lysate) and plant (wheat germ) systems, of RNAs differing only in the nucleotide present at position -3, confirmed Kozak's findings that the nucleotide at -3 had a major effect on translation efficiency in animal systems, but found no such effect in the plant system. In addition, comparison of 61 plant mRNA sequences revealed slight differences between plants and vertebrates in their initiation consensus sequences. Most notably, C was not the predominant upstream nucleotide and the requirement for a purine at -3 was not as strict (Lutcke *et al*, 1987).

Several other factors which have been implicated in the control of translation in eukaryotes have been found to have little or no effect on at least some plant viral mRNAs.

The m<sup>7</sup>G cap structure is not required for translation of many plant viral RNAs (for example, AMV RNA 4; Gehrke *et al*, 1983). Gallie *et al* (1987a,b) constructed RNAs containing a portion of the TMV leader sequence connected to reading frames of foreign genes and found significant enhancement of translation using both capped and uncapped mRNA. Cap structures have been reported as essential for the infectivity of *in vitro* transcripts of the BMV and TMV genomic RNAs (Ahluquist *et al*, 1984b; Meshi *et al*, 1986) but this seems to be due to the importance of the cap in preventing exonucleolytic degradation of RNA (Furiuchi *et al*, 1977), rather than a direct influence on translation.

The high A+U content and generally short length of most viral leader sequences (see for example Goelet *et al*, 1982; Symons, 1985 and references therein; Table 2.3, 2.4) reduces their potential to form stable secondary structures. Godefroy-Colburn *et al* (1985) found that the presence of a 5'-proximal hairpin structure in AMV RNA 3 had no effect on its translation efficiency. This agrees with the results of Kozak showing that secondary structural features are not required for initiation (Kozak, 1980a) and that translation is not diminished by the presence of moderately stable hairpin structures in the leader region (Kozak, 1986b). The absence of secondary structure in the 5' non-coding region of AMV RNA 4 correlates with its

independence of a cap structure and accords with the idea that plant viral RNAs may be able to avoid the cap-associated de-stabilization of secondary structure which has been proposed as an essential, preliminary step in translation initiation (Gehrke *et al*, 1983; Godefroy-Colburn *et al*, 1985).

This and other evidence taken together suggests that the unusual features of viral RNAs, and subtle differences in the translational machinery between plants and animals, may simplify the mechanism of plant viral mRNA translation, enabling them to avoid the restrictions imposed on their cellular counterparts.

### **Structure and function of the 3' non-coding regions.**

Probably the most prominent structural feature of the RNAs of CMV is their tRNA-like 3' ends. Aminoacylatable, tRNA-like structures have been found at the 3' termini of RNAs from various virus groups including Cucumo-, Bromo-, Tobamo-, Tymo- and Hordeiviruses (Hall, 1979; Joshi *et al*, 1983b). In CMV and the bromoviruses, these structures share close structural similarities including the same number of hairpins and looped-out regions (Symons, 1979; Ahlquist *et al*, 1981a). As an example, the 3'-terminal structure of CMV RNA 1 is shown in Figure 2.10A. An alternative form of this structure has been proposed (Joshi *et al*, 1983a) which involves pairing of formerly single-stranded loop-out regions separating the hairpins (Figure 2.10B) and results in a three-dimensional "pseudoknotted" L-shaped configuration closely resembling that of tRNAs (Figure 2.10D). In particular this folding causes coaxial stacking of adjacent hairpins to form an aminoacyl acceptor arm at one end of the molecule. Similar structures have been proposed for all viruses having tRNA-like 3' termini and supported by enzymatic cleavage and chemical modification studies (Rietveld *et al*, 1982, 1983; Joshi *et al*, 1983a; van Belkum *et al*, 1987).

Such a high degree of structural conservation implies that the 3'-terminal tRNA-like structures of viral RNAs have (an) important biological function(s). This has been investigated for the RNAs of BMV using a system for the production of biologically active RNAs from cloned cDNA (reviewed in Ahlquist *et al*, 1987) and template-dependent, BMV RNA-specific RNA replicase preparations (Miller

**FIGURE 2.10: Secondary structure models for the 3' termini of CMV RNAs.**

The structures shown are for RNA 1 but they do not vary significantly for the other CMV RNAs. Arrows indicate the positions of sequence changes in RNA 2.

- A. The tRNA-like secondary structure proposed by Symons (1979).
  
- B. An alternative form of the structure bearing a greater resemblance to tRNA, proposed by Joshi *et al* (1983a). The small letters are used to identify individual stems in the structure.
  
- C. Model for the three-dimensional structure of a typical tRNA (yeast tRNA<sup>Phe</sup>) showing the classical L shape.
  
- D. Three-dimensional representation of the structure shown in (B) emphasizing similarity with the structure of tRNA (C). Stems e and c of B form the aminoacyl arm and anticodon loop respectively. Stem b projects out of the side of the molecule. (Adapted from Rietveld *et al*, 1983)





and Hall, 1983). The 3'-terminal 134 nucleotides of BMV RNAs have been shown to contain all of the signals required for initiation of (-)strand RNA synthesis on (+)strand templates (Miller *et al* 1986) and for recognition by aminoacyl tRNA synthetase and tRNA nucleotidyl transferase (Joshi *et al* 1983a). Analysis of the effect of deletions of sections of the tRNA-like structure on aminoacylation and replication *in vitro* and infectivity *in vivo* (Dreher *et al* 1984; Bujarski *et al* 1985, 1986), revealed that structural features of the RNAs required for recognition by RNA replicase are distinct from those controlling aminoacylation. Hence, aminoacylation of the tRNA-like 3'-termini does not appear to be necessary for the initiation of viral RNA replication. A similar conclusion was reached by Joshi *et al* (1986) who showed that interaction of aminoacylated TYMV RNA with eukaryotic initiation factors was not required for initiation by RNA replicase. The fact that RNAs are aminoacylated *in vivo*, at least in the case of BMV (Loesch-Fries and Hall, 1982), and the strong conservation of a three-dimensional structure recognized by aminoacyl tRNA synthetases, implies that aminoacylation serves an important, as yet unidentified, function in the viral life-cycle.

The L-shaped model proposed for the 3'-terminal tRNA-like structure (Figure 2.10D) involves only approximately 130 nucleotides and is connected to the remainder of the genomic RNA via the corner furthest from the aminoacyl arm. This conformation would probably therefore allow the formation of the additional hairpins proposed for the 3' non-coding regions of CMV RNAs 1 and 2 (Figure 2.6), immediately 5' to the tRNA-like structure (residues 130-277 from 3' end). The function of these additional hairpins is not known. Similar structures are not present in the RNAs of BMV, which do however contain small, conserved hairpin structures located about 255 nucleotides from the 3' end of each of RNAs 1, 2 and 3 (Ahlgquist *et al*, 1984a). It is interesting that the two regions of the 3' non-coding region of CMV RNA 3 showing the greatest variation from RNAs 1 and 2, correspond to the stems of both hairpins and would prevent their formation in RNA 3. These structures could therefore be important in the differential control

of RNA 3 compared to RNAs 1 and 2. They bear an overall resemblance to some of the hairpins in AMV RNAs which are known to bind coat protein (reviewed in Jaspars, 1985) and to those proposed for the origin of assembly of TMV (Zimmern, 1977, 1983). This accords with the suggestion of French and Ahlquist (1987) that sequences in a similar region of BMV RNA 3 (approximately 160-200 nucleotides from the 3' end) may have a role in encapsidation. However, since signals for encapsidation must be present on all CMV RNAs, and the mechanism of assembly of the rod-shaped TMV (reviewed in Lomonosoff and Wilson, 1985) differs from that of icosahedral viruses such as CMV, these similarities may simply reflect the general characteristics of an RNA structure which functions by interacting with proteins.

### **Structure and function of the 5' Non-coding regions.**

The 5' non-coding regions of viral RNAs would be expected to contain primary or secondary structural signals controlling translation (see above) and initiation of (+) strand RNA synthesis at the 3' end of the (-) strand. Hairpin loop structures analogous to those present at the 5' ends of CMV and BMV RNAs 1 and 2 (Figure 2.8) have been proposed for similar locations in several other viral RNAs (e.g., Ravelonandro *et al.* 1983; Strauss and Strauss *et al.* 1983; Gustafson *et al.* 1987). The presence of a 5' hairpin structure in AMV RNA 3 has been substantiated by site-specific enzymatic cleavage (Ravelonandro *et al.* 1983) and shown to have no effect on translation of this RNA (Godefroy-Colburn *et al.* 1985). The fact that the block of 12 nucleotides conserved between the CMV and BMV RNAs and CMV sat-RNA (see also below) is present in the looped-out region of the proposed structure (Figure 2.8), suggests that both the primary sequence and the secondary structure of this region may have functional significance, possibly in initiation of (+) strand RNA synthesis. CMV and BMV RNA 3 contain a similar sequence in their intercistronic regions (Davies and Symons, 1988) and a sequence with even weaker homology occurs in the 5' non-coding regions of RNA 3 of BMV only (French and Ahlquist, 1987). Analysis of deletion mutants of BMV RNA 3 revealed that a section of the intercistronic region including this conserved

sequence was required for accumulation of RNA 3, suggesting that it may indeed have a role in the initiation of (+) RNA synthesis (French and Ahlquist, 1987). Since the conserved sequence is absent from the 5' end of CMV RNA 3, initiation on (-) RNA 3 template would require recognition by the RNA replicase at the internal (intercistronic) conserved sequence distant from the point of initiation. Internal polymerase recognition sites of this type are known to occur in Q $\beta$  RNA (Meyer *et al.* 1981).

The conserved sequence in the RNA 3 intercistronic regions (at least in BMV) is not involved in initiation of subgenomic RNA 4 synthesis (Miller *et al.* 1985; French and Ahlquist, 1987).

### **Interaction between the CMV genome and CMV sat-RNA.**

Very little is known about the primary and secondary structural features of sat-RNAs which are important for their biological properties. The effects of sat-RNA on CMV symptom development appear to be due to its interaction with the CMV genome and with the host plant (Francki, 1985b and references therein) and can vary from attenuation to induction of lethal necrosis (Kaper, 1982). Despite extensive nucleotide sequence homology between all CMV sat-RNAs so far sequenced, there are no conserved open reading frames (Garcia-Arenal, 1987) suggesting that sat-RNAs exert their biological effect at the nucleic acid level.

Primary and secondary structural conservation between CMV RNAs 1 and 2 and sat-RNA (Figure 2.8) suggests the possibility of functionally significant interactions between the satellite and the helper virus genome. For example, the conserved sequences in CMV (+)RNAs 1 and 2 and (-)sat-RNA could compete for a function required for viral RNA replication. If this region is involved in the initiation of (+) strand synthesis, as discussed above, then both (+)CMV RNAs and (-)sat-RNA may be recognized by the RNA polymerase. Alternatively, sat-RNA could base-pair with the conserved region in the genomic RNAs, thereby influencing viral and sat-RNA replication. Analysis of the interactions between sat-RNA and CMV genomic RNAs *in vitro* (Rezaian and Symons, 1986) did not detect base pairing in this region, although it may still occur *in vivo*, especially if

stabilized by interaction with a protein such as an RNA polymerase. Hybridization was however detected between sat-RNA and a region of the coat protein gene on RNAs 3 and 4 (Rezaian and Symons, 1986).

Interaction between sat-RNA and CMV RNAs 1 and 2, but not RNA 3, is consistent with studies of the effect of sat-RNA on CMV infections. The presence of sat-RNA decreases total virus yield and the proportion of RNAs 1 and 2 produced relative to RNA 3. (Kaper and Tousignant, 1977; Francki, 1985b). Furthermore, studies of different CMV strains and their pseudorecombinants have indicated that the rate of sat-RNA replication is controlled by CMV RNAs 1 and/or 2 (Mossop and Francki, 1979).

The sequences in sat-RNA involved in interactions with the genomic RNAs, both potential (this work) and observed (Rezaian and Symons, 1986), are conserved between all CMV satellites so far sequenced, although no correlation has been found between these or any other structural features of CMV sat-RNAs and their pathological characteristics (Garcia-Arenal *et al*, 1987). *In vitro* transcripts of cloned CMV sat-RNAs (Collmer and Kaper, 1986; Kurath and Palukaitis, 1987) will be particularly useful to determine those regions of the satellite which are important for its biological role and interaction with the helper virus genome.

### **Summary: Similarities and differences between CMV RNAs.**

From the foregoing discussion it will be apparent that certain structures are conserved in the 5' and 3' non-coding regions of all CMV RNAs (Figure 2.9) implicating their involvement in control of replication or encapsidation. The structures of RNAs 1 and 2 in particular are remarkably similar. The absence of certain structural features from RNA 3 may be important in the differential control of its expression. Differences between the RNAs at the level of translation would also be expected to have a role in the co-ordinated expression of viral proteins. The approaches which could be used to study the role of these structural features in CMV genome function are discussed in Chapter 5.

There are very strong overall similarities between the genome structures of

CMV and BMV, suggesting that these viruses have a close evolutionary relationship. It is therefore appropriate to extrapolate the results of detailed studies of BMV genome function to include CMV. Conversely, it will be interesting to determine those features of the genome of these two viruses which are responsible for their vastly different biological properties (Chapter 1, Table 1.1). and the fact that no satellite RNAs have been found associated with BMV.

Further similarities between CMV and BMV are evident when the amino acid sequence of their P1 and P2 proteins are compared. These studies and their implications are the subject of the next chapter.

CHAPTER THREE.

NUCLEOTIDE AND AMINO ACID SEQUENCE HOMOLOGY  
BETWEEN CMV AND OTHER VIRUSES.

## INTRODUCTION.

The classification of plant viruses into groups (see Appendix A) and the definition of relationships between them is basically dependent on their physical, chemical, and to a lesser extent, biological properties (Francki, 1983). In recent years, however, more information has become available on the genome structure of viruses, particularly from nucleotide sequencing studies, and this has provided new insight into their inter-relationships.

While the sequencing studies of CMV RNAs 1 and 2 (see Chapter 2) were in progress, Haseloff *et al* (1984) published the results of a study comparing the amino acid sequences of the P1 and P2 proteins of BMV and AMV. Corresponding proteins from each virus showed striking homology, suggesting that they may have similar functions. In addition, the P1 and P2 proteins were homologous to consecutive regions in a protein of Mr 183,000 (183k protein) encoded by tobacco mosaic virus (TMV), which is a monopartite, rod-shaped RNA virus (see Appendix A). The genome of TMV (about 6.4 kb, Goelet *et al*, 1982) encodes four proteins from 3 cistrons. The 5'-proximal cistron, encoding the 126k protein, is terminated with an amber codon which, when suppressed, results in translational read-through to produce the 183k protein. The remaining two genes encode the 30k protein and coat protein which are both translated from subgenomic mRNAs.

This chapter describes the comparison of the nucleotide sequences of CMV RNAs 1 and 2, and the predicted amino acid sequences of the P1 and P2 proteins, with the corresponding sequences from BMV, AMV and TMV. The results extend the relationships reported by Haseloff *et al* (1984) and provide further evidence for an evolutionary relationship between the tripartite viruses and TMV. The similarities between CMV and BMV were found to be particularly close.

Subsequently, similar comparisons have revealed significant homologies between the amino acid sequences of non-structural proteins encoded by plant, animal and even bacterial viruses. The results and implications of these protein

comparisons have been recently reviewed (Goldbach, 1986) and will be discussed further later. Many of the viruses found to be related by protein homologies have very different physical and biological properties and express their genomes by different strategies. Detailed information on the differences between these viruses is beyond the scope of this brief introduction and has been reviewed elsewhere (Francki *et al*, 1985; Strauss and Strauss, 1983; see also Appendix A). In any case, this information is not essential for an understanding of the homology studies which will be discussed below.



## METHODS

### **1. Source and nomenclature of sequences.**

The nucleotide sequences of RNAs 1 and 2 of CMV, BMV and AMV (strain 425-L) were obtained from the original references listed in Table 2.1 and that of TMV (vulgare strain) from Goelet *et al* (1982). All sequences were originally entered into the computer manually, but later confirmed by comparison with appropriate entries in the GenBank sequence database. The amino acid sequences of the proteins encoded by these viruses were derived from their nucleotide sequences. The translation products of RNA 1 and RNA 2 of the Tricornaviridae will be referred to here as P1 and P2, respectively. The products of TMV will be referred to as T1 (the 126k protein) and T2 (the C-terminal read-through region of the 183k protein, following the 126k amber stop codon) as suggested by Haseloff *et al* (1984).

### **2. Dot matrix comparison of nucleotide and amino acid sequences.**

Viral nucleotide and amino acid sequences were compared using the dot matrix program described by Reisner and Bucholtz (1983). Matches of at least 12 nucleotides or 4 amino acids with a minimum homology of 70% were plotted as dots on a graphics screen which was then photographed. Homologous regions were clearly visible as diagonal lines. The program also produced a listing of all matches found, giving their lengths and percentage homologies and showing the alignment of the two sequences.

The statistical significance of matches was assessed by performing repeated dot matrix comparisons between the CMV sequence (nucleotide or amino acid) and a computer-generated random sequence with the same composition as the test sequence. This produced a mean and standard deviation for the number of matches expected at random between the two sequences under the same comparison conditions and therefore allowed a statistical test of the significance of the observed matches (Reisner and Bucholtz, 1983).

### **3. Alignment of amino acid sequences.**

Amino acid sequences of homologous proteins were aligned by visual inspection based on the matches found by the dot matrix program (as above). Gaps were introduced where necessary to facilitate alignment.

### **4. Preparation of hydrophobicity plots.**

Hydrophobicity plots of viral proteins were generated by a computer program (A.H. Reisner, unpublished) using the amino acid hydrophobicity values of Nozaki and Tanford (1971). The value plotted at each point was the average hydrophobicity of the seven amino acids centred at that position.

## RESULTS.

### 1. Dot matrix comparison of CMV RNAs 1 and 2 and their protein products with those of other viruses.

Dot matrix analysis (Reisner and Bucholtz, 1983) was used to compare the nucleotide sequences of CMV RNAs 1 and 2 with the corresponding RNAs of BMV and AMV and with TMV RNA. The translation products of these RNAs were also compared using a similar analysis.

In all comparisons, homologies between sequences were evident as points or lines along the diagonal of the matrix. In some cases, when the sequences being compared showed lower homology, diagonal points were partly obscured by the relatively higher background of random matches. However, matches detected in all comparisons were shown to be statistically significant by testing against the number of matches expected using random sequences (see Methods). The probability of the observed (nucleotide and amino acid) homologies arising by chance was less than 1 in  $10^4$  in all cases.

The results of dot matrix comparisons with CMV RNA 1 and P1 protein are shown in Figure 3.1. Strong homology was evident between RNA 1 of CMV and BMV (Figure 3.1a) while AMV RNA 1 and TMV RNA showed much less homology with CMV RNA 1 (Figure 3.1b and c). Comparison of the translation products of these viral RNAs (Figure 3.1d-f) showed that the P1 proteins of BMV and, to a lesser extent AMV, are homologous to CMV P1, especially in the N-terminal and C-terminal regions. Homology between CMV P1 and TMV T1 proteins was difficult to detect by dot matrix analysis (Figure 3.1f) being apparently restricted to the C-terminal region. However, alignment of these proteins from all four viruses showed that the TMV T1 did share significant blocks of homology with the P1 proteins of CMV, BMV and AMV (see below).

Similar homologies were observed between RNAs 2 of CMV, BMV and AMV and their respective translation products (Figure 3.2). Homology was again greatest with BMV RNA 2 and P2 protein (Figure 3.2a and d), less with AMV RNA

2 and P2 (Figure 3.2b and e) and barely detectable with TMV RNA and T2 (Figure 3.2c and f). However, in contrast to RNA 1, homology with CMV RNA 2 and P2 was limited to the central region corresponding to about 400 amino acids.

A detailed analysis of the dot matrix results (Figures 3.1 and 3.2) revealed that homologous nucleotide and amino acid sequences were co-linear (data not shown).

## **2. Alignment of homologous viral proteins.**

Homologous regions of the CMV, BMV, AMV and TMV proteins were aligned based on the matches detected by dot matrix analysis.

The P1 proteins of CMV and BMV could be aligned over their entire lengths (Figure 3.3) with an overall amino acid identity, including gaps, of 43.8%. The AMV P1 and TMV 126k proteins showed essentially no homology with the central part of CMV P1. However, their C-terminal and N-terminal regions, each of approximately 300 amino acids could be aligned with CMV P1 (Figure 3.3). Thus the P1 and T1 proteins contain two regions of homology – hereafter referred to as the N-P1 and C-P1 domains – separated by a less homologous central region of approximately 300 amino acids. The percentage identity of each region of these proteins with CMV P1 is given in Table 3.1.

The homologous central (approx.) 400 amino acid regions of the P2 proteins of BMV and AMV have been aligned with CMV P2 in Figure 3.4 to give 55% and 37% amino acid identity, respectively (Table 3.2). Beyond this central region, however, homology is very limited (Table 3.2) and the sequences could not be aligned. The TMV T2 protein is 499 amino acids long, considerably shorter than the P2 proteins and hence almost its entire length could be aligned with the central regions of the P2 proteins (Figure 3.4), with 23.6% identity (Table 3.2).

Careful examination of the protein alignments in Figure 3.3 and 3.4 revealed several clusters of amino acids, some as long as 10-12 residues, which are almost completely conserved in the proteins of all three tripartite viruses and of TMV. Hence, homology between CMV and TMV, although difficult to detect by dot matrix analysis (Figure 3.1c,f; Figure 3.2c,f), corresponds to the most

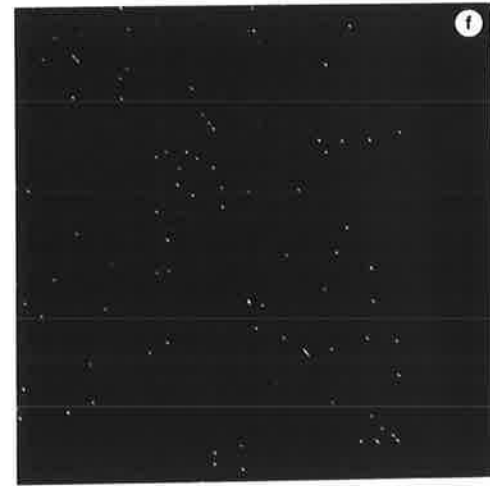
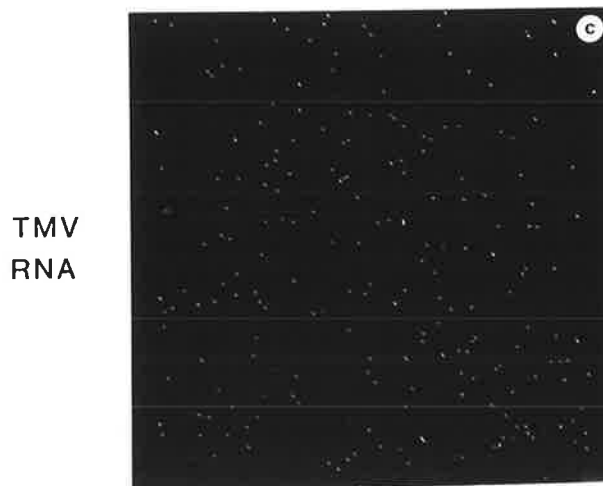
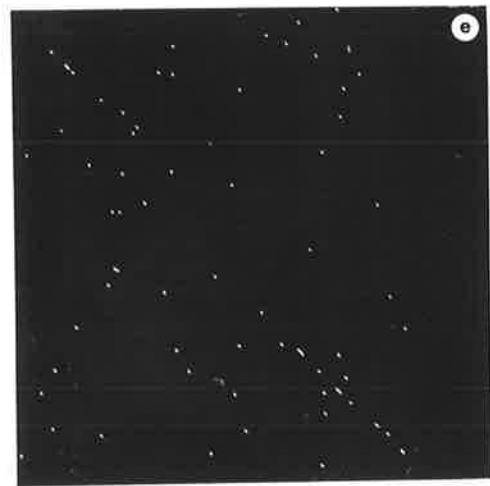
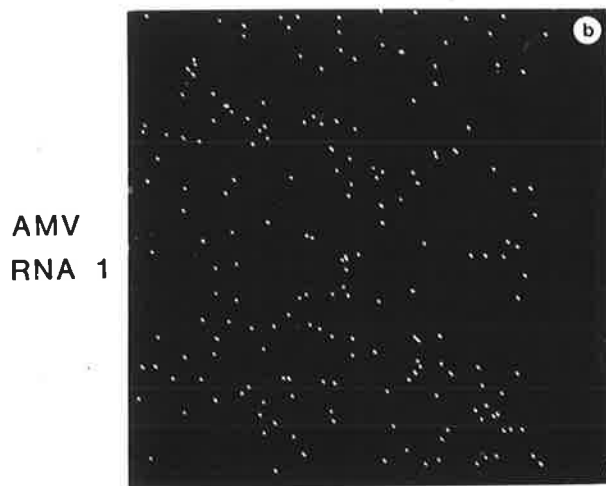
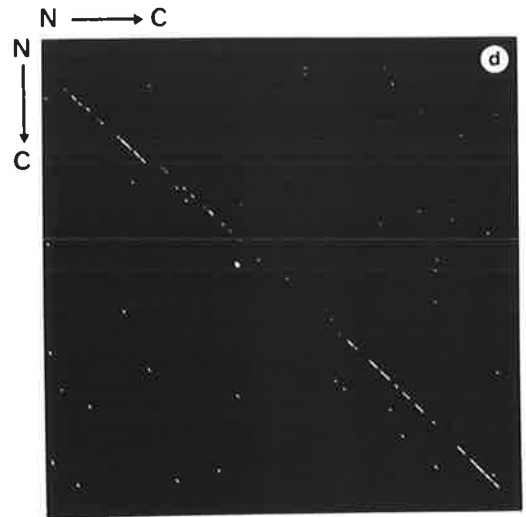
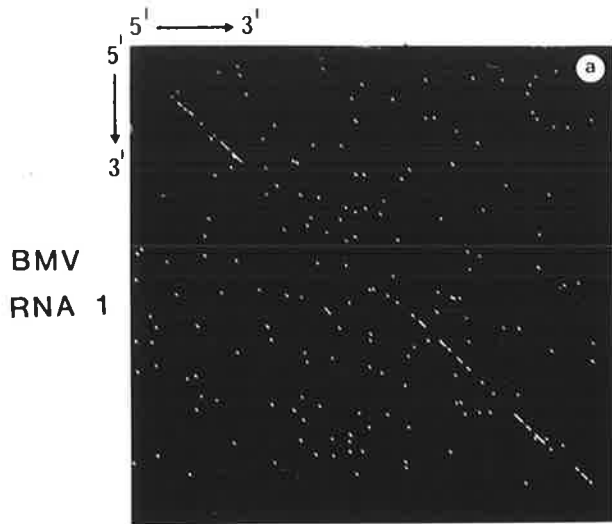
**FIGURE 3.1: Dot matrix comparison of CMV RNA 1 and P1 with the corresponding RNAs and translation products of BMV, AMV and TMV.**

The complete sequences of BMV and AMV RNA 1, and the T1-coding region of TMV (nucleotide residues 1-3419; Goelet *et al.*, 1982), were compared with the complete sequence of CMV RNA 1 (a-c). The amino acid sequences of the corresponding P1 and T1 proteins were also compared with CMV P1 (d-f). Lines plotted in the matrix correspond to matches of at least 12 nucleotides (a-c) or 4 amino acids (d-f) with a minimum homology of 70%. The 5' to 3' direction for the RNAs is shown in (a) and the N-terminal to C-terminal direction for the proteins is shown in (d).

CMV

RNA 1

TRANSLATION PRODUCT

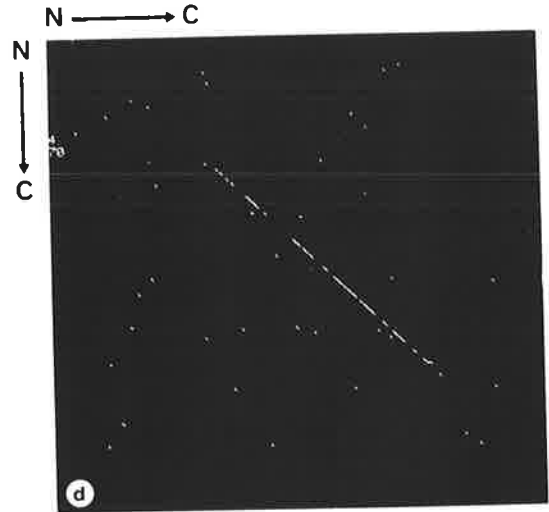
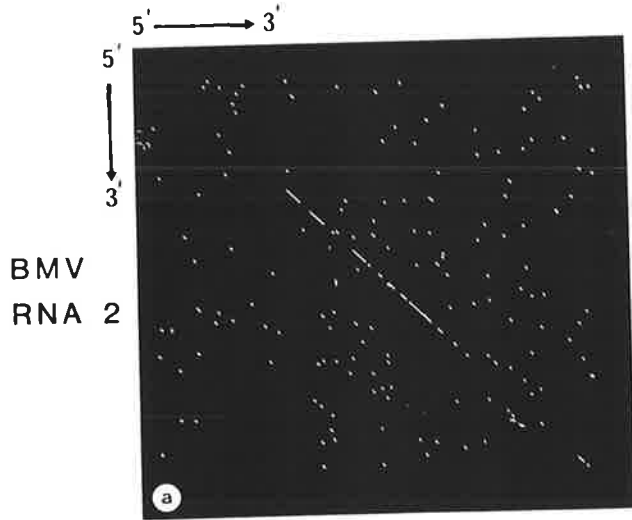


**FIGURE 3.2: Dot matrix comparison of CMV RNA 2 and P2 with the corresponding RNAs and translation products of BMV, AMV and TMV.** The complete sequences of BMV and AMV RNA 2, and the T2-coding region of TMV (nucleotide residues 3420-4919; Goelet *et al.*, 1982), were compared with the complete sequence of CMV RNA 2 (a-c). The amino acid sequences of the corresponding P2 and T2 proteins were also compared with CMV P2 (d-f). Lines plotted in the matrix correspond to matches of at least 12 nucleotides (a-c) or 4 amino acids (d-f) with a minimum homology of 70% (a-e) or 55% (f). The 5' to 3' direction for the RNAs is shown in (a) and the N-terminal to C-terminal direction for the proteins is shown in (d).

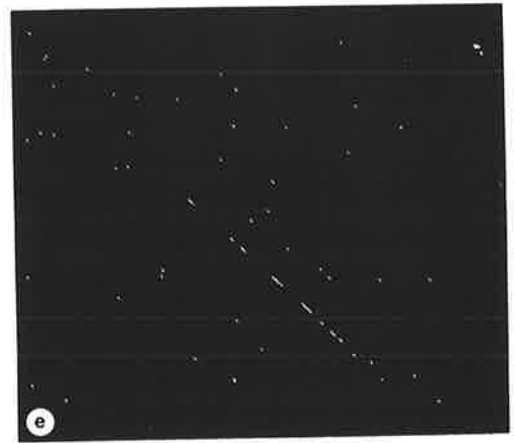
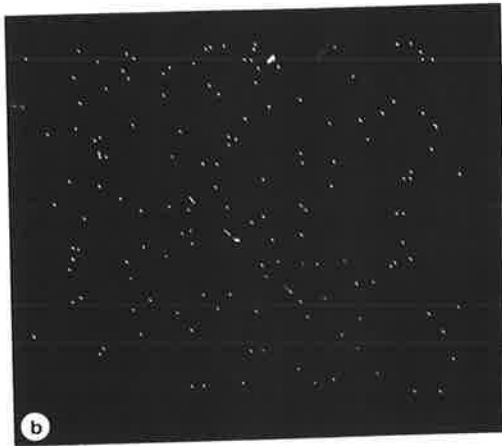
CMV

RNA 2

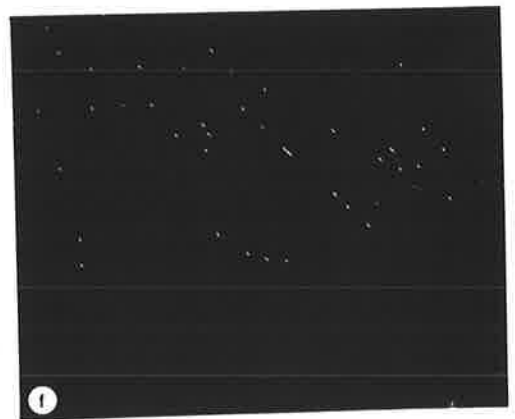
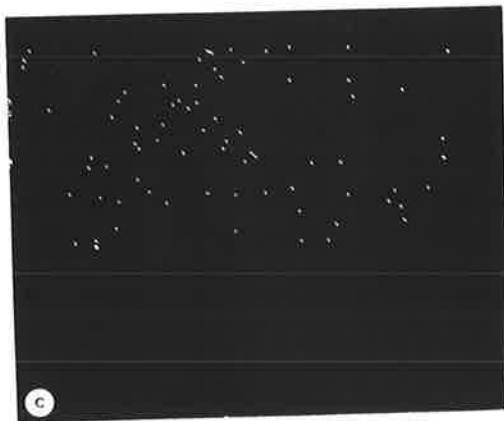
TRANSLATION PRODUCT



AMV  
RNA 2



TMV  
RNA





**FIGURE 3.3: Alignment of the CMV P1 protein with the P1 proteins of BMV and AMV and with the TMV 126k protein (T1).**

Homologies with the CMV sequence are shown as \*. Gaps (-) have been included to maximize the alignment. The numbers refer to the positions of amino acid residues from the N-terminus of each protein. The complete amino acid sequences of CMV P1 and BMV P1 were aligned with the homologous N-terminal and C-terminal regions of AMV P1 and TMV T2. The amino acid residues included in each region are given in the table below.

Protein	Total Length (residues)	Amino acids in region of protein		
		N-terminal	central	C-terminal
CMV P1	991	1-328	329-679	680-991
BMV P1	961	1-298	299-651	652-961
AMV P1	1126	19-316	-	803-1126
TMV T1	1116	26-277	-	800-1116

**TABLE 3.1: Percentage amino acid identity between CMV P1 and the corresponding proteins of BMV, AMV and TMV.**

The percentage identities (including gaps) were calculated from the sequence alignment of Figure 3.3. The N-terminal, central and C-terminal regions are defined in Figure 3.3 (above). TMV T1 and AMV P1 do not show homology in the central region and therefore percentage identity was not calculated (n.c.) for this region.

Protein	Amino acid identity (%) in region of protein		
	N-terminal	central	C-terminal
BMV P1	48.3	30.3	53.3
AMV P1	28.9	n.c.	32.0
TMV T1	20.6	n.c.	28.4

CMV 1 MATSSFNINELVASH-GDKGLLATALVDKTAHEQLEEQHQHQR---RGLKVYIRNVLDVKDSEVIRTRYGGKYDLHLAQQELAPHGLAGALK CETLDCLDFPPRSLGRQDLVLDVDFGGSW  
BMV 1 -MS\*\*IDLLK\*I\*EK--\*ADSQS\*QDI\*\*NQVAQ\*LSA\*IEYAK---\*SK\*INV\*\*K\*SIEEADAF\*D\*\*\*AF\*\*N\*T\*\*YH\*\*\*S\*\*\*\*\*VA\*HY\*\*\*\*\*S\*\*P---E\*P\*I\*\*\*\*\*  
AMV 19 SHA\*IQEMLRR\*VEKQAADTT\*IGK\*FSE\*-GRAYA\*DALPSDKGEV\*\*ISFS---\*\*ATQQNIL\*ANF-PGRRTVFSNSSSS\*CF\*A\*H\*\*L\*\*DFVYRC\*GN---TV\*\*SII\*L\*\*NF  
TMV 26 AYTQTATTSA\*LDTV\*\*NNS\*VNDLAKRRLYDTAV\*\*FNARD\*---\*P--\*\*NFSK\*IS-EEQTL\*A\*\*AYPEFQITFYNTQN\*V\*S\*\*\*G\*\*SL\*-\*EY\*MMQIP---YGS\*TY\*I\*\*NF

CMV 117 VTH-YLRGH-NVHCCSPCLGIRDKMRHTERLMSMRKVI LN DPQQDFGRQPDFCTKSAAECKVQA-----HFAISIHGGYDMGFRGLCEAMNAHGTTILKGTMM  
BMV 112 WH\*-FS\*RDKR\*\*S\*C\*V\*\*V\*\*AA\*\*E\*\*MCR\*\*\*-\*\*QESDD\*\*EV\*N\*\*LNR\*QD\*D\*\*\*-----DW\*\*C\*\*\*\*\*Q\*\*\*D\*\*HS\*\*VRV\*\*R\*\*V\*  
AMV 132 \*S\*MKVKR\*--\*\*\*\*C\*I\*DA\*\*GA\*L\*\*\*IL\*-L\*SYVRKHPEIV\*EA-\*Y\*MDTFQK\*SRR\*-----DY\*FA\*\*STS\*LDVGE\*ACSLDQK\*VMKFC\*\*  
TMV 113 AS\*-LFK\*RAY\*\*\*\*M\*N\*DV\*\*I\*\*\*EGQKD\*IEL-Y\*SRLERGGKTV\*N\*QKEAFDRYAEIPEDAVCHNTFQTMRHQPMQSGRVY\*IAL\*SI\*\*IPADEFGA\*LLRKNVHTCYAAFH

CMV 218 FDGAMFDDQGF-IPELKCQWRKIKSAFSEEDATCSAAKLNSSVFSRVRNGKTLIAFDVFEESTMSYVHDWNIKSFMTDQTYSFNG---MTYGIERCVIYAGVMYKIVGVPGMCPPE  
BMV 207 \*\*\*\*\*L\*\*RE\*\*-L\*L\*\*\*H\*QRDG\*GAD\*-----V\*K\*\*\*EN\*\*\*L\*\*I\*G\*QDLG\*\*F\*ESVHCID\*---T\*\*LL\*\*EMLKCN\*\*\*\*\*IATNLR\*\*R\*  
AMV 227 V\*AD\*L IHNE\*E--\*\*NPNVR\*EIDRKKD-----L\*H\*\*\*ID\*PNLG\*S\*RFSL\*HYL---\*\*NAVDLGHAA\*R\*\*\*KQDFG\*\*\*VIDLTYSL\*FV\*KM  
TMV 231 \*SENLLE\*SYVNL D\*INACFSR DGD-----KLT\*S\*AS\*\*\*LN\*C\*SYS\*\*

CMV 329 LIRHCIWFPMSKDYVGLKIPASDDLKWKTVRILLSTLRETEEIAMRCYNOKKNWMDLFKIILGVLSSKSTIVINGMSMQSGERIDLNDYHYIGFAILLHTKLKYEQLGKMYDMWNAS  
BMV 299 TL\*\*\*V\*\*EDISK\*\*\*VS\*\*EDWS\*NR\*\*C\*\*VAKT\*V\*\*V\*\*\*\*\*F\*\*KES\*E\*TENM\*AVASI\*\*A\*\*\*\*VI\*\*\*QAIMA\*\*\*L\*IE\*\*\*LVA\*\*LT\*NLYQ\*\*\*K\*TALR\*GMEWKG

CMV 449 IWKWFASMSRPFVFFSTVVKTLFPTLRPREEKEFLVKLSTFVTFNEECSDGKKEWDVISSAAFVATQAV---ADGTILAEKAKKLADRLAVPVEEVAIPEV-----SPTPVDQGT  
BMV 419 WCHH\*KTRFWGGDSSRAK\*GW\*R--\*\*AS\*FPLLR\*DSYADSFK\*LTRL\*NVEEF\*Q\*SVPISRLRTFWTEEDLF\*RLHEVQT\*\*TKRSKKKAK\*PPAAE\*\*QEEFHADAPESSSPESVS

CMV 560 ACGLETETSÉLDSLSAQTRSPIARIAERATAMLEYSAYEKQLHDTTVSNLQRIWCMAGGDNKRNSLESNLKVFVFDTYFSDALVNHFPTGRWMHPVPEGVVYVGVYNEKGLGPKLDSEL  
BMV 538 DDVKPVTDVVP\*AEVSEVPTDP\*GIS\*HG\*\*K\*FVR\*C\*R\*\*NNSE\*\*\*RHL\*DIS\*\*RGS---\*IAN\*SI\*E\*\*HRI\*DM\*\*\*\*LAN\*N\*LY\*--KKYD\*T\*\*\*\*\*H\*\*\*\*\*HAD\*T

CMV 680 YIVNGDCVISNSHD--LFSITKSLAPTGTISQVDGVAGCGKTTA KSMFNPST---DIIVTANKKSAQDVRYALFKSTDSKEACAF-VRTADSILLNDCPTVS---RVL-VDEVVLLHF  
BMV 652 \*\*\*DKT\*AC\*\*LR\*--IAEASAKVSV\*\*CD\*\*\*M\*\*\*\*\*DA\*RMGE---\*L\*\*\*\*R\*\*\*E\*\*\*M\*\*\*PD\*YNSKVALDV\*\*\*\*\*AIMHGV\*SCH---\*L\*--\*\*AG\*\*\*Y  
AMV 803 VF\*DQS\*\*FASAEAI PSLEKALG\*EAHFSVTI\*\*\*\*\*N\*\*QIARS\*GRDV\*L\*L\*S\*RS\*\*DELKETIDC\*PLT\*LHYI---\*\*C\*\*Y\*MSASAVKA---QR\*IF\*\*CF\*Q\*A  
TMV 800 SV\*YS\*MAKLRTL R--RLLRNGEPHVSSAKVVL\*\*\*P\*\*\*\*\*KE\*L\*RV\*FDE---\*L\*LVPG\*QA\*EMI\*RRAN\*SGIIV\*TKDN\*K\*V\*\*FMM\*FGKSTRCQFKR\*FI\*\*GLM\*\*T

CMV 790 GQLCAVMSKHLHAVRALCFGDSEQIAFSSRDASFDMRFS--KLIPEDETSADTTFRSPQDVVPLVRLMATKALPKGTRTKY-SDGAQSKVRKSVTS---RÁVASVSLVELDPTRFYITMTQ  
BMV 763 \*\*\*LV\*\*AALSCKSQV\*A\*\*\*T\*\*\*S\*K\*\*\*\*G\*KLLHG--N\*QY\*RRDVVHK\*Y\*C\*\*\*\*--IAAVNLL\*RKCGNRD\*\*\*Q\*WTSE\*\*\*SR\*L\*K---\*RIT\*GLQ\*TI\*\*N\*T\*L\*\*\*\*  
AMV 917 \*LVY\*AATLAGCSEVIG\*\*\*T\*\*\*P\*V\*\*NP\*\*VF\*HH---\*TG-KVERKLI\*W\*\*\*\*A\*A-----\*YC\*E\*YFYKNKKPVKTN\*R\*LR\*IEV---VPIN\*PVS\*\*RNTNAL\*LCH\*\*  
TMV 914 \*CVNFLVAMSLCEI\*YVY\*\*TQ\*\*PYIN\*VSG\*PYP AHFA\*\*EV\*\*VETRR\*\*L\*C\*A\*-----\*HY\*NRRYEGFV---MST\*S\*K\*\*\*SQEMVGA\*VINPISKPLHGKIL\*F\*\*

CMV 904 ADKASLITRAKELNLPKAFYTDRIKTVHESQISEDHVTLVRLKSTKCDLFKKFS-----YCLVAVTRHKVTRFYEYCGVLGGDLIANCIPLV<sup>COOH</sup>  
BMV 876 \*\*\*\*\*A\*Q\*\*\*\*DFPVS\*DWIDGH\*\*\*\*\*A\*\*\*\*\*V\*N\*\*\*\*\*HEE-----\*\*\*\*\*L\*\*\*\*\*K\*S\*E\*CFN\*E\*A\*\*\*\*F\*\*VK<sup>COOH</sup>  
AMV 1023 \*E\*\*\*V\*KAQTHLK-----GC\*N\*F\*T\*\*A\*\*KTF\*N\*YFC\*\*TR\*STS\*ATGRDPINGPCNG\*\*\*LS\*\*\*K\*\*K\*FTIAHSD\*V\*Y\*ACRDAGNTDSDSILARSYNHNF<sup>COOH</sup>  
TMV 1023 S\*\*EA\*LS\*-----GYSVDH\*\*\*\*V\*\*ETYS D\*S\*\*\*\*TP\*PVSI IAGD\*P-----HV\*\*\*\*LS\*\*TCSLK\*YTVVMDPLVS\*IRDLEKLSSYLDDMYKVDAGTQF<sup>COOH</sup>

**FIGURE 3.4: Alignment of the CMV P2 protein with the P2 proteins of BMV and AMV and with the read-through region of the TMV 183k protein (T2).**

Homologies with the CMV sequence are shown as \*. Gaps (-) have been included to maximize the alignment. The numbers refer to the positions of amino acid residues from the N-terminus of each protein. Only the central homologous region of CMV P2, BMV P2 and AMV P2 and essentially all of TMV T2 have been aligned. The amino acid residues included in the central region of each protein are given in the table below.

Protein	Total Length (residues)	Amino acids in central conserved region
CMV P2	839	288-694
BMV P2	822	289-644
AMV P2	790	301-714
TMV T2 <sup>†</sup>	499	34-448

<sup>†</sup> numbered from first amino acid following the 126k amber stop codon

**TABLE 3.2: Percentage amino acid identity between CMV P2 and the corresponding proteins of BMV, AMV and TMV.**

The percentage identities (including gaps) were calculated for the central homologous regions aligned in Figure 3.4. The terminal regions were compared by extending this alignment towards the (non-homologous) N- and C-termini. TMV T2 corresponds only to the central region of the P2 proteins and therefore identity could not be calculated for the terminal regions.

Protein	Amino acid identity (%) in region of protein		
	N-terminal	central	C-terminal
BMV P2	11.8	55.1	13.1
AMV P2	7.3	37.0	10.5
TMV T2	—	23.6	—

CMV 288 LPTHGNYDDSFHQVFVDSADYSTDMDHVRLRQSDLVAKI-PDGGHMLPVLNTGSGHQVRVGTKEVLTAIKKRNADVPELGDSVNLRSLSKAVAERFRLSYMNV-DALAKSNFVNVVSNFH  
BMV 239 \*\*\*\*AYF\*\*\*Y\*\*AL\*ENG\*\*\*M\*F\*RI\*\*K\*\*\*VDWYRD\*\*KYFQ-\*KM\*I\*\*AQR\*\*\*\*Q\*\*\*\*\*L\*\*\*\*\*M\*\*AI\*MKDTA\*\*I\*K\*\*\*STFL\*\*\*GEDCLRASMD\*MTKCL  
AMV 301 \*\*L\*HSI\*\*KYF\*EW\*ETS\*K\*L\*V\*PC\*IDL\*VFNNWQ-SSENCYE\*RFK\*\*ALST\*K\*\*QT\*A\*L\*\*\*\*\*MN\*\*N\*\*QIY--DVN\*V\*NSVVNK\*LTTVI\*-PDKLCM\*PDFI\*EGE  
TMV 34 \*\*GNSTMMNN\*DA\*TMRLT\*I\*LNVKDCI\*DM\*K\*S\*\*APKDQIKPLI\*MVR\*AAEMP\*QTGLL\*N\*V\*MI\*\*\*FNA\*\*\*SGII--DIENT\*SLVVDKFFDSYLLKEKR\*P\*KNVSLFSRE

CMV 406 A-YMQKWPSGLSYDD-L-PD-LHAEN-LQFYDHMIKSDVKPVVDTLNVDRPVPATITFHKKTI-TSQFSPLFISLFR-FQRCLRERVLPVGK-ISSLEM--TGFSVLNKHC-LEID  
BMV 358 E-\*HK\*\*---\*KHM-\*-\*QGVNVA\*\*TD\*CR\*Q\*\*L\*\*\*\*\*HLE\*A\*A\*\*\*\*\*S\*G-V\*\*N\*\*\*F\*TAC\*\*K-LSLA\*KS\*FIV\*I\*\*-----LKNVR\*\*NRYF\*\*A\*  
AMV 417 VS\*F\*DYIV-\*KNP\*PE\*YS\*P\*GVRSDS-\*K\*\*\*\*\*VL\*\*\*EDNS\*HLE\*\*M\*\*\*\*\*Y\*M D\*VM\*S-\*I\*LAAAA\*-LMLI\*\*DKITI\*S\*\*--FHQ\*FSIDAEAFDASF\*-FK\*\*\*  
TMV 152 SLNRWLEKQEQVTIGQLADF\*FVDLPAVD\*-R\*\*\*\*\*AQP\*QKLDTSIQTEY\*ALQ\*\*VY\*S\*\*INAI-\*G\*\*\*SE\*TRQLLDSVDSS\*FLFFTR\*TPAQI\*DFFGDLDShVPMdv\*\*L\*

CMV 515 LSKFDKSQGEFHLMIQEHILNDLGCPAP-ITKWWCDFHRFSYIKDKRAGVGMPI SFQRRITGDAFTYFGNTIVTMAEFAWCYD TDQFD--RLLFSGDDSLAFSKLPPV-GDPSK-FTTLFN  
BMV 466 \*\*\*\*\*L\*\*EF\*RE\*\*LA\*\*F\*\*\*-L\*N\*\*S\*\*\*\*D\*\*LS\*PH\*K\*\*\*SV\*\*\*\*\*L\*\*\*\*\*MI\*YAS\*LSDC\*--CAI\*\*\*\*\*I\*\*VK\*\*L\*TDM\*\*S\*\*\*  
AMV 531 F\*\*\*\*\*N\*L\*HL\*\*\*RF\*KY\*\*I\*NEFL\*L\*-FNA\*\*K\*R\*S\*SKN\*\*FFNVD\*\*\*\*\*L\*\*L\*\*\*\*\*L\*CLCHV\*\*LMDPNVKFVVA\*\*\*\*\*IGTVEELP-R\*QEFL\*\*\*\*\*  
TMV 270 I\*\*Y\*\*\*\*\*CAVEYE\*WRR\*L\*F-EDFLGEV\*KQG\*\*KTTL\*\*YT\*\*IKTC\*WY\*\*KS\*\*VTTFI\*\*\*VIIA\*CL\*SMLPMEKIIK--GA\*C\*\*\*\*\*LYFPKGCEFP\*VQHSANLMW\*

CMV 630 MEAKVMEPAVPIYCSKF--YSLM----S-LVTRFQSPTIREIQRLGTTKIPYSDNNDFLFAHFMSFVDRKLF  
BMV 581 \*\*I\*\*\*D\*S\*\*\*V\*\*\*LVETE\*----GN\*\*SV-PD\*-L\*\*\*\*\*AKR\*\*\*--LRDEQM\*R\*\*\*V\*\*C\*\*\*M\*\*  
AMV 649 L\*\*\*FPHNQ-\*F\*\*\*\*\*LI-TMPTTSGGKV\*LPIN\*LKLL\*-\*\*\*S\*\*V---NADI\*DEWYQSWIDIIGG\*  
TMV 387 F\*\*\*L\*FKKQYGF\*FCGRYVI-HHDRGCIVYY-----D\*LKLIS-K\*\*A\*HI--K\*WEHLEEFRR--LC\*VAVS

strongly-conserved regions when the amino acid sequences are aligned. Moreover, the degree of similarity between even the least homologous of the aligned regions (N-P1 domains of CMV and TMV; 20.6%, Table 3.1) is above that expected for alignment of random amino acid sequences (Doolittle, 1981). Taken together, these data therefore provide strong evidence that these proteins are indeed structurally related.

### **3. Similarities in the properties of the conserved proteins.**

Further similarities between the P2 proteins of CMV, BMV and AMV were evident when the distribution of acidic and basic amino acids in each protein were examined (Table 3.3). Within the central conserved region defined in Figure 3.4, there are roughly equal numbers of acidic and basic amino acid residues. In contrast, the N-terminal regions of all three proteins have a higher proportion of acidic residues, whereas the C-terminal regions show a preference for basic residues (Table 3.3). The P1 and P3 (3a) proteins of CMV, BMV and AMV do not have a similar distinct distribution of acidic and basic amino acids (results not shown). However, the coat proteins of all these viruses show some similarities in their distribution of charged amino acids, since they all contain basic N-termini (Davies and Symons, 1988). Clusters of basic amino acids have been found in several viral coat proteins and may be important in protein/RNA interactions (Harrison, 1984). The presence of a cluster of basic amino acids at one end of the P2 proteins suggests that they may also interact with RNA and is consistent with their proposed role in RNA replication (see Discussion).

Comparison of the hydrophobicity profiles of CMV and BMV proteins provided further evidence for a close similarity between these viruses. The N-P1 and C-P1 domains of the CMV and BMV P1 proteins both contained consecutive peaks and troughs in similar locations of the hydrophobicity plot (Figure 3.5A) whereas the profiles were different in the less conserved central domain. The P2 protein of these viruses also had very similar profiles, but only in their central conserved domain (Figure 3.5B). Hydrophobicity plots for the proteins of AMV and TMV could not be aligned in this way.

**TABLE 3.3: Distribution of acidic and basic amino acid residues in the P2 proteins of CMV, BMV and AMV.**

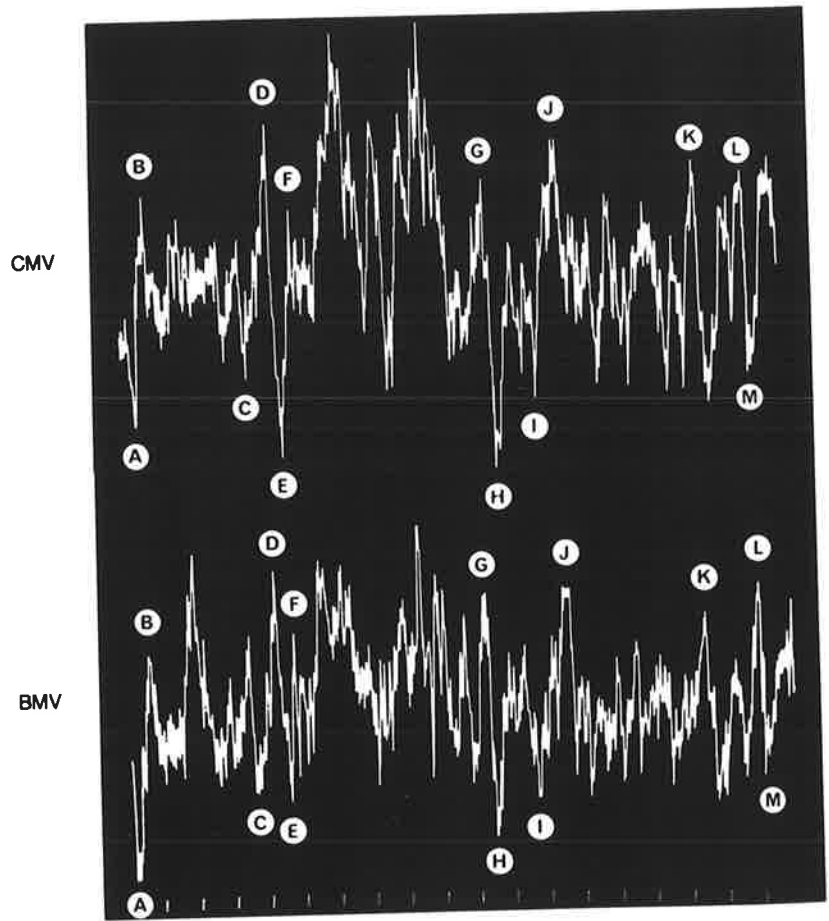
The total number of acidic (A; asp, glu) and basic (B; lys, arg) amino acids and their ratio was determined in each of the regions defined by the alignment of Figure 3.4.

P2 Protein of	Region of protein					
	N-terminal		central		C-terminal	
	A+B	A/B	A+B	A/B	A+B	A/B
CMV	69	2.29	90	1.00	42	0.56
BMV	63	3.2	108	1.00	57	0.78
AMV	75	2.75	96	1.23	25	0.39

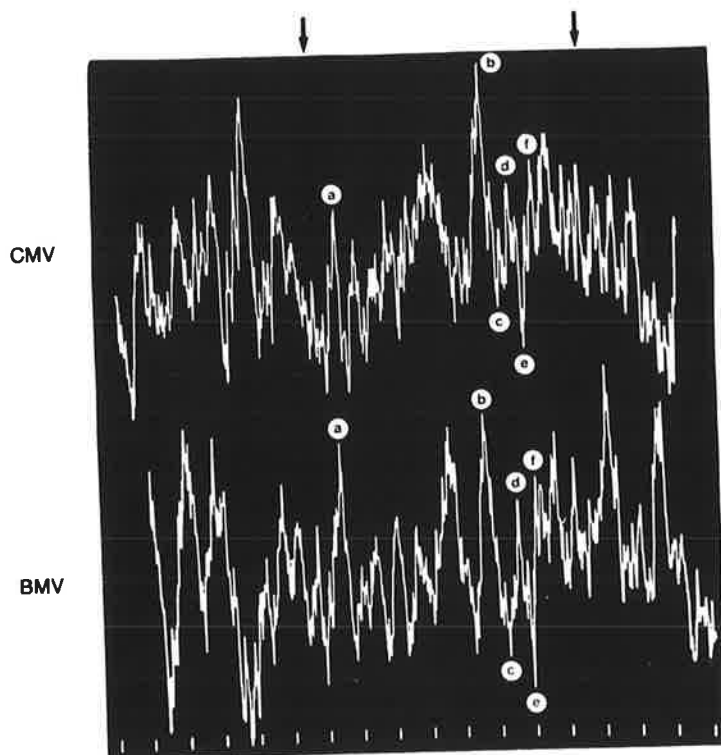
**FIGURE 3.5: Hydrophobicity plots of the P1 (A) and P2 (B) proteins of CMV and BMV.**

Each division of the scale corresponds to 50 amino acids. Similarities in the plots are identified by letters A-L in **A** and a-f in **B**. The P2 plots were aligned according to the amino acid alignment of Figure 3.4. The arrows indicate the boundaries of the central homologous regions as defined in the legend to Figure 3.4.

A



B





## DISCUSSION

The results of dot matrix comparison of the CMV P1 protein with other viral proteins (Figure 3.1) has revealed two domains of homology with the P1 proteins of BMV and AMV and with the TMV 126k protein (Figure 3.3). Similarly, the CMV P2 protein contains a central domain which is homologous to the corresponding proteins of BMV, AMV and to the TMV 183k readthrough region (Figures 3.2 and 3.4). The nucleotide sequences encoding these proteins are also conserved (Figures 3.1 and 3.2) although to a lesser extent than the amino acid sequences. The relative locations of these conserved domains within the genome of each virus are shown schematically in Figure 3.6. The separate RNAs of the tripartite genomes have been drawn end-to-end to emphasize the conserved order of the protein domains in all viruses.

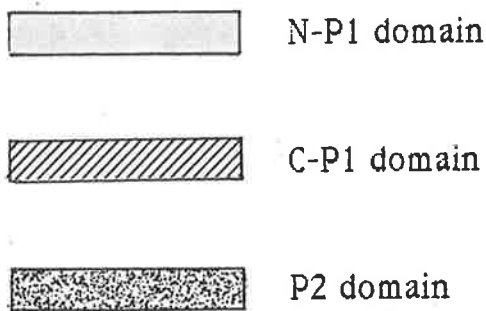
The amino acid sequence alignments in Figures 3.3 and 3.4 essentially concur with those of Haseloff *et al* (1984) for the proteins of BMV, AMV and TMV, and extend them to include CMV. The methods used to detect homology were slightly different in each case. Haseloff *et al* (1984) searched for amino acids which were identical or structurally similar, whereas the procedure used here revealed only identical amino acids. Although the former approach produces dot matrix plots with lower 'background', Doolittle (1981) has suggested that amino acid identity is the most significant indicator of relationships between proteins.

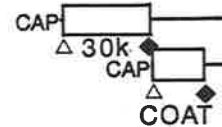
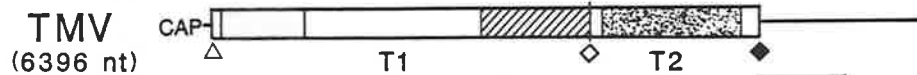
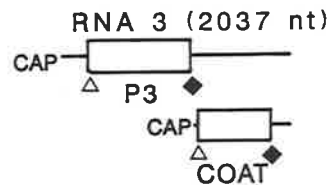
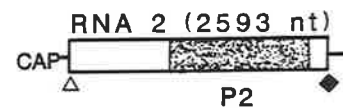
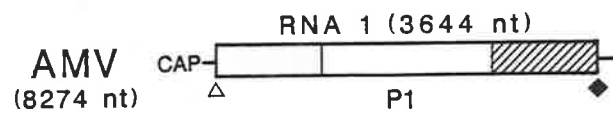
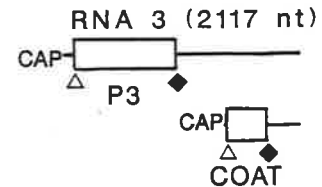
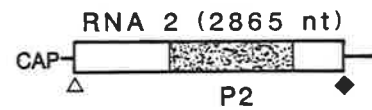
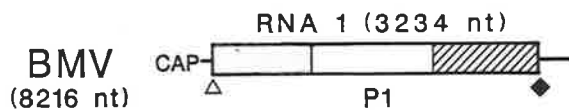
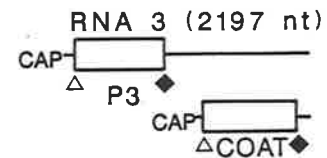
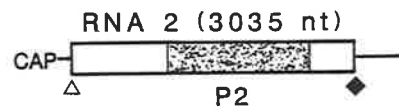
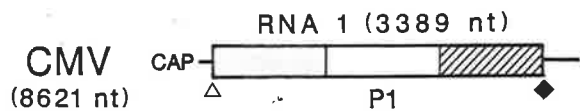
The similar distribution of acidic and basic amino acids in the P2 proteins of all three Tricornaviridae (Table 3.3) provides further evidence for their similarity. In addition, hydrophobicity profiles of the P1 and P2 proteins (Figure 3.5) are remarkably similar for CMV and BMV, although not for the other viruses.

All these results provide strong evidence for an evolutionary relationship between the Tricornaviridae, and between these viruses and TMV. The 3a proteins of CMV, BMV and AMV have also been found to show significant sequence homology (Murthy, 1983; Savithri and Murthy, 1983) although this is not shared by the 30k protein of TMV (Ahlquist *et al*, 1985). In all these

**FIGURE 3.6: Similarities between the proteins encoded by CMV, BMV, AMV, TMV and Sindbis virus.**

The genome of each virus has been drawn schematically on approximately the same scale. The genomic RNAs of each tripartite virus have been drawn end-to-end on the same line, while the subgenomic coat protein (CP) mRNA appears on the line below. The sizes of each RNA and the total length of each viral genome are given in nucleotides (nt). Proteins encoded by each virus are represented as open boxes. Similar shading within these boxes represents regions of amino acid homology. The homologous regions in the proteins of CMV, BMV, AMV and TMV have been derived from the alignments in Figures 3.3 and 3.4. The alignment of these viruses with Sindbis virus and the basic structure of the figure were adapted from Ahlquist *et al* (1985). ( $\Delta$ ): initiation codons; ( $\blacklozenge$ ): termination codons; ( $\diamond$ ): suppressable termination codons.





comparisons, BMV shows the most extensive homology with CMV, while AMV and TMV are progressively less homologous. This can be seen, for example, in the percentage amino acid identity between the aligned sequences of each protein and that of CMV (Table 3.1 and 3.2). Thus it appears that, in evolutionary terms, CMV is most closely related to BMV while AMV and TMV show progressively more divergence.

Protein structure similarities between these viruses could be due to their descent from a common ancestor or to convergent evolution of proteins having similar functions. The results reported here, and subsequent studies by several authors (reviewed in Goldbach, 1986; discussed below), strongly suggest that, despite differences in their genome organization, these viruses have indeed evolved from a common ancestor and that the amino acid sequences of their translation products have been conserved as a result of selection pressure to maintain functions vital to the survival of the virus (Haseloff *et al*, 1984; Ahlquist *et al*, 1985). Strong evolutionary constraint at the protein level is apparent in the higher degree of sequence conservation at the amino acid level compared to the nucleotide level (Figures 3.1 and 3.2).

#### **Further homologies with the Tricornaviridae and TMV.**

In addition to homology between the proteins of BMV, AMV and TMV, Haseloff *et al* (1984) also detected similarities between the P2 homologous domains from these viruses and a non-structural protein, designated ns72 or nsP4, encoded by the animal alphaviruses, Sindbis virus and Middleburg virus. These viruses are similar to TMV in that they have a monopartite single-stranded (+)RNA genome and employ readthrough of a suppressable stop codon as part of their genome strategy. Thus the protein nsP4, produced by readthrough of the preceding cistron (nsP3) in alphaviruses, is homologous to the readthrough region of the TMV 183k protein (Figure 3.6). Analysis of the amino acid sequences of other non-structural proteins of Sindbis virus (Ahlquist *et al*, 1985), revealed homology between the nsP1 and nsP2 proteins and the N-P1 and C-P1 domains respectively in TMV and the Tricornaviridae. Hence, all these viruses

share homologous proteins which are arranged in a similar linear order along their genomes (Figure 3.6).

Even more recently, several other viruses have been found to encode proteins which are homologous to the P1 and P2 conserved domains. The genomic RNA of carnation mottle virus (CarMV; Guilley *et al*, 1985) and RNA 1 of tobacco rattle virus (TRV; Hamilton *et al*, 1987) both contain suppressable stop codons and express proteins which are homologous to the P2 conserved domain. A similarly homologous protein is encoded by RNA  $\gamma$  of barley stripe mosaic virus (BSMV; Gustafson *et al*, 1987) and RNA 1 of beet necrotic yellow vein virus (BNYVV; Bouzoubaa *et al*, 1987) although these are not expressed as readthrough products. Furthermore, CarMV and BNYVV also encode proteins which contain regions with homology to the C-P1 domain, although the N-P1 domain appears to be absent. Thus, viruses with diverse genome structures and strategies, and classified in vastly different virus groups (see Appendix A), contain proteins which are homologous and occur in similar relative locations within their genomes.

#### **Homologies between other plant viral proteins.**

Independently of the similarities reported above, extensive homology has been reported between proteins encoded by cowpea mosaic virus (CPMV) and animal picornaviruses such as poliovirus and foot-and-mouth disease virus (Franssen *et al*, 1984; Argos *et al*, 1984). These viruses also have similarities in the structure and mode of expression of their genome and in their capsid morphology (reviewed in Goldbach, 1986). In addition, proteins encoded by tobacco vein mottling virus (TVMV; Domier *et al*, 1986) and tobacco etch virus (TEV; Allison *et al*, 1986) have also been found to share homology with CPMV and the picornaviruses (Domier *et al*, 1987). The proteins conserved in all these viruses are homologous to the 2C, VPg, 3C-protease and 3D-polymerase proteins of poliovirus and are encoded in the same conserved order in each virus. A recent immunocytochemical study (Bienz *et al*, 1987) has suggested that the poliovirus 2C protein is involved in attaching the viral replication complex to vesicular membranes which are the site of RNA replication in infected cells.

Hence all the proteins conserved between the picornaviruses and the aforementioned plant viruses are probably involved in RNA replication.

### **Functional significance of homologous proteins in Tricornaviridae.**

Evidence for the function of the homologous proteins in Tricornaviridae and TMV is much more circumstantial. Infection of protoplasts with RNAs 1 and 2 alone of BMV (Kiberstis *et al*, 1981) or AMV (Nassuth and Bol, 1981; Nassuth *et al*, 1983) results in synthesis of viral RNA, suggesting that functions encoded by these RNAs are involved in replication. This is in agreement with the behaviour of temperature-sensitive (ts) mutants of AMV and CCMV (a Bromovirus) RNAs 1 and 2 which affect replication of viral RNA (Sarachu *et al*, 1985; Huisman *et al*, 1985, van Vloten-Doting, 1985). In addition, Berna *et al* (1986) used antisera against the non-structural proteins of AMV to study their accumulation in extracts from virus-infected tobacco leaves. The amount of P1 and P2 proteins correlated with the appearance of replicase activity, whereas the P3 (3a) protein did not. However, there is no direct evidence for the involvement of TMV- or CMV-encoded proteins in replication.

### **Homology between RNA polymerases from plant, animal and bacterial viruses.**

The strongest evidence for the involvement of the P2 proteins (and their homologues in TMV and Sindbis) in replication has come from a comparison of these proteins with the RNA-dependent RNA polymerases of CPMV and the picornaviruses (Kamer and Argos, 1984). An alignment of all these proteins revealed a pair of conserved aspartic acid residues (D-D) flanked on each side by hydrophobic residues and an additional conserved block about 20-30 amino acids upstream. This 'D-D' domain was present not only in the the plant viral and picornaviral proteins compared, but also in the putative polymerases of bacteriophage MS2, hepatitis B virus and cauliflower mosaic virus and in the reverse transcriptases of several retroviruses (Kamer and Argos, 1984). This same conserved domain has subsequently been found in proteins encoded by many different animal, insect and plant viruses. To illustrate its extreme

conservation, an alignment of this region in the eleven plant viral proteins in which it has been found, together with some animal and insect viral proteins, is given in Figure 3.7. It has been proposed that this region is an active site or recognition site essential to the mechanism of all viral RNA polymerizing enzymes (Kamer and Argos, 1984)

### **Implications for the function of CMV-encoded proteins.**

The results of comparisons between the amino acid sequences of the CMV P1 and P2 proteins and those of other viruses presented in this chapter, together with the results of more recent comparative analyses reviewed above, provide some clues to the possible function of the CMV P1 and P2 proteins. Inclusion of the 'D-D' polymerase domain in the central homologous region of CMV P2 strongly suggests that this protein is an RNA-dependent RNA polymerase, or the core subunit thereof. The close evolutionary relationship between CMV and BMV implies that corresponding proteins from both viruses have similar functions and hence that CMV P1 is also involved in replication. Further understanding of the role of individual proteins in the replication process requires a detailed study of the RNA replicase induced during CMV infection. Investigations described in Chapter 4 were therefore aimed at obtaining more information on the function of CMV-encoded proteins in replication of viral RNA.

The presence of three functional domains (N-P1, C-P1 and P2) on a single polypeptide in TMV, but on two separate polypeptides in the case of CMV, has led to the suggestion that the P1 and P2 proteins may associate *in vivo* to form a biologically functional complex (Ahlquist *et al.* 1985). Studies of mutant tripartite viruses have suggested that the translation products of RNAs 1 and 2 act together to provide functions necessary for RNA replication (van Vloten-Doting, 1985). An association between these proteins is in agreement with the results of studies on pseudorecombinant viruses which have shown that RNA 3 is readily interchangeable between viruses, whereas RNAs 1 and 2 are not (Habibi and Francki, 1974c; Rao and Francki, 1981; van Vloten-Doting, 1985).

A further, and not necessarily alternative, possibility is that a single

**FIGURE 3.7: Alignment of amino acid sequences around the 'D-D' domain in viral RNA polymerases.**

The amino acid (aa) sequences of the 11 plant viral proteins so far shown to contain the conserved 'D-D' polymerase domain have been aligned with putative and known RNA polymerases from some representative animal and insect viruses.

- (A) viruses with homology to the Tricornaviridae;
- (B) viruses with homology to cowpea mosaic virus;
- (C) other animal viruses; and
- (D) the insect virus, black beetle virus (BBV).

The residues in bold type are conserved in all sequences. Those positions marked (▼) are those at which at least 20 of the 22 sequences have amino acids with similar chemical properties, as defined by Domier *et al* (1987); i.e., acidic and polar (D,E,N,Q), basic (K,R), hydrophobic (A,C,F,H,I,L,M,V,W,Y), polar (T,S), or strong turn formers (D,G,N,P).

The proteins have been numbered from their N-termini except for SNBV and MIDV which have been numbered after the opal termination codon.

The alignment of sequences is expanded from that of Hamilton *et al* (1987).

References for the plant viruses in A have been given in the text. Sequences in B are from Domier *et al* (1987). Sources of animal virus and BBV sequences are as cited by Hamilton *et al* (1987).

Abbreviations: Plant virus name abbreviations appear in the table at the front of this thesis. Animal virus names have been abbreviated as follows:

SNBV: Sindbis virus; MIDV: Middleburg virus; SFV: Semliki Forest virus; POLIO: poliovirus; FMDV: foot-and-mouth disease virus; EMCV: encephalomyocarditis virus; COX: coxsackievirus; RHINO 2 and RHINO14: rhinoviruses 2 and 14; Hep A: hepatitis A virus.



Virus Protein Position

**A**

Virus	Protein	Position	Sequence	aa-	Sequence
CMV	P2	570-615	QRR <b>T</b> GDA <b>F</b> T <b>T</b> Y <b>F</b> GN <b>T</b> I <b>V</b> T <b>M</b> A <b>E</b> -11	aa-	RL <b>L</b> F <b>S</b> G <b>D</b> D <b>S</b> L <b>A</b> F <b>S</b> K <b>L</b>
BMV	P2	521-566	QRR <b>T</b> GDA <b>F</b> T <b>T</b> Y <b>F</b> GN <b>T</b> L <b>V</b> T <b>M</b> A <b>M</b> -11	aa-	CA <b>I</b> F <b>S</b> G <b>D</b> D <b>S</b> L <b>I</b> I <b>S</b> K <b>V</b>
AMV	P2	586-633	QRR <b>T</b> GDA <b>L</b> T <b>Y</b> L <b>G</b> N <b>T</b> I <b>V</b> T <b>L</b> A <b>C</b> -13	aa-	F <b>V</b> V <b>A</b> S <b>G</b> D <b>D</b> S <b>L</b> I <b>G</b> T <b>V</b> E
TMV	183k	1442-1487	QR <b>K</b> S <b>G</b> D <b>V</b> T <b>T</b> F <b>I</b> G <b>N</b> T <b>V</b> I <b>I</b> A <b>A</b> C-11	aa-	K <b>G</b> A <b>F</b> C <b>G</b> D <b>D</b> S <b>L</b> L <b>F</b> F <b>K</b>
TRV	194k	1511-1556	Q <b>Q</b> K <b>S</b> G <b>D</b> A <b>D</b> T <b>Y</b> N <b>A</b> N <b>S</b> D <b>R</b> T <b>L</b> C <b>A</b> -11	aa-	M <b>V</b> T <b>Y</b> G <b>G</b> D <b>D</b> S <b>L</b> I <b>A</b> F <b>P</b> R
CarMV	85k	526-569	CR <b>M</b> S <b>G</b> D <b>M</b> N <b>T</b> A <b>L</b> G <b>N</b> C <b>L</b> L <b>A</b> C <b>L</b> I-9	aa-	R <b>L</b> I <b>N</b> G <b>D</b> D <b>C</b> V <b>L</b> I <b>C</b> E <b>R</b>
BSMV	γa	586-631	Q <b>Q</b> K <b>S</b> G <b>N</b> C <b>D</b> T <b>Y</b> G <b>S</b> N <b>T</b> W <b>S</b> A <b>A</b> L <b>A</b> -11	aa-	F <b>C</b> V <b>F</b> G <b>G</b> D <b>D</b> S <b>L</b> I <b>L</b> F <b>D</b> O
BNYVV	237k	1895-1940	V <b>K</b> T <b>S</b> G <b>E</b> P <b>G</b> T <b>L</b> L <b>G</b> N <b>T</b> I <b>L</b> M <b>G</b> A <b>M</b> -11	aa-	C <b>M</b> A <b>M</b> K <b>G</b> D <b>D</b> G <b>F</b> K <b>R</b> O <b>A</b> N
SNBV	nsP4 <sup>a</sup> (op+)	490-537	MM <b>K</b> S <b>G</b> M <b>F</b> L <b>T</b> L <b>F</b> V <b>N</b> T <b>V</b> L <b>N</b> V <b>V</b> I-13	aa-	C <b>A</b> A <b>F</b> I <b>G</b> D <b>D</b> N <b>I</b> I <b>H</b> G <b>V</b> V
MIDV	nsP4 <sup>a</sup> (op+)	490-537	MM <b>K</b> S <b>G</b> M <b>F</b> L <b>T</b> L <b>F</b> V <b>N</b> T <b>M</b> L <b>N</b> M <b>T</b> I-13	aa-	C <b>A</b> A <b>F</b> I <b>G</b> D <b>D</b> N <b>I</b> V <b>H</b> G <b>V</b> K
SFV	nsP4	485-532	MM <b>K</b> S <b>G</b> M <b>F</b> L <b>T</b> L <b>F</b> I <b>N</b> T <b>V</b> L <b>N</b> I <b>T</b> I-13	aa-	C <b>A</b> A <b>F</b> I <b>G</b> D <b>D</b> N <b>I</b> V <b>H</b> G <b>V</b> I

**B**

CPMV	pp <sup>b</sup>	1494-1554	G <b>I</b> P <b>S</b> G <b>F</b> P <b>M</b> T <b>V</b> I <b>V</b> N <b>S</b> I <b>F</b> N <b>E</b> I <b>L</b> -26	aa-	G <b>L</b> V <b>T</b> Y <b>G</b> D <b>D</b> N <b>L</b> I <b>S</b> V <b>N</b> A
TEV	pp	2586-2634	G <b>N</b> N <b>S</b> G <b>Q</b> P <b>S</b> T <b>V</b> D <b>N</b> T <b>L</b> M <b>V</b> I <b>A</b> -15	aa-	V <b>Y</b> Y <b>V</b> N <b>G</b> D <b>D</b> L <b>L</b> I <b>A</b> I <b>H</b> P
TVMV	pp	2531-2584	G <b>N</b> N <b>S</b> G <b>Q</b> P <b>S</b> T <b>V</b> D <b>N</b> T <b>L</b> M <b>V</b> V <b>L</b> A-19	aa-	K <b>F</b> F <b>A</b> N <b>G</b> D <b>D</b> L <b>I</b> I <b>A</b> I <b>S</b> P
POLIO	pp	285-336	G <b>M</b> P <b>S</b> G <b>C</b> S <b>G</b> T <b>S</b> I <b>F</b> N <b>S</b> M <b>I</b> N <b>N</b> L <b>I</b> -17	aa-	K <b>M</b> I <b>A</b> Y <b>G</b> D <b>D</b> V <b>I</b> A <b>S</b> Y <b>P</b> H
FMDV	pp	295-346	G <b>M</b> P <b>S</b> G <b>C</b> S <b>A</b> T <b>S</b> I <b>I</b> N <b>T</b> I <b>L</b> N <b>N</b> I <b>Y</b> -17	aa-	T <b>M</b> I <b>S</b> Y <b>G</b> D <b>D</b> I <b>V</b> V <b>A</b> S <b>D</b> Y
EMCV	pp	2120-2171	G <b>L</b> P <b>S</b> G <b>C</b> A <b>A</b> T <b>S</b> M <b>L</b> N <b>T</b> I <b>M</b> N <b>N</b> I <b>I</b> -17	aa-	K <b>V</b> L <b>S</b> Y <b>G</b> D <b>D</b> L <b>L</b> V <b>A</b> T <b>N</b> Y

**C**

COX		286-337	G <b>M</b> P <b>S</b> G <b>C</b> S <b>G</b> T <b>S</b> I <b>F</b> N <b>S</b> M <b>I</b> N <b>N</b> I <b>I</b> -17	aa-	R <b>M</b> I <b>A</b> Y <b>G</b> D <b>D</b> V <b>I</b> A <b>S</b> Y <b>P</b> W
RHINO14		2003-2054	G <b>M</b> P <b>S</b> G <b>C</b> S <b>G</b> T <b>S</b> I <b>F</b> N <b>S</b> M <b>I</b> N <b>N</b> I <b>I</b> -17	aa-	K <b>I</b> L <b>A</b> Y <b>G</b> D <b>D</b> L <b>I</b> V <b>S</b> Y <b>P</b> Y
RHINO 2		1974-2025	G <b>V</b> P <b>S</b> G <b>C</b> S <b>G</b> T <b>S</b> I <b>F</b> N <b>T</b> M <b>I</b> N <b>N</b> I <b>I</b> -17	aa-	K <b>I</b> I <b>A</b> Y <b>G</b> D <b>D</b> V <b>I</b> F <b>S</b> Y <b>I</b> H
HEP A		2037-2091	S <b>M</b> P <b>S</b> G <b>S</b> P <b>C</b> T <b>A</b> L <b>L</b> N <b>S</b> I <b>I</b> N <b>N</b> I <b>N</b> -20	aa-	R <b>I</b> L <b>C</b> Y <b>G</b> D <b>D</b> V <b>L</b> I <b>V</b> F <b>S</b> R

**D**

BBV	102k	646-700	G <b>V</b> K <b>S</b> G <b>S</b> S <b>T</b> T <b>T</b> P <b>H</b> N <b>T</b> Q <b>Y</b> N <b>G</b> C <b>V</b> -20	aa-	I <b>G</b> P <b>K</b> C <b>G</b> D <b>D</b> G <b>L</b> S <b>R</b> A <b>I</b> I
-----	------	---------	---	-----	--

viral-coded protein may perform more than one function. The fact that some viruses (for example BSMV and BNYVV, see above) encode proteins homologous to the P2 and C-P1 domains but not the N-P1 domain suggests that the two domains of P1 may be functionally distinct. In the case of BSMV and BNYVV, the N-P1 function may not be required or may be fulfilled by another protein with homologous function but not primary sequence.

Finally, since the P2 or "polymerase" domain is conserved between the polyprotein-expressed viruses (picornaviruses and CPMV) and the tripartite viruses (Kamer and Argos, 1984), it is tempting to speculate that other proteins encoded by these viruses may be functionally, if not structurally, similar. Polypeptides encoded upstream of the (putative) polymerases in both the picornaviruses (2C, VPg, 3C-protease) and the tripartite viruses (P1 and P2), are known to be involved in replication. Although CMV does not encode a VPg and is not known to encode a protease, it could encode a protein homologous to the poliovirus 2C, which is responsible for attachment of the poliovirus replication complex to cellular membranes (Bienz *et al.* 1987). Since viral RNA replication is known to occur on a membrane-bound replication complex for most RNA viruses, including CMV (see Chapter 4), it is not unreasonable to suggest that the CMV P1 protein may be involved in binding this complex to the membrane.

CHAPTER FOUR.

CHARACTERIZATION OF CMV-INDUCED  
RNA-DEPENDENT RNA POLYMERASE.

## **INTRODUCTION**

### **1. The replication of RNA virus genomes.**

The replication of viral RNA genomes requires the presence of an enzyme capable of synthesizing RNA from an RNA template. Replication of viral (+) RNA requires specific recognition of the parental (+) strand and the synthesis of a complementary (-) strand which then acts as a template for the production of new (+) viral RNAs (Hall *et al*, 1982). An enzyme capable of replicating RNA in this way is referred to as an "RNA Replicase". However in many studies of viral RNA replication, the enzymes studied have not been true replicases, since they are not capable of completing the replication cycle of (+)-(-)-(+), as outlined above. In such cases a more correct description of the enzyme is an RNA-dependent RNA polymerase (RdRPase) which simply copies RNA without strict constraints on the nature of the final product.

The RNA replicases of many bacterial, animal and plant viruses have been extensively studied and the literature concerning these enzymes is too large to be reviewed in detail here. The brief review which follows is intended only to highlight the most significant studies and the best-characterized examples, with particular reference to involvement of host- and virus-coded proteins in the replication of viral RNAs.

### **2. Bacteriophage Q $\beta$ RNA replicase.**

The classical and well-studied example of a viral RNA replicase is that of bacteriophage Q $\beta$  (reviewed in Blumenthal and Carmichael, 1979). This enzyme is composed of a virus-encoded 65k subunit and 3 host-encoded proteins – ribosomal protein S1 (70k) and protein synthesis elongation factors EF-Tu (45k) and EF-Ts (35k). An additional "host factor" is also required during part of the replication cycle. Many plant and animal viral replicases have features in common with Q $\beta$  replicase making it a useful basic model for understanding the enzymes from different sources.

### **3. Animal viral RNA replicases.**

The RNA replicases of animal viruses, particularly those of picornaviruses and alphaviruses, have been extensively studied (see Strauss and Strauss, 1983). Although it is not necessary to review these studies here, it is interesting to note that animal viral RNA replicases apparently also require both host- and virus-coded proteins (for example, see Andrews and Baltimore, 1986; Lubinski *et al.*, 1987) and occur as membrane-bound complexes (for example, see Friedman *et al.*, 1972; Lazarus and Barzilai, 1974; Traub *et al.*, 1976; Etchison and Ehrenfeld, 1981).

### **4. Plant viral RNA replicases.**

The results of studies on plants infected with many different viruses (reviewed in Hall *et al.*, 1982; Fraenkel-Conrat, 1986) have revealed some features which are apparently common to all plant viral RNA replicases.

The "native" form of these enzymes is a membrane-bound complex comprising viral RNA and proteins of (probably) both host and viral origin. Solubilization of the enzyme from particulate fractions in most cases leads to a change in its properties, presumably due to the loss of polypeptides responsible for regulation and specificity of the replication complex. In addition, all plants so far examined contain a host-coded RNA-dependent RNA polymerases (RdRPases) which is stimulated by virus infection. These enzymes are single polypeptides of Mr 130,000 to 150,000 found usually in the soluble fraction of cells although in some cases they may also be components of the membrane-bound replication complex (The notable exceptions are the RNA replicases of TYMV and CPMV which have been definitely shown not to contain the host RdRPase). As discussed below, the role of the host RdRPase and viral-coded proteins in viral RNA replication differs between replicases induced by different viruses and in most cases has not been fully characterized.

#### **Brome mosaic virus RNA replicase.**

The RNA replicase isolated from BMV-infected barley by Hall and colleagues probably resembles the native membrane-bound enzyme complex more closely

than any other system studied. Template-dependent and BMV RNA-specific RNA polymerase activity was solubilized from particulate fractions using the non-ionic detergents NP-40 (Hardy *et al*, 1979) and dodecyl- $\beta$ -D-maltoside (12-M) (Bujarski *et al*, 1982). It has been suggested that 12-M stabilizes the enzyme by acting as a substitute for glycolipid components of the membrane which are apparently essential for complete enzyme activity (Bujarski *et al*, 1982). Removal of endogenous RNA from these RNA replicase preparations using micrococcal nuclease (Miller and Hall, 1983) rendered them completely dependent on added template and allowed the investigation of structures in BMV RNA required for initiation of (+) and (-) strand synthesis (for example, Miller *et al*, 1985, 1986; Bujarski *et al*, 1985). A 110k protein was present in 12-M-solubilized preparations from BMV-infected but not healthy plants which co-migrated, and shared tryptic peptides, with the translation product (P1) of BMV RNA 1 (Bujarski *et al*, 1982). This and other evidence (see Chapter 3) suggests that BMV P1 protein is a component of the viral RNA replicase.

#### **Alfalfa mosaic virus RNA replicase.**

Using the same 12-M detergent solubilization procedure, Houwing and Jaspars (1986) obtained an RNA replicase preparation from the particulate fraction of AMV-infected plants which was completely template dependent and showed a strong preference for AMV RNA. Similar preparations from mock-inoculated leaves did not contain this replicase activity, although it is not clear whether or not the host RdRPase is a component of the virus-specific replicase complex. This is in contrast to earlier investigations of AMV replicase which identified distinct viral-specific and host-encoded activities but found few qualitative differences between the enzymes present in healthy and AMV-infected plants (Weening and Bol, 1975; Clerx and Bol, 1978; LeRoy *et al*, 1977; Chiffot *et al*, 1980).

#### **Tobacco mosaic virus RNA replicase.**

The particulate fraction of TMV-infected cells contains an RNA replicase capable of synthesizing double-stranded replicative forms and replicative

intermediates of TMV RNA (Bradley and Zaitlin, 1971; Young and Zaitlin, 1986; Watanabe and Okada, 1986) as well as host-coded RdRPase which synthesizes low molecular weight RNAs of mostly (-) polarity. Soluble and solubilized-particulate RNA polymerase appear to be similar, if not identical to the host-coded RdRPase, showing template-dependence but copying a wide variety of templates (Zaitlin *et al.* 1973; Romaine and Zaitlin, 1978). Even though the TMV-encoded 126k and 183k proteins have been implicated in viral RNA replication (see Chapter 3), there is no direct evidence of their involvement as components of TMV RNA replicase.

#### **Turnip yellow mosaic virus RNA replicase.**

The chloroplast membrane has been identified as the site of TYMV RNA replication using high resolution autoradiography and subcellular fractionation (Lafleche *et al.*, 1972) and, more recently, by immunocytochemistry (Garnier *et al.* 1986). Solubilization of the replication complex from chloroplast membranes with a nonionic detergent (Lubrol W) facilitated its subsequent purification using a PEG-Dextran two-phase system and several chromatographic steps to yield a preparation containing two major protein subunits of 115k and 45k (Mouches *et al.* 1984). Antibodies raised against replicase preparations enabled the identification of the 115k and 45k proteins as viral- and host-encoded, respectively (Mouches *et al.* 1984; Candresse *et al.*, 1986). The host-coded RdRPase is present in TYMV-infected cells, but has different properties from, and is not involved in TYMV RNA replicase.

#### **Cowpea mosaic virus RNA replicase.**

Studies on CPMV RNA replicase have been recently reviewed (Goldbach and van Kammen, 1985) and therefore only a summary of the conclusions of these studies will be presented here.

The membrane fraction of CPMV-infected cowpeas contains both the host-encoded RdRPase and the viral RNA replication complex. The host enzyme was solubilized using  $Mg^{2+}$ -deficient buffers and shown to consist of a single protein of Mr 130,000. Triton X-100 was required to solubilize the CPMV RNA replication complex which was then purified and shown to contain a

virus-encoded 110k protein and 68k and 57k proteins of host origin. Hence, like the TYMV enzyme (see above), CPMV RNA replicase is distinct from the host-coded RdRPase. These studies made extensive use of antibodies raised against the host-coded 130k enzyme and resulted in the development of the antibody-linked polymerase assay (ALPA) technique (van der Meer *et al*, 1983) which is generally applicable for the identification of polypeptides with polymerase activity.

### 5. CMV-induced RNA-dependent RNA polymerase.

CMV-induced RdRPase activity was initially detected in crude soluble extracts of infected cucumbers (Gilliland and Symons, 1968; May *et al*, 1969) and later also in the particulate fraction (May *et al*, 1970). The soluble polymerase was purified approximately 100-fold by Clark *et al* (1974) and the chromatographic methods were extended by Kumarasamy and Symons (1979a) to achieve 10,000-fold purification. The major component of this highly-purified preparation was a protein of Mr 100,000 (100k), although several other minor polypeptides were present which were also found in corresponding preparations from healthy plants.

Solubilization of RdRPase activity from the particulate fraction (PF) by incubation with magnesium sulphate enabled subsequent purification of the solubilized, particulate RNA polymerase (Gill *et al*, 1981; Gordon *et al*, 1982) to yield a preparation with similar specific activity and polypeptide composition as the highly-purified soluble enzyme. A protein of 100k co-purified with RdRPase activity, while 110k and 35k proteins were also sometimes seen in trace amounts. The 100k protein was present in extremely small amounts in these preparations and could only be detected by labelling using [<sup>3</sup>H]potassium borohydride in a reductive methylation reaction (Kumarasamy and Symons, 1979b). No RdPPase activity was found in soluble or particulate fractions of healthy cucumber seedlings subjected to the same purification procedures.

Similar polypeptide compositions have been reported for CMV-induced RdRPase preparations obtained in other laboratories. The RNA polymerase



purified by Takanami and Fraenkel-Conrat (1982) contained two proteins of 100k and 112k and a minor component of 50k. Khan *et al* (1986) used a chromatographic purification scheme different to that of Gill *et al* (1981) to obtain RNA polymerase preparations containing a major component of 100k and minor components of 110k and 10k. However, they gave no information on the recovery and purification of enzyme obtained at each step of their procedure, presumably because of its low level. Silver staining (Merril *et al* 1981) was required to detect the small amounts of RNA polymerase proteins present. These authors suggest that the major 100k RNA polymerase is produced by cleavage of the 110k protein, resulting in a 10k fragment.

### **The particulate replication complex.**

Although completely template-dependent, highly-purified CMV RdRPase shows little preference for CMV RNA as template and produces transcripts which are heterogeneous in size, with a mean length of 150 nucleotides (Kumarasamy and Symons, 1979a; Kumarasamy, 1980; Takanami and Fraenkel-Conrat, 1982). In contrast, the particulate fraction (PF) of infected plants contains RNA polymerase activity which is not dependent on added template. Investigation of well-washed PF (Gill, 1983; Jaspars *et al*, 1985) showed that that it was able to synthesize, *in vitro*, RNAs which were largely double-stranded and remained associated with the PF. When denatured, these RNAs had the same length and polarity as the four RNAs of CMV suggesting that the particulate enzyme was elongating pre-initiated positive strands within a replication complex.

The PF also synthesized low molecular weight transcripts of plant RNA and of viral RNA of both plus and minus polarity, which were presumably the products of the 100k protein contained within the PF. The ability of PF to synthesize these low molecular weight products was not removed by treatment with magnesium sulphate which is known to release a large proportion (although not all) of the 100k protein (Jaspars *et al*, 1985).

All of these results suggest that the CMV RNA replicase, like similar enzymes in other systems, is a membrane-associated complex containing several proteins

one of which is the 100k protein shown to have RdRPase activity in highly-purified fractions. Since this protein alone does not have the template specificity or catalytic properties required for an RNA replicase, it must be associated with other proteins, probably viral-coded (see below), which have roles in the regulation and specificity of viral RNA replication.

### **The role of viral- and host-coded proteins.**

If CMV replicase is similar to that from bacteriophage Q $\beta$  and other systems, it will contain both host- and viral-coded subunits. A comparison of the RdRPase induced in cucumbers by infection with CMV and the unrelated tobacco ringspot virus (Peden *et al*, 1972), revealed differences in the time course and properties of the enzymes induced by the two viruses, suggesting that the RNA polymerase was not a host enzyme. Conversely, the close similarity in size of the 100k, 110k and 35k proteins in highly-purified RdRPase preparations (see above), with the *in vitro* translation products of RNAs 1, 2 and 3, respectively, suggests that these RNAs encode components of the replicase. Gordon *et al* (1982) investigated this possibility by comparing the electrophoretic mobilities of the polypeptides in RdRPase preparations from cucumber seedlings infected with different strains of CMV, with the translation products of RNAs from the corresponding CMV strains. While all RdRPase preparations contained the 100k, 110k and 35k proteins, the electrophoretic mobilities of the *in vitro* translation products differed between strains suggesting that these products were not components of the polymerase. This conclusion was supported by differences in the one-dimensional peptide maps produced by cleavage of the 100k and 110k polymerase proteins with *S. aureus* V8 protease and cyanogen bromide, when compared with similar digests of the RNA 1 and 2 translation products. From these results it was concluded that the full-length translation products of CMV RNAs were not present in highly-purified RdRPase preparations, and hence, that the major 100k protein is host-coded.

Several lines of evidence now suggest that this conclusion should be re-examined more closely.

Firstly, amino acid sequence homology between the non-structural proteins of CMV and other viruses, as reported in the previous chapter, reveals that the CMV P2 protein contains a conserved domain present in all viral RNA polymerases, and implies that the CMV P1 protein is also involved in replication. This is further supported by the presence of a protein with similar mobility and tryptic peptides as BMV P1 in RNA replicase preparations from BMV-infected barley (Bujarski *et al*, 1982; see above). Based on similarities in their genome structure (Chapter 2) and non-structural proteins (Chapter 3), CMV and BMV are very closely related. It therefore appears unlikely that CMV P1 and P2 proteins are not components of CMV replicase.

Secondly, Khan *et al* (1986) have reported the presence of trace amounts of RdRPase in healthy cucumbers. Although they did not purify or characterize the enzyme in detail, they estimated its molecular weight by sedimentation on glycerol gradients to be 140,000, suggesting that it is distinct from the CMV-induced enzyme of Mr 100,000.

Finally, the methods used by Gordon *et al* (1982) do not provide unequivocal evidence for the host-coded nature of the 100k protein. Sequencing studies of CMV RNAs 1 and 2 (Chapter 2) revealed large discrepancies between the sizes of their translation products predicted from the nucleotide sequence and estimated by electrophoretic mobility following *in vitro* translation. This discrepancy makes conclusions based on comparative electrophoretic mobilities of CMV-encoded proteins difficult to justify. In addition, although peptide maps of the 100k protein and the RNA 1 translation product contained some fragments differing in size, other fragments were of similar size. In the absence of further evidence, such as two-dimensional peptide mapping results, the non-identity of these two proteins cannot be definitely stated.

#### **Re-investigation of the Mr 100,000 protein.**

In the light of all the above considerations, it is possible that the CMV-induced 100k RdRPase is a viral coded protein. Considering the variation in electrophoretic mobility observed for these proteins, it could be either the P1 or

P2 protein, with or without post-translational modification. Alternatively, the 100k protein may be a host encoded protein present at undetectably low levels in healthy plants but induced by CMV infection. This would not preclude the additional involvement of the (as yet undetected) viral-coded P1 and/or P2 proteins in RNA replicase. It is clear that further characterization of the 100k protein is essential to an understanding of CMV replication.

Although many approaches could theoretically be used to obtain information about the 100k protein (see Chapter 5), amino acid sequencing of the purified protein would unequivocally identify it as virus- or host-encoded. The work described in this chapter was therefore aimed at developing a scheme for the purification of large amounts of 100k protein from CMV-infected cucumbers. This purified protein could then be used to obtain at least partial amino acid sequence data. Approximately 10  $\mu\text{g}$  (100 pmoles) of 100k protein would be required for amino acid sequencing using current gas-phase sequencer and micro-bore HPLC detection technology (M. Snoswell, personal communication). There are no estimates of the amount of 100k protein or CMV-encoded non-structural proteins present in CMV-infected plants. However, Berna *et al* (1986) used antibodies raised against synthetic peptides to detect AMV non-structural proteins in the early stages of infection in tobacco plants. They estimate that the amount of P1 and P2 proteins per gram of leaves is 18-50 ng and 10.5-30 ng, respectively. If the CMV-induced 100k protein is indeed viral-coded, then it would probably be present at similar levels. Using a conservative estimate of 10 ng/g, it follows that 1 kg of leaves would contain 10  $\mu\text{g}$  of the required protein. If, however, the purification procedure had an overall recovery of 40% (corresponding to 4 purification steps, each with 80% recovery), then 2.5 kg of leaves would be required to obtain 10  $\mu\text{g}$  of purified protein. Hence it would be feasible to obtain the required amount of 100k protein, provided that suitably efficient purification procedures could be developed.

### **Problems with previous approaches.**

Previous schemes for the purification of RdRPase from cucumbers (Kumarasamy and Symons, 1979a; Gill *et al.*, 1981; Gordon *et al.*, 1982; Takanami and Fraenkel-Conrat, 1982; Khan *et al.*, 1986) have achieved high degrees of purification with only very small amounts of protein. The amount of 100k protein present in the most highly-purified fractions was too low to measure accurately (Gordon *et al.*, 1982; Khan *et al.*, 1986) but was visualized using  $^3\text{H}$ -labelling (Kumarasamy and Symons, 1979b) and silver staining (Merril *et al.*, 1981), which can detect less than 10 ng of protein in a single band.

These purification schemes employed conventional liquid chromatography and relied heavily on ion exchange (phosphocellulose and various DEAE matrices) and affinity chromatography using group-specific affinity ligands such as heparin, polynucleotides (poly-C and poly-U) and the triazine dye, Cibacron Blue F3GA. However, the recovery of enzyme from some columns, particularly affinity columns, was low (e.g., less than 30%; Gill *et al.*, 1981; Gordon *et al.*, 1982). In addition, the chromatographic materials employed in all these studies required the use of relatively low pressures and slow flow rates, which would preclude their application to a large scale approach.

### **6. Experimental approach.**

To obtain the maximum possible amount of 100k RdRPase, every stage of the extraction and purification procedure must be optimized. This involves not only developing an efficient large-scale chromatographic purification scheme with high recoveries at each step, but also requires re-evaluation of the methods used for extraction of enzyme from leaves. In addition, the optimal conditions for the stability of the enzyme need to be determined to avoid large losses of activity during storage and purification. Hence, the work described here systematically addressed the following aspects of the purification of CMV RdRPase.

1. The relative yield of RdRPase activity obtained from CMV-infected cucumbers using different extraction procedures.
2. The time-course of RdRPase activity during CMV infection.

3. The stability of enzyme activity during storage and incubation in buffer conditions similar to those used for chromatography.
4. The behaviour and relative recovery of RdRPase during chromatography on different column materials including those suitable for large scale application.
5. The effect of salt gradient conditions and buffer composition on chromatographic behaviour and recovery of RdRPase.

The numerous problems encountered during these studies did not permit the large-scale purification of 100k protein for amino acid sequence analysis. However, the information gained has important implications for the properties of the CMV-induced RdRPase and for the future feasibility of such a project.

## MATERIALS AND METHODS

### MATERIALS

#### **1. Seeds and Soil.**

Cucumber seeds, *Cucumis sativus*, were obtained from Arthur Yates and Co. Initially the 'Supermarket' cultivar was used but more recent virus preparations and all large scale polymerase extractions used cv. Green Gem. Preliminary experiments revealed no significant difference in symptom expression or virus yield between these two cultivars. All plants were grown in 'University of California mix' (U.C. mix) soil.

#### **2. Radiochemicals.**

[ $\alpha$ -<sup>32</sup>P]-UTP at specific activity of 3,000 Ci/mmol was obtained from BRESATEC, Adelaide, South Australia.

#### **3. General Laboratory Chemicals**

All chemicals used were of analytical reagent grade and obtained mostly from Sigma or B.D.H.

Polyethylene glycol (M.W. 6,000-7,500) obtained from Sigma (product number P2139) was used exclusively for PEG precipitations. PEG from other sources was unsatisfactory for precipitation of enzyme extracts and resulted in dark green (rather than clear) supernatants containing enzyme activity.

#### **4. Chromatographic Materials**

Phosphocellulose (Whatman P11) was a generous gift of Dr. Steve Wilton, BRESATEC, and was washed before use with solutions of 0.1N HCl and 0.1N NaOH in 50% ethanol as described by Gill(1983).

Poly (C)-cellulose was prepared by coupling polycytidylic acid (Sigma) to CF-11 cellulose (Whatman) using the method of Carmichael (1979).

Phenyl-Sepharose CL-6B, Blue Sepharose CL-6B, Sephadex G-15 and G-25, Sephacryl S-200, Sephacryl S-300 and all prepacked columns for Fast Protein Liquid Chromatography (FPLC) were obtained from Pharmacia.

BLUE-TRISACRYL M (IBF) was packed into a Pharmacia HR 5/5 column for use on FPLC.

### 5. Buffers for column chromatography

All buffers for column chromatography were prepared using water of high purity (Milli-Q system) and filtered and degassed by passage through a 0.22  $\mu\text{m}$  filter (Millipore type GVWP) under vacuum. Glycerol and acetic acid were AR grade and organic solvents were HPLC grade.

Urea (Merck, Ultra Pure) used in column buffers was deionised in aqueous solution using Amberlite MB-1 mixed bed resin (Sigma) and then buffered with 50 mM Tris-HCl, pH 8.5 to inhibit formation of cyanate ions.

When required, 2-mercaptoethanol (2-ME; Sigma) was added to filtered, degassed buffers immediately before use. Excess buffer remaining after approximately 12 hours was discarded.

Due to the significant change in pH of Tris-HCl buffers with temperature, buffers for use at 4°C were adjusted to pH 7.9 at 25°C so that the pH would be 8.5 at 4°C.

## **METHODS.**

### 1. Virus and Plants

CMV was propagated in cucumber (*Cucumis sativus*) or *Nicotiana* spp. and purified 10-14 days post-inoculation (p.i.) by the method of Peden and Symons (1973).

Dried samples of CMV (Q-strain) were initially inoculated onto *Cucumis sativus* cv. Supermarket. CMV for use as inoculum in RNA polymerase extractions (see below), was propagated in *Nicotiana clevelandii*, *N. glutinosa* or *N. tabacum* L. cv. White Burley, which gave a higher yield of virus than cucumbers. Since many other viruses, including the closely-related cucumovirus TAV, can also multiply in *Nicotiana* spp., care was taken to ensure the purity of CMV preparations. Only virus which had been purified from cucumbers was used as inoculum for *Nicotiana* spp. CMV was not passaged in *Nicotiana* spp., but



maintained in the glasshouse only in *C. sativus*, which is not a host for TAV (Habibi and Francki, 1974c).

For extraction of RNA polymerase, cucumber seedlings were grown in a glasshouse for 9-11 days at which stage the cotyledons were 1.5-2.0 cm long and primary leaves had not yet developed. After a 24-32 hour dark period, the cotyledons were lightly dusted with carborundum powder and inoculated by rubbing both sides with a suspension of purified virus, 0.5-1.0 mg/ml in 50 mM borate, 5 mM EDTA, pH 9. Both the inoculated cotyledons and the small primary leaves were harvested 6-9 days p.i. (unless otherwise stated) and used immediately for RNA polymerase extraction.

Healthy control plants were grown for the same number of days under the same conditions.

Total CMV RNA for use as template in RNA polymerase assays was extracted from purified virus as described by Peden and Symons (1973).

## 2. Extraction of CMV-induced RNA-dependent RNA Polymerase

**Method A : Extraction and PEG precipitation of crude soluble RNA-dependent RNA polymerase.**

A crude extract of RNA-dependent RNA polymerase (RdRPase) was prepared from CMV-infected cucumber seedlings essentially according to the procedure of May *et al* (1969), as modified by Kumarasamy (1980), and precipitated with polyethylene glycol (Kumarasamy and Symons, 1979a). All procedures were performed at 4°C.

Cotyledons and primary leaves were homogenized in a Waring blender for 60-90 seconds in 2 ml Extraction Buffer A (50 mM Tris-HCl, pH 8.5, 100 mM  $\text{NH}_4\text{Cl}$ , 5 mM  $\text{MgSO}_4$ , 5%(v/v) glycerol, 2 mM EDTA, 90 mM 2-ME, 50% saturated  $(\text{NH}_4)_2\text{SO}_4$ ) per gram of leaf tissue. The inclusion of 2 mM EDTA and 5%(v/v) glycerol in this extraction buffer was found by Kumarasamy (1980) to stimulate enzyme activity in the initial extract. The homogenate was squeezed through Miracloth (Calbiochem) or fine nylon cloth and centrifuged at 10,000g for 10

minutes. The pellet was resuspended in Extraction Buffer A (4 ml per gram of leaves) using a Potter-Elvehjem homogenizer and again centrifuged at 10,000g for 10 minutes. The resultant pellet was resuspended in Extraction Buffer A without  $(\text{NH}_4)_2\text{SO}_4$  (2 ml per gram of leaves) and centrifuged as before to yield a pale green ammonium sulphate supernatant ( $\text{AmSO}_4$ -SN) containing RdRPase activity.

Three volumes of this supernatant were mixed with one volume of 44% (v/v) polyethylene glycol, 1 M  $\text{NH}_4\text{Cl}$ , 50 mM 2-ME and stirred gently for 15 minutes. Precipitated enzyme was recovered by centrifugation at 20,000g for 15 minutes and solubilized from the pellet by homogenization in 0.075 ml Resuspension buffer (20 mM Tris-HCl, pH 8.5, 5%(v/v) PEG, 1 mM EDTA, 2 M  $\text{NH}_4\text{Cl}$ , 50 mM 2-ME) per gram of leaves. The clear, slightly pale-green PEG-Enzyme supernatant was obtained following centrifugation of the suspension at 20,000g for 15 minutes.

Small samples of  $\text{AmSO}_4$ -SN were mixed with an equal volume of Storage Buffer (20 mM Tris-HCl, pH 8.5, 1 mM EDTA, 55% (v/v) glycerol) and stored at  $-80^\circ\text{C}$  for assay at a later stage.

#### **Method B : Extraction, solubilization and PEG precipitation of particulate RNA-dependent RNA polymerase.**

The RdRPase activity present in the particulate fraction of CMV- infected cucumber seedlings was extracted and solubilized essentially as described by Gill *et al* (1981), with minor modifications. All procedures were performed at  $4^\circ\text{C}$ .

Cotyledons and primary leaves were homogenized in a Waring blender for 60-90 seconds in 2 ml Extraction Buffer B (50 mM Tris-HCl, pH 8.5, 100 mM  $\text{NH}_4\text{Cl}$ , 90 mM 2-ME, 10%(w/v) sucrose) per gram of leaf tissue. The homogenate was squeezed through fine nylon cloth and centrifuged at 22,000g for 25 minutes. The pellet was washed twice by resuspension in and centrifugation from (22,000 g, 25 min) 4 ml Extraction Buffer B per gram of original leaf material. Pellets were resuspended using a Potter-Elvehjem homogenizer.

The final particulate fraction (PF) pellet was resuspended in Extraction Buffer B without sucrose but containing 0.15 M  $\text{MgSO}_4$  (1 ml per gram of leaves) and the homogenized suspension incubated at about 45°C until the temperature had reached 37°C (this normally took 10-15 mins). It was then incubated at 37°C for 5 minutes and centrifuged at 20,000g (Sorvall GSA rotor; 11,000 rpm) for 30 minutes.

The clear, pale-green supernatant containing the solubilized particulate enzyme was carefully removed from the sloppy dark green pellet and dialyzed for 16-20 hr, with one change, against 1L of Extraction Buffer B. The dialyzed enzyme was then precipitated by the addition of an equal volume of 44% (v/v) polyethylene glycol, 1 M  $\text{NH}_4\text{Cl}$ , 50 mM 2-ME and gentle stirring for 20 minutes. Precipitated enzyme was recovered by centrifugation at 20,000g for 20 minutes and solubilized from the pellet by homogenization in 0.1 ml Resuspension buffer per gram of leaves. The clear, colourless or very slightly pale-green PEG-Enzyme supernatant was obtained following centrifugation of the suspension at 20,000g for 15 minutes.

Small samples of PF and  $\text{MgSO}_4$ -solubilized supernatant were mixed with an equal volume of Storage Buffer and stored at -80°C for assay at a later stage.

#### **Method B: Large-scale Preparations.**

For large scale extractions of RdRPase, the following modifications were made to Method B to allow more convenient manipulation of large amounts of leaves:

(1) The conditions for centrifugation of crude homogenate and washing of PF were changed from 22,000g/25 min to 13,500g/30 min to allow the use of the large capacity GS-3 rotor (Sorvall) which has a maximum speed of 9,000 rpm (13,500g).

(2) The enzyme solubilized from the PF with  $\text{MgSO}_4$  was precipitated immediately with an equal volume of PEG/ $\text{NH}_4\text{Cl}$  solution without prior dialysis.

Neither of these modifications were found to have a significant effect on the

RdRPase activity recovered in either the PF, solubilized or PEG-precipitated steps of the extraction (results not shown).

### 3. Storage of PEG-Enzyme

PEG-Enzyme produced by both of the above methods was snap frozen in a dry ice/ethanol bath and stored at  $-80^{\circ}\text{C}$ . Before freezing, small samples were removed to allow assay of RNA polymerase activity without thawing the entire PEG-Enzyme preparation.

### 4. Assay of RNA-dependent RNA polymerase activity.

RNA-dependent RNA polymerase activity was assayed by measuring incorporation of  $[\alpha\text{-}^{32}\text{P}]\text{-UTP}$  into acid insoluble product essentially as described by Jaspars *et al* (1985), but with several modifications. Since enzyme fractions eluted from columns contained differing amounts of salt and other components, the exact composition of individual assays varied. In the following method the most commonly used assay composition is given together with the range of concentration of each component in parentheses. Variations in assay conditions, where relevant, are given in the appropriate figure legends. It was particularly important to control the  $\text{NH}_4\text{Cl}$  concentration since this can have a marked effect on enzyme activity (Kumarasamy, 1980; Gill, 1983).

Reaction mixtures of 50  $\mu\text{l}$  contained 10  $\mu\text{l}$  (range 2-20) of enzyme fraction and the following final composition: 15 mM (9-20) Tris-HCl, pH 8.5, 5 mM KCl, 13 mM magnesium acetate, 80 mM (20-200)  $\text{NH}_4\text{Cl}$ , 60 mM (57-67) 2-ME, 5  $\mu\text{g}$  CMV RNA, 2% (0-6) v/v glycerol, 0.6 mM each of ATP, CTP and GTP, 0.2 mM UTP (unlabelled) and 2.5  $\mu\text{Ci}$  (2.5-5)  $[\alpha\text{-}^{32}\text{P}]\text{-UTP}$  (3,000 Ci/mmol).

Phospho(enol)pyruvate and pyruvate kinase used in assays by Jaspars *et al* (1985) were omitted here without significant effect on RdRPase activity.

Reactions were incubated at  $37^{\circ}\text{C}$  for 60 minutes and 45  $\mu\text{l}$  samples from each reaction were spotted onto GF/A filter discs or squares of Whatman 3MM paper. Discs or squares were washed batchwise 4 times for 10-15 minutes in 1N HCl, 2% (w/v)  $\text{Na}_2\text{HPO}_4$ , 2% (w/v)  $\text{Na}_4\text{P}_2\text{O}_7$ , rinsed briefly in ethanol, dried in an

oven (120°C) for 5 minutes and then counted in Triton-Toluene scintillation fluid (0.22% (w/v) PPO, 0.022% (w/v) POPOP, 37.5% (v/v) Triton-X100 in toluene) using a Beckman LS7500 liquid scintillation spectrometer. Counts incorporated in blank assays (i.e. containing no enzyme) were subtracted from all experimental values.

Enzyme activity was calculated using the percentage incorporation of  $^{32}\text{P}$ -UMP into acid-insoluble product and the known starting concentration of UTP in the reaction. One Unit of RdRPase activity is defined as catalyzing the incorporation of 1 nmole of UTP per minute into acid-insoluble material.

### 5. Chromatography trials of Method A PEG-Enzyme.

Samples of Method A PEG-Enzyme were fractionated by phosphocellulose chromatography and the active fractions then used for trials of other columns. All columns were run at 4°C.

Phosphocellulose chromatography was performed as described by Gill *et al* (1981) except that PEG-Enzyme was diluted 1/10 in TGEM buffer (20 mM Tris-HCl, pH 8.5, 30% glycerol, 1 mM EDTA, 30 mM 2-ME) to give a final  $\text{NH}_4\text{Cl}$  concentration of 0.2 M before loading, as suggested by Gill (1983). The column bed volume was 2-10 ml and the flow rate was approximately 10 cm/h. RdRPase was eluted in TGEM buffer containing 0.6 M  $\text{NH}_4\text{Cl}$  and could be stored in this buffer at -20°C for at least 5 days without significant loss of activity.

Poly (C)-cellulose chromatography was performed exactly as described by Gill *et al* (1981), using a 3 x 0.8 cm column running at 4 ml/hour.

A 4 x 0.8 cm column of Blue Sepharose CL-6B was equilibrated and run as described by Gill *et al* (1981) for Cibacron Blue F3GA agarose, except that the pH of the buffer was 8.5 rather than 7.3. The lower pH was not necessary to affect binding of RdRPase to the column. Before loading, samples of phosphocellulose-fractionated enzyme (see above) were diluted in TGEM buffer to give a final  $\text{NH}_4\text{Cl}$  concentration of 0.1 M, and magnesium acetate was added to 10 mM.

Columns of Sephacryl S-200 (25 x 1.6 cm) or S-300 (3.2 x 38 cm) were run in 20 mM Tris-HCl, pH 8.5, 5% glycerol, 1 mM EDTA, 30 mM 2-ME, and 0.2 M

$\text{NH}_4\text{Cl}$ , at flow rates recommended by the manufacturer. Samples of phosphocellulose-fractionated enzyme were precipitated with 2 volumes of saturated ammonium sulphate, in the presence of 200  $\mu\text{g}/\text{ml}$  BSA as carrier, and resuspended in column buffer (above). Alternatively, samples were concentrated using Centriflo CF50A membrane cones (Amicon), which also allowed exchange of buffer to the correct column buffer. Final sample volumes were about 5% of the column bed volume.

#### **6. Preparation of Method B Peg-Enzyme for column chromatography.**

PEG-Enzyme prepared by Method B (see above) was centrifuged at 20,000g/4°C for 15 minutes to remove any residual particulate material. The supernatant was then passed through a sterile nitrocellulose filter (Sartorius) with pore size of 0.45  $\mu\text{m}$  or 0.22  $\mu\text{m}$ . There was no loss of enzyme activity due to either centrifugation or filtering (results not shown). PEG-Enzyme prepared in this way was stored in aliquots at -80°C and used for chromatography immediately upon thawing without further filtering or dialysis.

#### **7. Hydrophobic interaction chromatography of PEG-Enzyme.**

Hydrophobic interaction chromatography was performed using 2-10 ml columns of Phenyl-Sepharose CL-6B running at 25-50 cm/h or a pre-packed Phenyl-Superose HR 5/5 column running on the FPLC system at 0.25-0.5 ml/min. All columns were run at 4°C. Various buffer conditions were investigated (see Results). Most columns were run in PGM buffer (0.1M  $\text{KH}_2\text{PO}_4/\text{Na}_2\text{PO}_4$ , pH 7.0, 10% glycerol, 5 mM 2-ME). This buffer was prepared by mixing solutions of  $\text{KH}_2\text{PO}_4$  and  $\text{Na}_2\text{HPO}_4$  in the proportions required to obtain pH 7.0.

Samples of filtered PEG-Enzyme (0.25-0.75 ml) were diluted with an equal volume of 2 x PGM buffer containing 2-3 M  $\text{NH}_4\text{Cl}$  and loaded onto a column equilibrated in PGM with the same  $\text{NH}_4\text{Cl}$  concentration. Enzyme was eluted by decreasing the  $\text{NH}_4\text{Cl}$  concentration, finishing with PGM alone. Columns were cleaned after each use by washing successively with water, 0.1M NaOH and

ethanol. Phenyl-Superose columns were additionally washed with 0.1% TFA in acetonitrile.

### **8. Chromatography of PEG-Enzyme on FPLC ion exchange media.**

Ion exchange chromatography was performed with the FPLC system using prepacked MONO Q HR 5/5 and MONO S HR 5/5 columns (bed volume 1 ml).

Prior to use, columns were equilibrated in TGM containing  $\text{NH}_4\text{Cl}$  at the required starting concentration, usually 0.1 M. The flow rate was 0.5-1.0 ml/min.

Samples of filtered PEG-Enzyme (0.1-0.5 ml) were diluted with TGM to give the required starting  $\text{NH}_4\text{Cl}$  concentration, usually 0.1 M. Samples were pumped onto the column using a "Superloop" loading chamber. The exact salt conditions used for elution varied between experiments and are detailed in the appropriate figure legends. For experiments at room temperature, ice was placed in the fraction collector bowl, so that fractions were collected into pre-chilled tubes.

Ion exchange columns were cleaned after each use by several successive 2 ml injections of cleaning solution (0.1 M NaOH, 1 M NaCl, 20% methanol) each followed by 5 ml of water, 5 ml of 2 M  $\text{NH}_4\text{Cl}$  and a further 5 ml of water.

### **9. Chromatography of PEG-Enzyme on BLUE-TRISACRYL M.**

A 1 ml (HR 5/5) column of BLUE-TRISACRYL M was equilibrated with TGM buffer containing 10 mM magnesium acetate and 0.2 M  $\text{NH}_4\text{Cl}$ . Samples of filtered PEG-Enzyme (0.15-0.3 ml) were diluted 1/10 with TGM buffer and magnesium acetate was added to 10 mM. After loading, the column was washed with the same buffer used for equilibration. Enzyme was eluted with the same buffer without magnesium acetate. The column was washed extensively after each use with 50 mM Tris-HCl, pH 8.5, 6M urea and then cleaned by several successive 2 ml injections of 0.2 M NaOH, each flushed with water and  $\text{NH}_4\text{Cl}$ , as described for ion exchange columns (see above).

### 10. Measurement of protein concentration.

The concentration of protein in enzyme preparations and column fractions was determined by the method of Bradford (1976) modified for use in 96-well microtitre plates (Rylatt and Parish, 1982). In contrast to other methods for determining protein concentration, this method shows minimal interference by 2-ME, glycerol or polyethylene glycol at concentrations used in extraction and purification procedures. Samples were clarified by centrifugation in a microcentrifuge for 2 minutes before assaying.



## RESULTS

### 1. Method for extraction of RdRPase from CMV-infected cucumber seedlings.

Several investigations were undertaken to determine the simplest and most efficient method for extraction of RdRPase activity from CMV-infected tissue. Initial extractions used Method A and both the  $\text{AmSO}_4$ -SN and PEG-Enzyme preparations produced by this method were investigated for their suitability as starting materials for chromatographic purification of RdRPase. Several problems were encountered with this procedure (see below) and therefore a systematic comparison with Method B was undertaken.

#### **Extraction using Method A and attempted column purification of Method A $\text{AmSO}_4$ -SN.**

Method A, the ammonium sulphate extraction procedure of May *et al* (1969), was initially chosen to extract RdRPase activity since it is a simple and rapid procedure which has been used previously to obtain crude polymerase preparations from virus-infected cucumbers (May *et al*, 1969; Clark *et al*, 1974; Takanami and Fraenkel-Conrat, 1982; Khan *et al*, 1986). The inclusion of 50% ammonium sulphate ( $\text{AmSO}_4$ ) in the extraction buffer facilitates the removal of contaminating plant proteins from the RdRPase which is insoluble in 50%  $\text{AmSO}_4$ . In particular this procedure has been reported to remove uridyl terminal transferase (Takanami and Fraenkel-Conrat, 1982) and essentially all plant ribonucleases (May *et al*, 1969).

Several attempts were made to load the green-coloured  $\text{AmSO}_4$ -SN, obtained by Method A, directly onto a column for further purification. It was hoped that a chromatographic procedure could be used to remove the green-coloured material and simultaneously affect some degree of purification. Initial attempts to separate green material from the enzyme by size exclusion chromatography on

Sephadex G-15 or G-25 resulted in them both eluting together in the void volume (results not shown), suggesting that green-coloured contaminants were associated with the high molecular weight enzyme. Increasing the ionic strength of the column eluant to 1 M with  $\text{NH}_4\text{Cl}$  did not prevent this association.

Phosphocellulose chromatography of  $\text{AmSO}_4$ -SN also failed to remove contaminating green material (results not shown). Hence it was not feasible to load  $\text{AmSO}_4$ -SN directly onto columns without prior treatment.

#### **Attempted column purification of Method A PEG-Enzyme.**

Precipitation of  $\text{AmSO}_4$ -SN with 11% PEG (see Method A) and solubilization of the RdRPase using a high salt resuspension buffer removed most of the green material leaving a PEG-Enzyme preparation which was clear and pale-green. This enzyme usually became turbid when frozen and re-thawed, or dialyzed according to the original method of Kumarasamy and Symons (1979a). The removal of this cloudy precipitate by centrifugation or filtration through a nitrocellulose filter resulted in some loss of RdRPase activity (results not shown). Nevertheless, Method A PEG-Enzyme was applied to several trial columns without clarification. The samples were diluted prior to loading, rather than dialyzed, since Gill (1983) reported that this resulted in higher recovery of RdRPase activity following chromatography.

Table 4.1 summarizes the results of chromatography trials with the Method A PEG-Enzyme. Phosphocellulose chromatography gave recoveries similar to those reported previously (Kumarasamy and Symons, 1979a; Gill *et al.*, 1981). Any remaining green material did not co-migrate with enzyme activity. However, attempts at further purification were not successful. Blue Sepharose CL-6B gave erratic results and no activity was recovered during loading or elution of poly (C)-cellulose columns. RdRPase was eluted, with reduced activity, in the void volume of Sephacryl S-200 and S-300 columns, suggesting that it was aggregating to have a high apparent molecular weight (greater than  $M_r 1.5 \times 10^6$ ).

**TABLE 4.1: Chromatography of Method A PEG-Enzyme preparations.**

PEG-Enzyme prepared by Method A was diluted 1/10 in TGEM buffer (see Methods) and applied to a column of phosphocellulose equilibrated in TGEM containing 0.2 M  $\text{NH}_4\text{Cl}$ . The active fractions, eluted with the same buffer containing 0.6 M  $\text{NH}_4\text{Cl}$ , were loaded onto other columns as shown, either immediately or after storage of the phosphocellulose fractions at  $-20^\circ\text{C}$  for 1-5 days. Further details are given in the text and in Materials and Methods. Recovery of activity is expressed as percentage of that loaded. The degree of purification refers to the increase in specific activity over the loaded sample. Results are the average of at least three separate experiments except those for poly(C)-cellulose which was tried only twice.

COLUMN	SAMPLE	RECOVERY	PURIFICATION
Phosphocellulose	PEG-Enzyme	50-75%	15-20 X
poly (C) - cellulose	phosphocellulose fractions	----no activity recovered----	
Blue-Sepharose CL-6B	phosphocellulose fractions	50-60% <sup>s</sup>	N.D.
Sephacryl S-200 and S-300	phosphocellulose fractions	<50% ----in void volume----	

<sup>s</sup>: Recovery of activity from Blue-Sepharose columns was highly variable and was 50-60% at best. Recoveries of less than 20% were occasionally obtained.  
 ND: not determined.

**Requirements for an alternative extraction method.**

Although these chromatography trials were not successful, they highlighted some important requirements for a suitable purification method. To enable rapid processing of large amounts of sample, higher flow rates than those possible with gels such as phosphocellulose would be needed. Ideally this would involve the use of fast-flow Sepharose and FPLC media. To use these columns required PEG-Enzyme preparations which were essentially clear and colourless and could be filtered prior to loading. The alternative extraction Method B for preparation of PEG-Enzyme was therefore investigated.

Method B is essentially that of Gill *et al* (1981) who established the optimal conditions for the solubilization of RdRPase from the particulate fraction (PF) in the presence of magnesium sulphate. Solubilization requires the presence of both magnesium and sulphate ions and occurs almost instantaneously once the temperature of the PF suspension has reached 37°C. The major advantage of this technique is that it releases only 5-7% of the protein in the PF, thus affording a considerable purification of RdRPase at an early stage. It has been used as the initial step in the purification of particulate CMV-induced RdRPase (Gill *et al*, 1981; Gordon *et al*, 1982).

**Time-course of RdRPase-induction in CMV-infected plants.**

To compare the two extraction methods and determine the time-course of enzyme induction during viral infection, RdRPase was extracted from CMV-infected cucumbers at varying stages of infection using both Methods A and B (Figure 4.1). The yield of RdRPase activity per gram of leaf tissue reached a peak at 6-9 days post-inoculation (p.i.) and declined sharply by 12 days. The same pattern was observed using both methods, but the yield of PEG-Enzyme activity was significantly higher (2.5 to 3.5-fold at 6-9 days p.i.) with Method B (compare Figure 4.1, A and B). Furthermore, the specific activity of PEG-Enzyme obtained by Method B was, on average, five times higher than that obtained with Method A (Figure 4.1C).

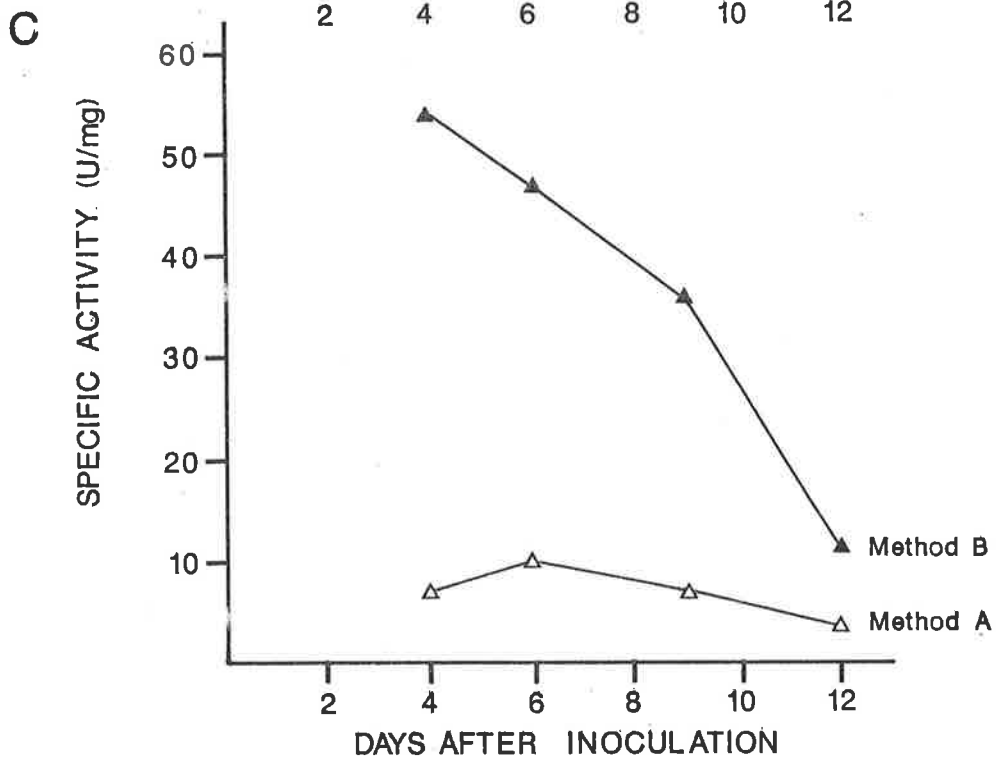
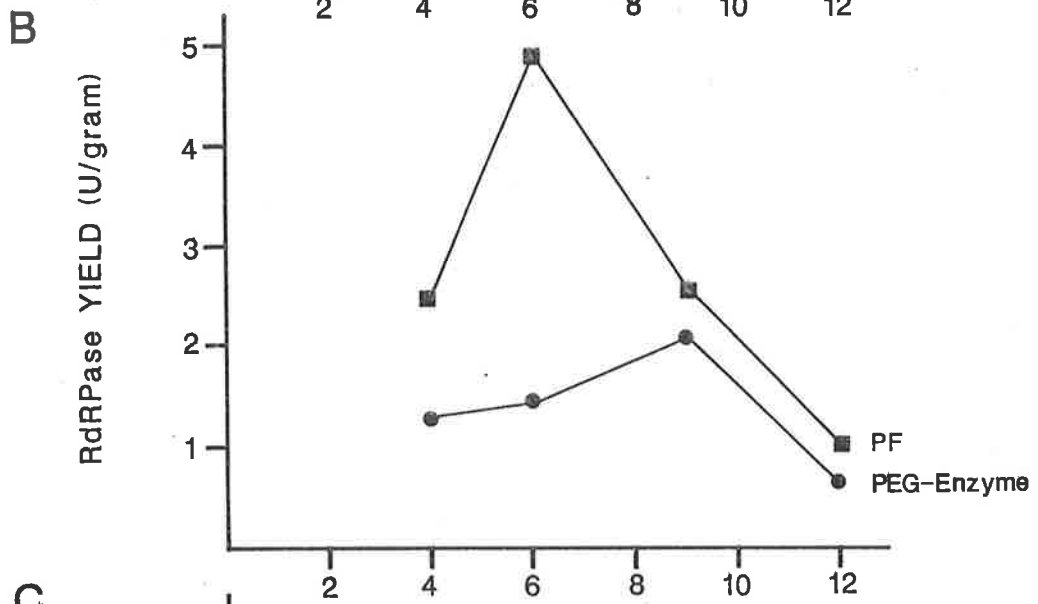
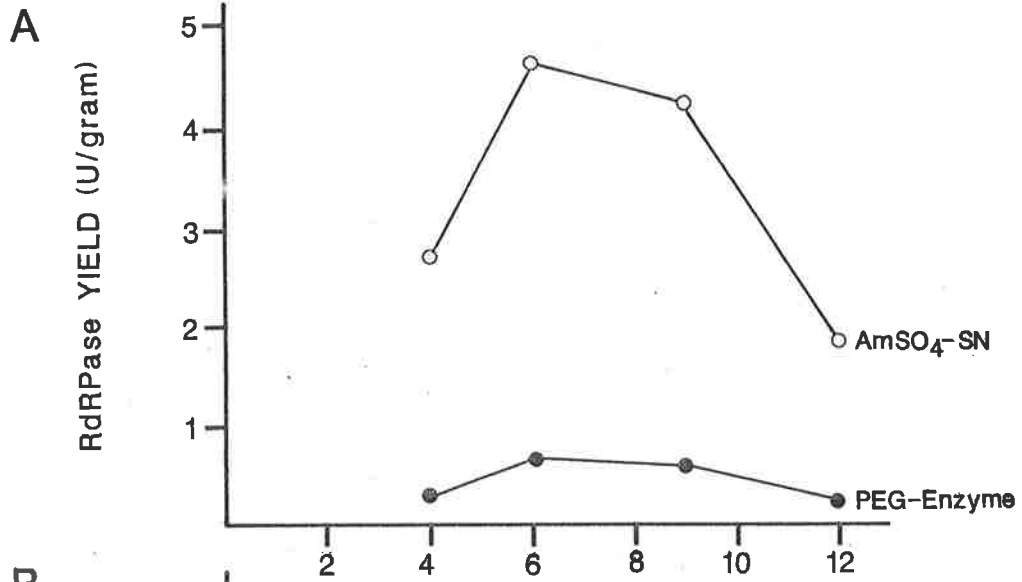
Since the enzyme yield in the crude preparations (  $\text{AmSO}_4$ -SN or PF, for

**FIGURE 4.1: Time-course of RdRPase activity during CMV infection.**

RdRPase was extracted from CMV-infected cucumbers by Method A (A) and Method B (B) at various stages of infection. The yield of enzyme activity (Units per gram of leaf tissue) was determined at the  $\text{AmSO}_4$ -SN (○), PF (■) and PEG-Enzyme (●) steps.

The specific activity (C: Units per mg protein) of PEG-Enzyme preparations was also determined for Method A (△) and Method B (▲).

A large batch of ten-day old cucumber seedlings were inoculated with purified CMV after a 24-hour dark period. Samples (approx. 100g) of infected cotyledons, and primary leaves if present, were harvested 4, 6, 9 and 12 days after inoculation and RdRPase immediately extracted using both Method A (approx. 50g) and Method B (approx. 50g). Samples of PF,  $\text{AmSO}_4$ -SN (both diluted with an equal volume of storage buffer) and PEG-Enzyme were snap-frozen and stored at  $-80^\circ\text{C}$ . Separate samples were stored at  $-20^\circ\text{C}$  for assay of protein concentration. Duplicate RdRPase assays were performed simultaneously on all stored samples using 2  $\mu\text{l}$  of PEG-Enzyme or 10  $\mu\text{l}$  of other fractions in 50  $\mu\text{l}$  assays. To enable comparisons, the  $\text{NH}_4\text{Cl}$  concentration was adjusted to 80 mM in all assays.



Methods A and B respectively) was very similar, the major problem with Method A appeared to be the recovery by PEG precipitation. Increasing the final concentration of PEG in the precipitation step from 11% to 22% (w/v), as recommended by Khan *et al* (1986) for the precipitation of CMV RdRPase, did not increase the recovery in Method A PEG-Enzyme (results not shown).

### **Comparison of Methods A and B.**

In addition to the data presented in Figure 4.1, several other RdRPase preparations were obtained using both extraction methods, confirming the time-course results and allowing an accurate comparison between Methods A and B. Based on all available data, summarized in Table 4.2, it is clear that Method B is far superior for extraction of CMV-induced RdRPase. Not only is the yield of enzyme higher per gram of starting material, but the resultant PEG-Enzyme has a considerably higher specific activity than that prepared by Method A. The clear light green-coloured solubilized fraction facilitates preparation of clear colourless PEG-Enzyme which can be conveniently filtered ready for column chromatography without loss of activity.

All subsequent investigations therefore used Method B PEG-Enzyme as their starting material.

### **2. Large-scale preparation of Method B PEG-Enzyme.**

To obtain large amounts of RdRPase for subsequent purification, PEG-Enzyme was prepared from 13 batches (each 95-250g) of CMV-infected cucumbers, 6-9 days post-inoculation, using Method B with modifications made for large-scale extraction (see Methods). A total of approximately 2.2 kg of leaves were extracted, yielding 210 ml of PEG-Enzyme. Each preparation was assayed for RdRPase activity and protein concentration. Analogous preparations were obtained from approximately 2 kg of healthy cucumbers and their protein concentrations determined.

These data were averaged and used to obtain information about the extraction method, as presented in Table 4.3. Solubilization of RdRPase from the PF afforded a 5-fold purification and PEG precipitation a further 2-fold

**TABLE 4.2: Comparison of Method A and Method B  
PEG-Enzyme preparations.**

Summary of the properties of PEG-Enzyme prepared 6-9 days p.i. by Methods A and B. Figures are the average of 6 separate preparations including those from Day 6 and Day 9 in Figure 4.1.

	<b>Method A</b>	<b>Method B</b>
<b>RdRPase Yield</b> (Units/gram)	0.32	1.58
<b>Recovery<sup>a</sup></b> from crude fraction (%)	11.6	64.6
<b>Specific Activity</b> (U/mg)	4.05	27.6
<b>Protein concentration</b> (mg/ml)	0.92	0.61
<b>Clarity</b>	CLOUDY <sup>b</sup>	CLEAR
<b>Colour</b>	PALE-GREEN	COLOURLESS <sup>c</sup>
<b>Activity remaining<sup>d</sup></b> after filtering (%)	≈50	≈100

- Refers to percentage recovery of RdRPase activity from AmSO<sub>4</sub>-SN PEG-Enzyme in case of Method A and PF to PEG-Enzyme in Method B.
- Method A PEG-Enzyme was often cloudy in appearance especially if it had been frozen and re-thawed. Dialysis exacerbated the formation of this cloudy precipitate which, when removed by centrifugation or filtration, decreased RdRPase activity.
- Most Method B PEG-Enzyme preparations were colourless, although some were very pale green.
- Refers to percentage of activity remaining after passage of PEG-Enzyme through a nitrocellulose filter with pore size of 0.45 μm. Filtering is a necessary step in the preparation of samples for chromatography, especially in the case of FPLC.



**TABLE 4.3:** Large-scale extraction of RdRPase activity by Method B.

The yield of RdRPase activity and the protein concentration were determined for each step of 13 large-scale Method B RdRPase extractions. These figures were averaged and used to calculate the mean Specific Activity (Units per mg protein), recovery and purification factor (relative to particulate fraction) at each step. Samples of PF, MgSO<sub>4</sub>-soluble supernatant (both diluted with an equal volume of Storage Buffer) and PEG-Enzyme from each preparation were stored at -80°C until they could be assayed at the same time. The NH<sub>4</sub>Cl concentration in all assays was adjusted to 80 mM. Preparations from healthy plants were not assayed for RdRPase activity, but the protein concentration at each step was determined (average of 9 separate large-scale preparations).

Step	RdRPase yield (Units per gram of leaves)	Recovery (%)	Protein concent'n (mg/ml)	Specific activity (U/mg)	Purification factor	HEALTHY Protein concent'n (mg/ml)
Particulate fraction	2.5	100	0.41	5.9	1	0.25
MgSO <sub>4</sub> -soluble supernatant	7.3	296	0.24	30.3	5.1	0.12
PEG-Enzyme	2.8	114	0.49	59.0	10	0.38

purification, resulting in PEG-Enzyme with an average specific activity of 59 U/mg (Table 4.3). The large increase in RdRPase activity in  $MgSO_4$ -solubilized supernatant is probably due to the removal of inhibitors present in crude PF. Dialysis of this solubilized supernatant increased its activity even further, but PEG precipitation of dialyzed samples gave PEG-Enzyme preparations with similar activity to those prepared from undialyzed samples (results not shown). PEG-Enzyme was therefore routinely prepared from undialyzed solubilized supernatant.

Large-scale preparations from healthy plants were not assayed for RdRPase activity, but were found to have lower protein concentrations at all steps of the procedure (Table 4.3). No RdRPase activity was detected in any healthy preparations previously obtained (results not shown).

### **3. Stability of PEG-Enzyme.**

#### **Stability under storage.**

Initial preparations of (Method A or B) PEG-Enzyme, stored at  $-15^{\circ}C$ , were found to lose enzyme activity when stored for more than 2-3 months and, therefore storage at  $-80^{\circ}C$  was adopted routinely. Furthermore, repeated freezing and thawing of PEG-Enzyme was avoided by aliquoting samples for assay and chromatography into separate tubes. Under these conditions, PEG-Enzyme could be stored for at least 6 months with no appreciable loss of activity (results not shown).

Since purification procedures were to be carried out at  $4^{\circ}C$  or room temperature, and with diluted PEG-Enzyme, stability of enzyme activity was investigated under these conditions.

#### **Stability of undiluted PEG-Enzyme.**

Incubation of undiluted PEG-Enzyme at  $4^{\circ}C$  and at  $25^{\circ}C$  (room temperature) resulted in approximately 20% loss of RdRPase activity over a period of 8 hours (Figure 4.2A). However, there was little difference in stability between these two temperatures. RdRPase was much less stable at higher temperatures. For example, less than 10% activity remained after incubation at  $30^{\circ}C$  for 8 hours

(results not shown).

### **Stability of PEG-Enzyme diluted for chromatography**

PEG-Enzyme diluted 1/20 in TGM buffer containing 30 mM 2-ME showed essentially no loss of activity over 8 hours at 4°C whereas equivalent samples in buffer without 2-ME showed a slight decrease in activity (less than 15%) over the same period (Figure 4.2B). Almost identical results were obtained when this experiment was repeated at 25°C or without glycerol (results not shown). Since buffer containing 30 mM 2-ME absorbs significantly at 280 nm, 5 mM 2-ME was used in all chromatography buffers to provide adequate reducing conditions but still enable an absorbance profile to be obtained. Assay of free thiol groups in samples of TGM buffer (containing 5 mM 2-ME) at 25°C and 4°C using the method of Habeeb (1972) showed that 2-ME was not significantly oxidized over the period of time used for these experiments.

Hence, RdRPase in PEG-Enzyme preparations shows acceptable stability at 4°C or at room temperature when diluted in TGM buffer over the period of time typically required for chromatography.

### **4. Hydrophobic interaction chromatography of PEG-Enzyme.**

Hydrophobic interaction chromatography (HIC) was tested as a possible first step in the purification of RdRPase, since high salt conditions such as those present in PEG-Enzyme (2M  $\text{NH}_4\text{Cl}$ ), favour hydrophobic interactions and facilitate binding to HIC columns. This would allow PEG-Enzyme to be loaded directly without dialysis or dilution. Reducing hydrophobic interactions by decreasing the salt concentration or by changing the buffer composition, could then be used to elute bound proteins. This technique has been used successfully for large-scale purifications (see Pharmacia Separation News, vol. 13, number 3, 1986) and provides a separation mode which is complementary to ion exchange. Phenyl-Sepharose CL-6B and Phenyl-Superose columns (FPLC) were therefore tested using various buffers and salt conditions. Both columns gave similar results.

**FIGURE 4.2: Stability of PEG-Enzyme under various incubation conditions.**

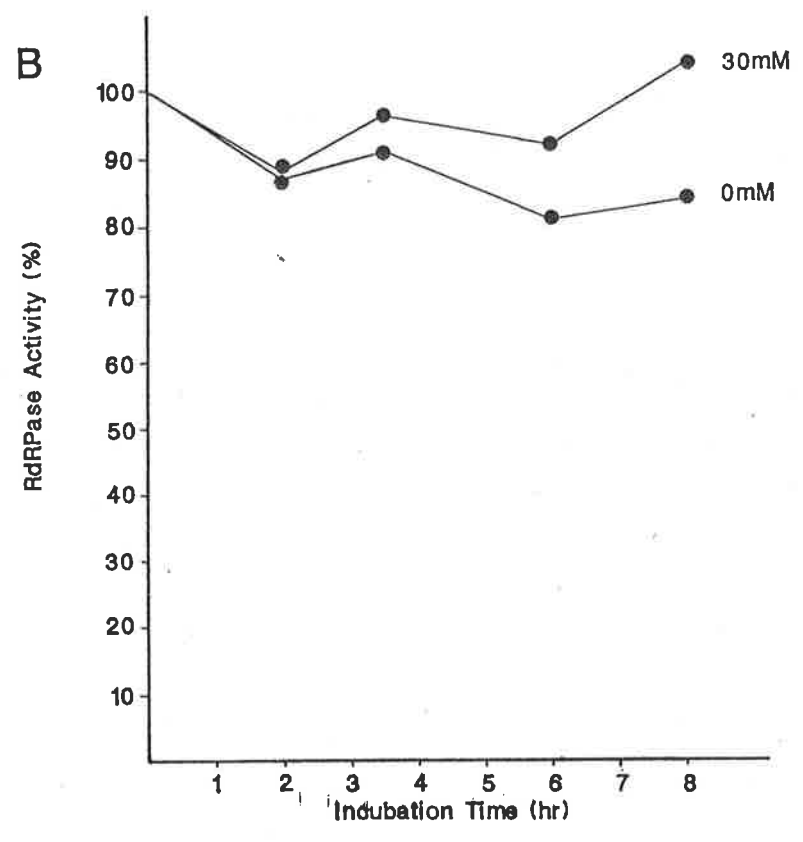
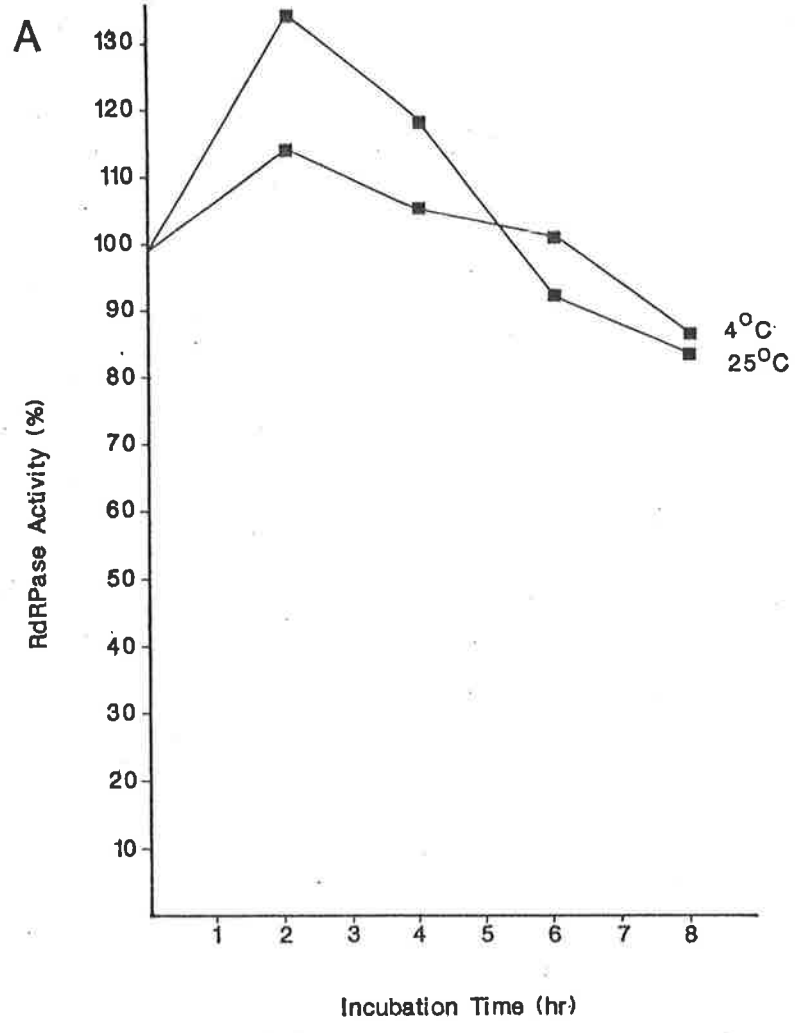
**A:** Stability of undiluted PEG-Enzyme incubated at 25°C or at 4°C.

**B:** Stability of PEG-Enzyme incubated at 4°C following dilution 1/20 in TGM buffer containing 30 mM 2-ME or without 2-ME.

Duplicate 10 µl aliquots of PEG-Enzyme were removed from storage (-80°C) at two-hourly intervals and either incubated immediately (**A**) or diluted with TGM buffer (**B**) before incubating at the specified temperature. After 8 hours, the final aliquots were thawed and all samples assayed immediately for RdRPase activity. Duplicate 50 µl assays, each containing 2 µl of PEG-Enzyme (**A**) or 10 µl of diluted PEG-Enzyme (**B**) were performed for each aliquot. Each point in the figure therefore corresponds to the average of 4 assay results, expressed as a percentage of enzyme activity in the "0 hours" sample. Average acid- precipitable radioactivity (cpm) in the "0 hours" samples were as follows:

A; 4°C sample- 14,131 ; 25°C sample- 16,662.

B; 0 mM 2-ME sample-17,143 ; 30 mM 2-ME sample-15,612.



Initially, columns were run in 20 mM Tris, pH 8.5, 30% glycerol, 1 mM EDTA, 5 mM 2-ME, containing various  $\text{NH}_4\text{Cl}$  concentrations. Under these conditions, however, RdRPase did not bind to the column, even when the initial  $\text{NH}_4\text{Cl}$  concentration was 2.75 M. Decreasing the glycerol concentration to 10%, or lowering the pH to 7.8 did not prevent some enzyme from flowing through the column. Complete binding was only achieved when the sample and column both contained PGM buffer (0.1M  $\text{KH}_2\text{PO}_4/\text{Na}_2\text{HPO}_4$ , pH 7.0, 10% glycerol, 5 mM 2-ME) with at least 2M  $\text{NH}_4\text{Cl}$ . This buffer was expected to assist binding to HIC columns since the presence of  $\text{PO}_4^{3-}$ ,  $\text{Na}^+$  and  $\text{K}^+$  ions, the lower concentration of glycerol, and the lower pH all increase hydrophobic interactions.

RdRPase could be eluted by lowering the  $\text{NH}_4\text{Cl}$  concentration using either gradient or step elution. However, results were not consistent, since the enzyme eluted as 2 or 3 poorly-defined peaks at variable  $\text{NH}_4\text{Cl}$  concentrations even when identical aliquots of the same PEG-Enzyme preparation were loaded. Overall recovery was less than 40%. No RdRPase activity was recovered when PEG-Enzyme was loaded directly onto the column, regardless of the eluant buffer used. Incubation of PEG-Enzyme in PGM buffer showed no significant loss of RdRPase activity over 8 hours at 4°C (results not shown).

Hence, even though RdRPase could be effectively bound to a phenyl-containing column, unfavourable interactions with the column appeared to result in loss of activity and erratic elution profiles. This unfortunately precluded the use of HIC for RdRPase purification.

### **5. FPLC ion-exchange chromatography of PEG-Enzyme.**

Ion exchange chromatography of PEG-Enzyme was investigated using MONO Q (anion exchange) and MONO S (cation exchange) columns on the FPLC system. These columns allow rapid and reproducible separations of proteins and have very high capacities (25 mg protein per 1 ml column) despite their small size. In addition, purification protocols developed on a small scale using these

columns can be readily scaled up using large columns of Q- and S-Sepharose Fast Flow media which contain the same ionic groups as MONO Q and MONO S, respectively, and have extremely good flow properties (flow rates up to 400 cm/h; eg., greater than 100 ml/min for a 5 cm diameter column).

Initial attempts to use buffer containing phosphate pH 7.0 (PGM buffer) for the MONO S column (as recommended for a cation exchange column) gave poor recovery of RdRPase activity. TGM buffer was therefore used for both columns. Under these conditions RdRPase showed very similar behaviour on MONO Q and MONO S columns (Figure 4.3). Repeated investigations using these columns showed that recovery of RdRPase activity was primarily dependent on the slope of the salt gradient used for elution. Recoveries greater than 75% were generally obtained from both MONO Q and MONO S columns using steep gradients such as those shown in Figure 4.3, whereas flatter gradients gave much lower recoveries (see below). Good recoveries were also obtained using stepwise elution, provided that the step in  $\text{NH}_4\text{Cl}$  concentration was sufficiently large, such as that described below (Figure 4.4). The MONO Q column was chosen as a test system for further characterization of factors affecting chromatographic behaviour of RdRPase.

#### **Effect of temperature and buffer composition.**

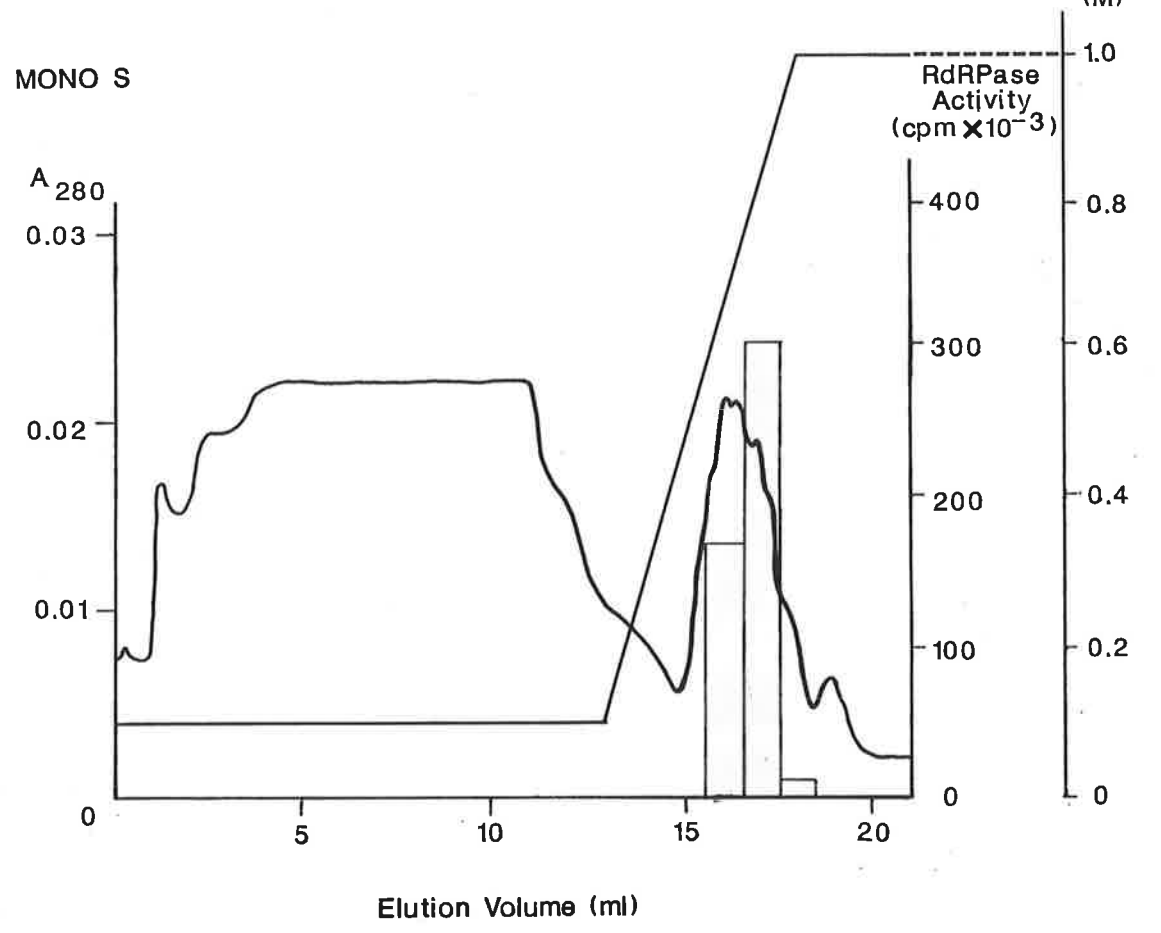
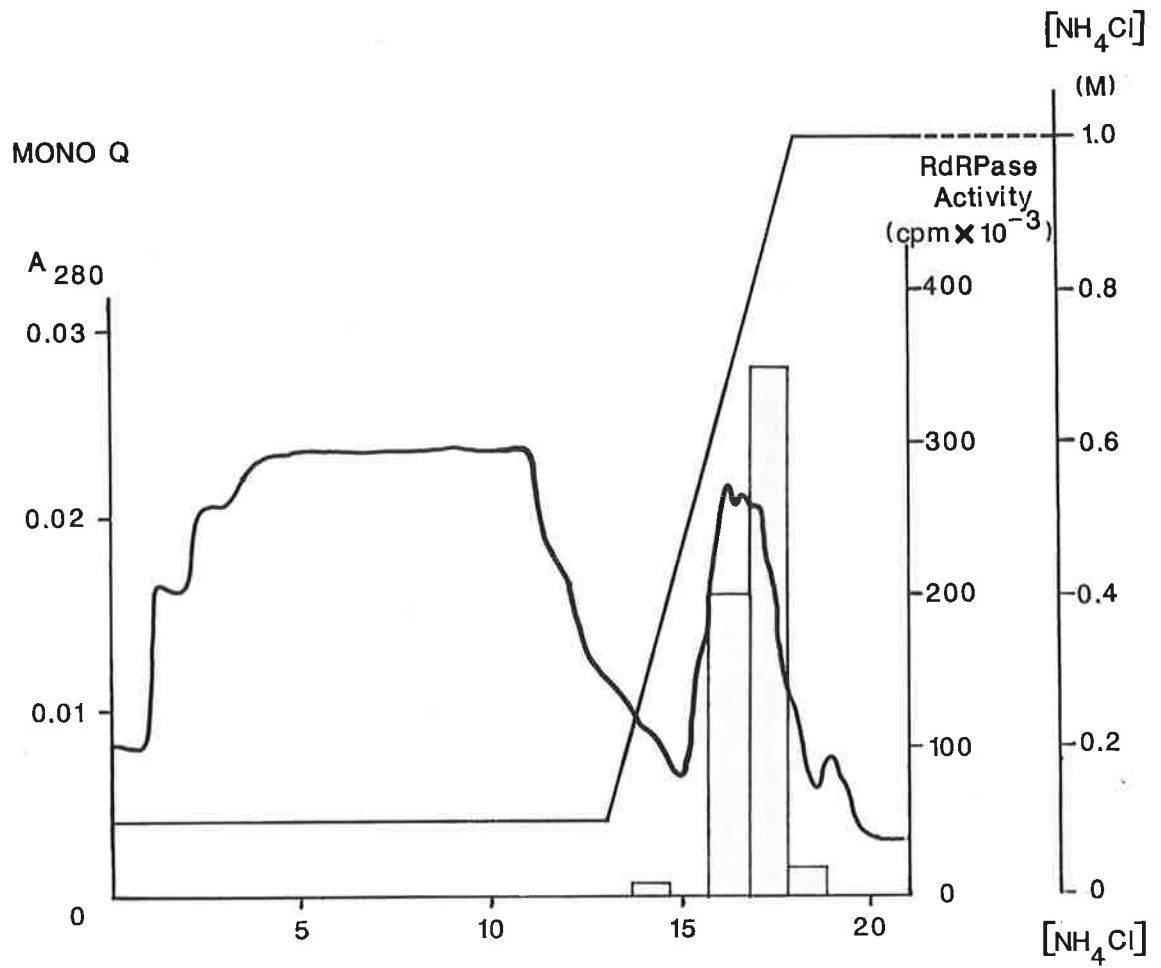
To study the effect of temperature and variations in buffer composition on the recovery of enzyme activity, a stepwise elution scheme was used. At room temperature, RdRPase could be eluted from a MONO Q column by a step in  $\text{NH}_4\text{Cl}$  concentration from 0.1 to 0.4 M (Figure 4.4A) with 50-80% recovery of activity (Table 4.4). When the same experiment was repeated at 4°C, RdRPase activity did not elute until the  $\text{NH}_4\text{Cl}$  concentration was increased from 0.4 to 1.0 M (Figure 4.4B). Recovery was slightly higher at 4°C (Table 4.4). This difference was not due to a change in the pH of the eluant buffer, since buffers were prepared so that their pH would be 8.5 at the appropriate temperature.

Further investigations at 4°C were aimed at determining the effect of different buffer components on enzyme activity. Omission of 2-ME from the

**FIGURE 4.3: Chromatography of PEG-Enzyme on MONO Q and MONO S columns.**

Identical 0.5 ml samples of PEG-Enzyme (0.76 mg/ml) were diluted with TGM buffer to 10 ml and loaded onto MONO Q (upper panel) and MONO S (lower panel) columns equilibrated at room temperature (approx. 25°C) with TGM buffer containing 0.1 M  $\text{NH}_4\text{Cl}$ . RdRPase was eluted using a linear gradient of  $\text{NH}_4\text{Cl}$  in TGM buffer from 0.1 M to 1.0 M over 5 ml (gradient rate=190 mM/ml) as shown. The flow rate was 1 ml/min. Fractions containing RdRPase (shaded boxes) were detected by assaying 10  $\mu\text{l}$  aliquots from each 1 ml fraction. Recovery of enzyme activity, relative to the PEG-Enzyme sample loaded, was 78% and 52% for the MONO Q and MONO S columns, respectively.





**FIGURE 4.4: Stepwise elution of RdRPase from MONO Q columns.**

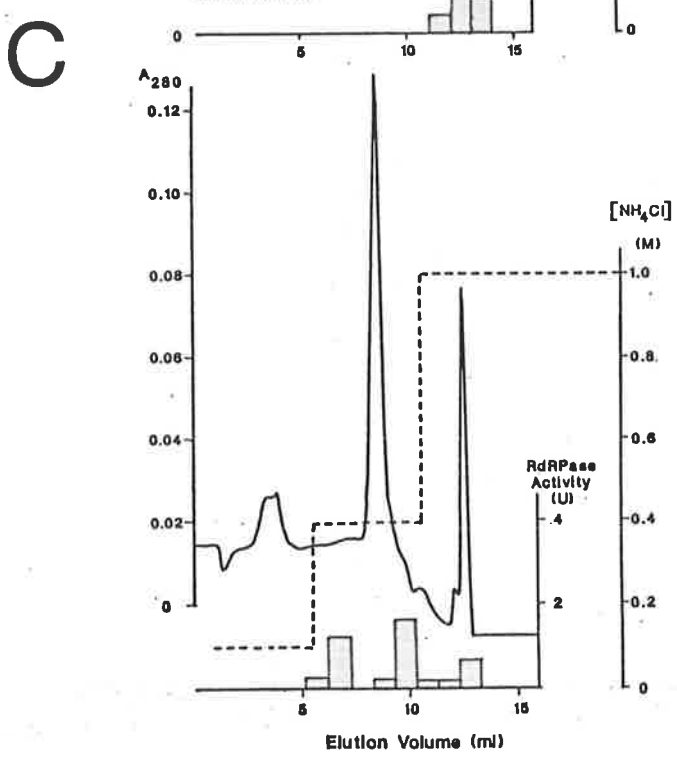
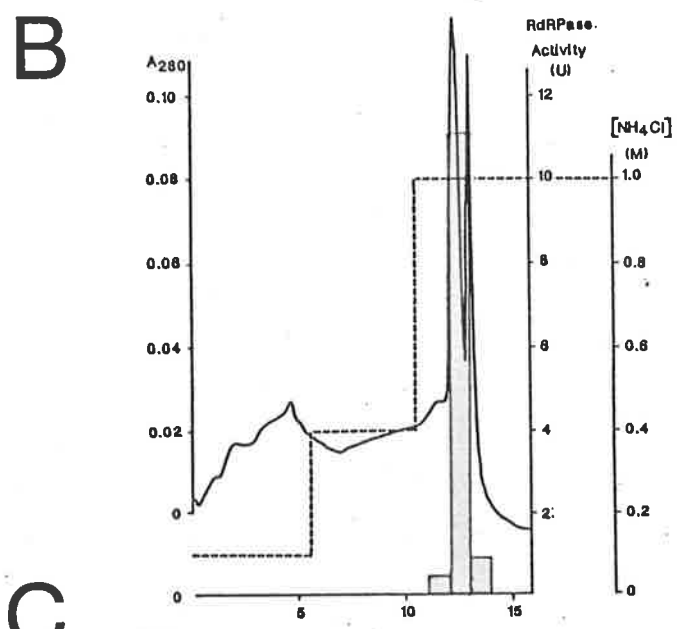
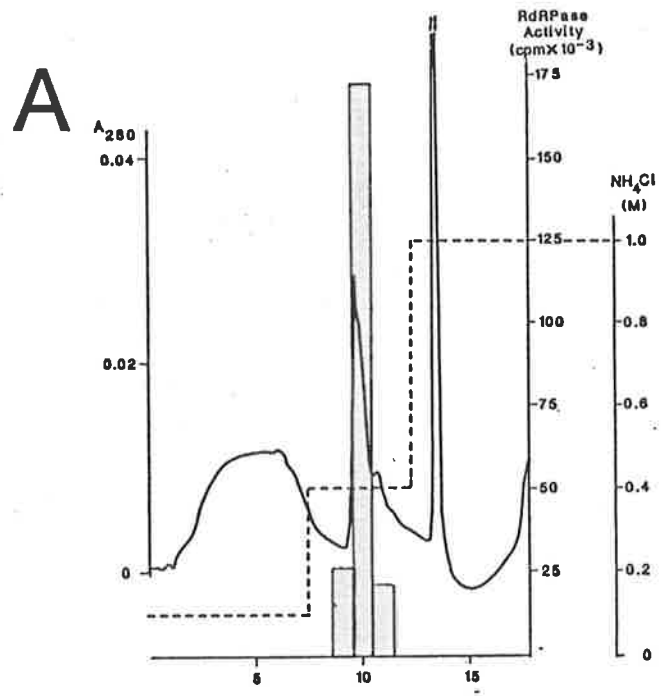
Samples of PEG-Enzyme (0.76 mg/ml) were diluted 1/20 in TGM buffer (A,B) or TGM buffer without glycerol (C) and loaded onto a MONO Q column equilibrated with the same buffer containing 0.1 M NH<sub>4</sub>Cl. Columns were run at room temperature (approx. 25°C; A) or at 4°C (B,C). The total sample volume was 5 ml in (A) and 2.5 ml in (B) and (C). All columns were developed with the same stepwise elution scheme involving successive 5 ml washes with TGM + 0.4 M NH<sub>4</sub>Cl and TGM + 1.0 M NH<sub>4</sub>Cl (dotted line). The flow rate was 1 ml/min. Fractions containing RdRPase (shaded boxes) were detected by assaying aliquots from each 1 ml fraction. The NH<sub>4</sub>Cl concentration was adjusted to 80 mM in all assays. Recovery of RdRPase activity in each case is given in Table 4.4 (below). Since different detectors were used at room temperature and at 4°C, the absorbance profiles can be used for qualitative comparison only.

**TABLE 4.4: Effect of temperature and buffer composition on recovery of RdRPase from MONO Q columns using stepwise elution.**

	Temperature (°C)	Buffer	[NH <sub>4</sub> Cl] to elute RdRPase <sup>a</sup>	Recovery (%) <sup>b</sup>
A	25	TGM	0.1-0.4	50-80
B	4	TGM	0.4-1.0	60-80
C	4	TM (no glycerol)	0.1-0.4	< 20
	4	TG (no 2-ME)	0.4-1.0	< 1

a. Refers to NH<sub>4</sub>Cl concentration required to elute activity in the stepwise elution scheme shown in Figure 4.4.

b. Relative to PEG-Enzyme loaded.



column buffer produced results resembling those obtained using complete buffer (Figure 4.4B), but less than 1% of activity was recovered (Table 4.4). When glycerol was omitted, recovery was also markedly reduced (Table 4.4) and the enzyme eluted over several fractions following the 0.4 M  $\text{NH}_4\text{Cl}$  wash (Figure 4.4C). Hence omission of glycerol and chromatography at room temperature both resulted in elution of RdRPase much earlier than observed at 4°C. Since all other variables (source of enzyme, pH, flow rate, etc.) were identical in these experiments, the differences observed could only be due to the omission of glycerol or the increase in temperature, both of which are factors known to increase hydrophobic interactions (see Discussion).

#### **Effect of salt gradient elution.**

To further investigate the effects of salt gradient elution on recovery of RdRPase activity, identical samples of PEG-Enzyme were subjected to MONO Q chromatography at both 4°C and room temperature using either a steep salt gradient (300 mM  $\text{NH}_4\text{Cl}/\text{ml}$ ) or a flat gradient (30 mM  $\text{NH}_4\text{Cl}/\text{ml}$ ). The results confirmed earlier observations that the salt concentration required for elution of RdRPase is dependent on temperature (Figure 4.5 compare A and C with B and D). At room temperature, approximately 0.2 M  $\text{NH}_4\text{Cl}$  was sufficient to elute RdRPase, whereas at 4°C approximately 0.4 M was required. Use of the flat gradients resulted in good separation of the proteins in PEG-Enzyme preparations but gave lower recoveries of enzyme activity (Table 4.5) compared with those obtained using steep gradients. Excellent recoveries (85-100%) were obtained using steep gradients at 4°C, although resolution of protein peaks was relatively poor (Figure 4.5 A and B).

Recovery of enzyme activity was consistently lower at room temperature than at 4°C (Table 4.5). Since PEG-Enzyme diluted in TGM buffer shows no difference in stability when incubated at 4°C or at 25°C (Figure 4.2), the difference in recovery observed here must be a consequence of instability caused by chromatography.

**FIGURE 4.5: Gradient elution of RdRPase from MONO Q columns.**

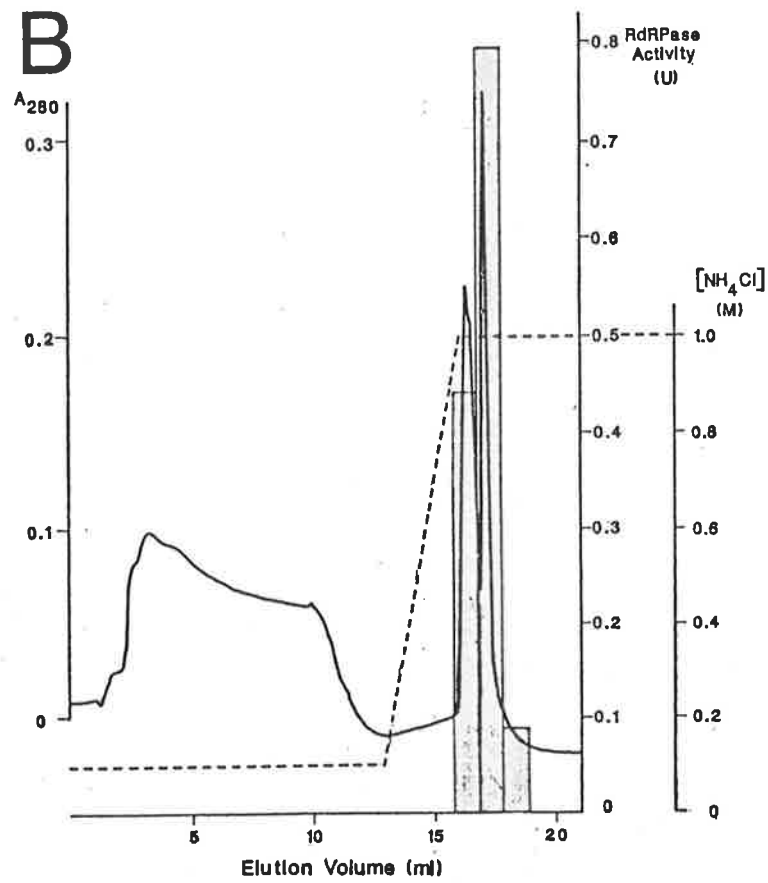
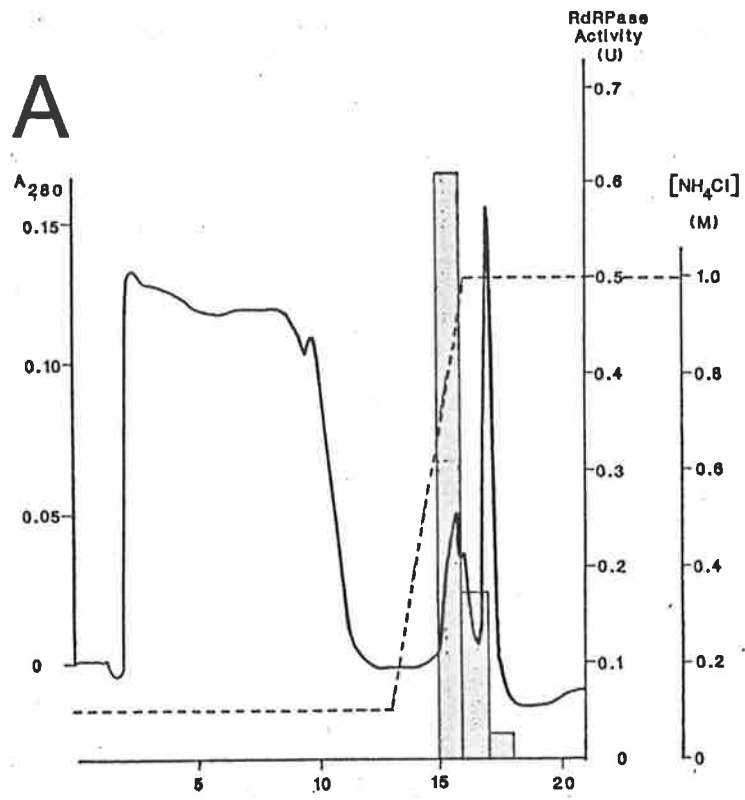
Identical 0.4 ml samples of PEG-Enzyme were diluted to 8 ml with TGM buffer and loaded onto a MONO Q column equilibrated with TGM buffer containing 0.1 M  $\text{NH}_4\text{Cl}$  at room temperature (approx. 25°C; A,C) or at 4°C (B,D, see next page). RdRPase was eluted using a gradient of  $\text{NH}_4\text{Cl}$  in TGM buffer from 0.1 M to 1.0 M (dotted line) over 3 ml (gradient rate =300 mM/ml; A,B) or 30 ml (gradient rate=30 mM/ml; C,D). The flow rate was 0.5 ml/min. Fractions containing RdRPase (shaded boxes) were detected by assaying 10  $\mu\text{l}$  aliquots from each 1 ml fraction. Recovery of RdRPase activity is given in Table 4.5. Since different detectors were used at room temperature and at 4°C, the absorbance profiles can be used for qualitative comparison only.

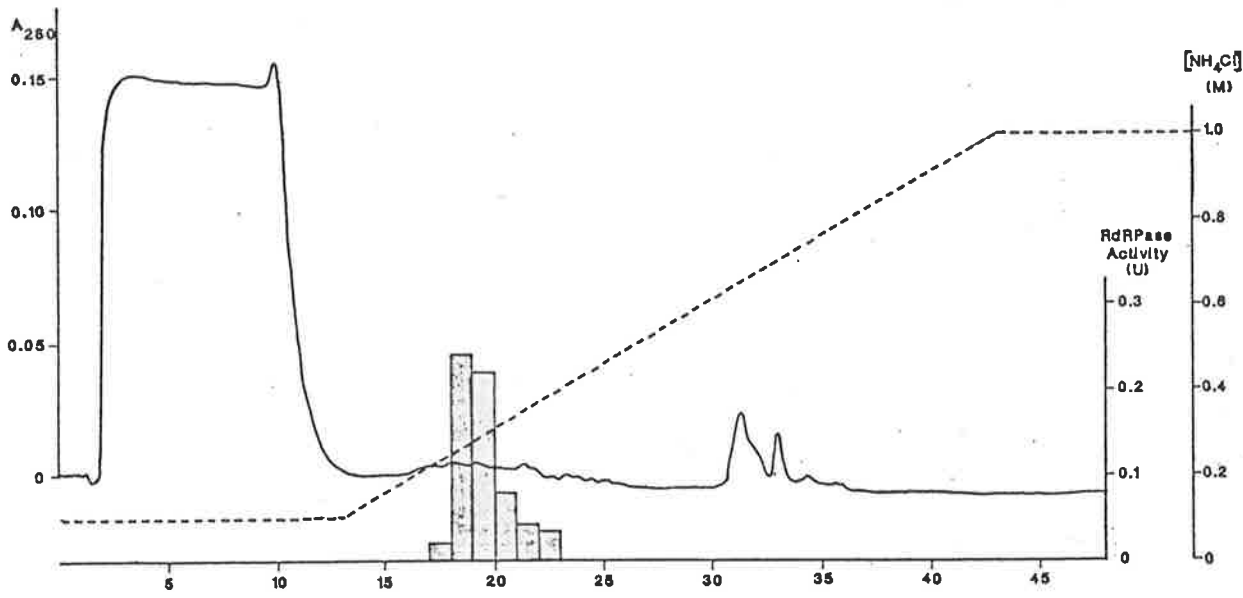
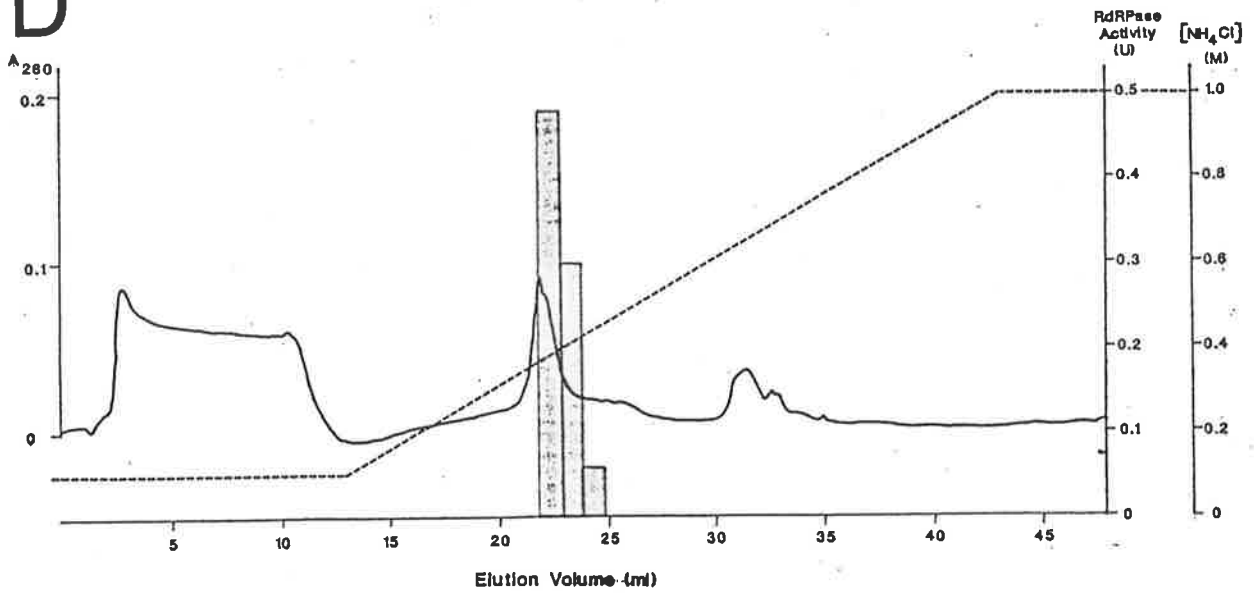
**TABLE 4.5: Effect of  $\text{NH}_4\text{Cl}$  gradient on recovery of RdRPase from MONO Q columns.**

Temperature (°C)	Recovery of RdRPase(%) <sup>a</sup>	
	Gradient rate ([ $\text{NH}_4\text{Cl}$ ]/ml) <sup>b</sup>	
	300	30
25	53	20
4	89	72

a. Relative to PEG-Enzyme loaded.

b. Refers to the rate of change of  $\text{NH}_4\text{Cl}$  concentration (mM per ml) in the gradient elution scheme shown in Figure 4.5



**C****D**

### **6. Chromatography of PEG-Enzyme on BLUE-TRISACRYL M.**

BLUE-TRISACRYL M was used as an alternative to Blue-Sepharose CL-6B or other chromatographic media containing Cibacron Blue F3GA, since the TRISACRYL support is reported to have good flow properties and a chemical structure which minimizes non-specific hydrophobic interactions (LKB handbook "ULTROGEL, MAGNOGEL AND TRISACRYL: Practical guide for use in affinity chromatography and related techniques", 2nd edition, pp 73-74). Various buffer conditions were investigated for loading and elution of PEG-Enzyme. The best recovery (40-45%) was obtained by loading PEG-Enzyme in TGM buffer containing 0.2 M  $\text{NH}_4\text{Cl}$  and 10 mM magnesium acetate and eluting RdRPase by removal of magnesium acetate from the buffer (results not shown). Loading at lower  $\text{NH}_4\text{Cl}$  concentration with or without magnesium sulphate, and elution using salt gradients did not improve recovery and made little difference to the profile of eluted proteins (results not shown).



## DISCUSSION

The overall aim of the investigations described here was to devise a strategy for the purification of as much RdRPase as possible from large amounts of CMV-infected cucumbers. Hence, the procedure for extraction of the enzyme, the time course of its activity during infection, its stability during storage and chromatography, and its behaviour on different chromatographic media have been investigated. The results provide solutions to some of the problems which must be overcome to enable large-scale preparation of RdRPase.

### **Method for extraction of RdRPase.**

A comparison of two alternative methods for the extraction of RdRPase from CMV-infected plants revealed that Method B, involving solubilization of enzyme from PF with magnesium sulphate, was superior to the ammonium sulphate extraction Method A. The PEG-Enzyme produced had a higher yield of RdRPase activity per gram of leaves, a higher specific activity and was clear and essentially colourless, therefore allowing it to be filtered prior to chromatography without loss of activity (Table 4.2). The effectiveness of this method was largely due to the ability of magnesium sulphate to solubilize a large proportion of RdRPase, but only a small proportion of protein, from PF (Gill *et al*, 1981) thus providing significant purification (Table 4.3). This procedure was successfully used to obtain a large amount of PEG-Enzyme for chromatographic studies.

### **Time-course of RdRPase activity.**

To maximize the yield of RdRPase from virus-infected plants they must be extracted at a stage of infection when the enzyme is at its highest level. Nearly all previous studies of CMV-induced RdRPase have used cotyledons at 7-12 days post-inoculation (p.i.) as their starting material (eg. Kumarasamy and Symons, 1979a; Gill *et al*, 1981; Khan *et al*, 1986). In contrast, Jaspars *et al* (1985) found that synthesis of double-stranded CMV RNAs and low molecular weight RNAs (the presumed product of the 100k protein, see Introduction) by PF was maximal at 2-3 days p.i. The time-course of RdRPase activity in CMV-infected cotyledons was therefore investigated using methods which extract soluble (Method A) and

solubilized particulate ( Method B) enzyme (Figure 4.1). Both methods showed a peak of enzyme activity at 6-9 days which is consistent with previous time-course studies (May *et al*, 1970; Peden *et al*, 1972). It seems likely that particulate CMV replicase activity, which is responsible for replication of viral RNA, is induced early in infection (Jaspars *et al*, 1985). If the 100k protein is a component of the replicase complex, it may only be released into the soluble phase or become "available" for solubilization after it has been used for replication of viral RNAs. This would account for its peak of activity at 6-9 days.

#### **Factors affecting enzyme stability.**

The major problem encountered during studies of RNA replicases of CMV and other viruses is their instability during purification and storage. Low temperatures (4°C) and buffers containing glycerol (up to 30%) and a reducing agent are essential for the preservation of enzyme activity (Kumarasamy, 1980; Kumarasamy and Symons, 1979a; Khan *et al*, 1986). Glycerol is known to have a strong stabilizing effect on proteins. This appears to be due to its interaction with the water molecules which form a hydration shell around proteins in solution. Denaturation of protein within the resultant structured solvent is thermodynamically less favourable than it would be in the absence of glycerol (Gekko and Timascheff, 1981). Many enzymes also require a reducing environment. This can be maintained by the addition of a sulphhydryl- containing reagent (such as DTT or 2-ME) and EDTA to remove traces of divalent cations which promote oxidation. All the solutions used in this work contained 10-30% glycerol and at least 5 mM 2-ME, and extraction and storage buffers also contained 1 mM EDTA. However, EDTA was omitted from chromatography buffers since it binds to anion exchange columns and can compete with binding of proteins.

Initial attempts to purify CMV RdRPase prepared by Method A (Table 4.1) were repeatedly frustrated by low or barely acceptable recoveries and wide variation in chromatographic behaviour between different preparations. Since successful large-scale purification required reliable procedures with reproducibly

high yields, some of the factors affecting the stability and recovery of RdRPase were investigated. Although the purified protein is not intended for enzymatic studies, maintenance of activity is still important so that it can be correlated unequivocally with the 100k protein in highly-purified preparations.

In agreement with Clark *et al* (1974), PEG-Enzyme preparations were found to be unstable when stored at  $-15^{\circ}\text{C}$ , but could be stored at  $-80^{\circ}\text{C}$ , with no apparent loss of activity, for at least 6 months. This allows PEG-Enzyme from many separate extractions to be "stock-piled" in the freezer and pooled together at a later stage for large-scale chromatography.

Over short periods of time (e.g., 8 hours), PEG-Enzyme was found to be relatively stable at  $4^{\circ}\text{C}$  or at room temperature (Figure 4.2A). When diluted in the buffer used for chromatography (TGM) it was even more stable again (Figure 4.2B). Omission of 2-ME or glycerol had little effect. From these results it appeared that loss of activity would not be a serious problem over the periods of time required for chromatography.

However, when PEG-Enzyme samples were loaded onto columns, significant losses of enzyme activity occurred, especially in the absence of glycerol or 2-ME (Table 4.4). Recovery was lower at  $25^{\circ}\text{C}$  than at  $4^{\circ}\text{C}$  (Table 4.4 and 4.5). These results are inconsistent with the stability of diluted PEG-Enzyme during incubation under the same conditions (see above), implying that instability is a direct result of the chromatographic process. There are several possible ways in which enzyme inactivation could be occurring. Firstly, it could be due simply to dilution inactivation of the enzyme as it is separated from other proteins on the column. In a dilute form it would probably be more dependent on the stabilizing effects of glycerol and 2-ME. However, preliminary results of MONO Q chromatography using larger samples (not shown) suggest that recovery of activity is not dependent on sample size. Secondly, contaminating proteases could be digesting the enzyme during its purification. Partially purified RdRPase preparations have been found to contain trace amounts of protease activity for which an inhibitor could not be found (Clark *et al*, 1974). Proposed cleavage of

the 110k protein present in polymerase preparations to form the active 100k protein (Khan *et al*, 1986) would require the presence of a protease and is consistent with the finding of Gill (1983) that 110k protein is degraded during storage and manipulation of purified enzyme fractions. However, if protease activity contributes to instability during chromatography, it should also have been observed during incubation of PEG-Enzyme samples in chromatography buffer.

Finally, unfavourable non-specific interactions may be occurring between the column and the enzyme, resulting in its denaturation. Hydrophobic interactions, for example, are strengthened by higher temperatures and lower concentrations of glycerol, which are also the conditions under which enzyme recoveries are reduced (Table 4.4). Further evidence for the occurrence of these interactions is presented later.

Recovery of RdRPase from MONO Q columns was strongly dependent on the slope of the salt gradient used for elution (Figure 4.5, Table 4.5). These results are in agreement with earlier studies (Clark *et al*, 1974; Kumarasamy and Symons, 1979a; Kumarasamy, 1980) which reported lower recoveries of polymerase activity from ion-exchange columns using gradient rather than stepwise elution. The simplest explanation of all these results is that the enzyme is composed of more than one subunit, each differing slightly in ionic properties, which are separated (at least partially) by gradient elution. This could be tested by mixing aliquots from adjacent gradient-eluted fractions and testing for enhancement of activity in RdRPase assays. However, this would require loading large samples of enzyme to overcome the extensive dilution of individual polypeptides during chromatography.

The results of Figure 4.5 also clearly illustrate the importance of gradient elution for resolution of individual protein peaks. Gradients with slopes intermediate between the two extremes used here would provide an effective compromise between the high recoveries obtained using stepwise or steep gradient elution and the improved resolution afforded by a flatter gradient. The recoveries obtained here using gradient elution from MONO Q columns (Table 4.5)

are higher than those previously reported using any other columns (Clark *et al.*, 1974; Kumarasamy, 1980; Takanami and Fraenkel-Conrat, 1982), suggesting that the MONO Q column will be a valuable step in future purification procedures.

### **Evidence for hydrophobic interactions.**

The difference observed in the salt concentration required to elute RdRPase from MONO Q columns at different temperatures and in the presence of different concentrations of glycerol (Figures 4.4 and 4.5) suggests the involvement of hydrophobic as well as ionic interactions. Increased temperature, lower concentrations of glycerol and higher concentrations of salt are all known to strengthen hydrophobic interactions, which are the result of thermodynamically favourable attraction between "patches" of hydrophobic residues on the surface of proteins (Ochoa, 1978; Scopes, 1982). Hydrophobic patches can also interact with chromatographic media, especially if they contain non-polar chains used to attach ligands or charged groups. Interactions of this type are important for hydrophobic interaction and affinity chromatography, but undesirable and usually absent in other forms of chromatography. In the case of RdRPase, however, behaviour on MONO Q columns is apparently dependent on a balance between hydrophobic and ionic interactions. Increased temperature and lower concentrations of glycerol resulted in earlier elution, i.e., weaker binding to the column (Figure 4.4). This is the opposite to what would be expected if increased hydrophobic interactions were simply causing a stronger attraction between the enzyme and the column. However, it could be explained by proposing that the enzyme is "sticky" and has a tendency to aggregate with itself or with other proteins via hydrophobic interactions. At 4°C, in the presence of glycerol, the enzyme would bind to the MONO Q column via hydrophobic and ionic interactions and would be eluted as the salt concentration is increased (Figure 4.4B). However, if the hydrophobic forces are strengthened by, for example, higher temperatures or the removal of glycerol, the hydrophobic patches on the protein may interact with each other rather than with the column, leading to an overall weaker binding to the column. A lower salt concentration would then be required to elute the

enzyme as observed in Figure 4.4 (A and C). Unfavourable hydrophobic association with columns would also explain the extremely low recoveries obtained when enzyme samples without glycerol were loaded on HIC columns (see Results) and when samples were run on Superose gel filtration columns with low glycerol buffers (results not shown). Experiments could be devised to test these proposals further. However, for the purpose of purification, the most important conclusion is that hydrophobic interactions may affect the chromatographic behaviour of RdRPase and contribute to its instability in the absence of glycerol.

#### **Development of a chromatographic purification scheme.**

To obtain substantial amounts of CMV-induced RdRPase for amino acid sequencing studies requires a well-designed and highly efficient purification scheme. The columns must give reproducibly high enzyme recoveries and allow large quantities of sample to be loaded. Because of the instability of the enzyme, high flow rates are important so that the purification can be completed as rapidly as possible.

Since the starting material, PEG-Enzyme, is in a buffer containing high salt, hydrophobic interaction chromatography (HIC) is an obvious choice for initial column purification. However, because of the variable results obtained with HIC columns, an alternative initial purification step must be found. The best alternative is ion-exchange chromatography which would require dilution of the enzyme to reduce the salt concentration before loading. Although dialysis could be used, Gill (1983) found that higher enzyme yields were obtained following chromatography if the sample was diluted rather than dialyzed. In addition, results of ion-exchange chromatography are more reproducible if a dilute sample is used (Scopes, 1982). Results reported here (Table 4.1) and elsewhere (Kumarasamy and Symons, 1979a; Gill *et al.*, 1981; Takanami and Fraenkel-Conrat, 1982) have shown that stepwise chromatography on phosphocellulose is reproducible and affords good purification although its poor flow properties make it unsatisfactory for large-scale column application. It could, however, be effectively used for batch adsorption, which would allow a large volume of sample

to be bound and eluted quickly. Although acting basically as a cation exchanger, phosphocellulose also has "pseudo-affinity" properties, especially towards enzymes with phosphate ester substrates (Scopes, 1982). This could explain its successful application to polymerase purification.

FPLC ion exchange media, particularly MONO Q, proved to be highly effective for chromatography of RdRPase. Recovery of enzyme activity was reproducibly high despite apparent hydrophobic interactions between the enzyme and the column. In addition, excellent flow properties and the ability to manipulate the elution conditions to obtain a high degree of purification, indicate that chromatography on MONO Q (or Q-Sepharose Fast flow) is an ideal step for RdRPase purification.

A recent comparison of techniques used for protein purification (Bonnerjea *et al*, 1986) found that affinity chromatography was by far the most effective purification step, although it also had the lowest average yield (61%). It is therefore desirable to include at least one affinity step in the RdRPase purification scheme. This should be preceded by other steps (for example, ion exchange) to achieve a partial purification and decrease the volume of sample to be loaded onto the affinity column. Various affinity adsorbents have been used during purification of RdRPsases, mostly containing ligands such as heparin and poly(C) or dye-ligands such as Cibacron-Blue F3GA, which have affinity for a group of related proteins and hence are termed "group-specific" adsorbents.

Poly(C)-cellulose has been used previously (Kumarasamy and Symons, 1979a; Gill *et al*, 1981) for purification of CMV RdRPase since the enzyme can copy poly(C) with high efficiency (May and Symons, 1971). However, the recovery of activity was previously very low and in this study no enzyme activity was recovered from poly(C) columns (Table 4.1). Much better results have been obtained using dye-ligand chromatography on Blue Sepharose CL-6B (Gordon *et al*, 1982; Takanami and Fraenkel-Conrat, 1982; this work), although results were variable and large losses of activity sometimes occurred (Table 4.1). TRISACRYL-BLUE M was found to be more reproducible than Blue Sepharose

CL-6B, albeit with lower enzyme recovery. It is likely that higher recovery and a greater degree of purification of bound RNA polymerase could be obtained using affinity elution with ATP or poly(C) instead of, or combined with, an increase in the salt concentration.

Potentially the most useful column for purification of viral RNA polymerase is one containing viral RNA. Mouches *et al* (1984) used a TYMV RNA-Sepharose column to reasonably good effect during purification of TYMV RNA replicase. This column was prepared by coupling TYMV RNA to cyanogen bromide-activated Sepharose 4B (Schofield and Williams, 1977). However, the binding capacity and efficiency of an RNA affinity column could be greatly improved by using a column of avidin-Sepharose to which biotinylated RNA has been bound. This technique has been used to obtain 3,400-fold purification of lac repressor protein from crude *E. coli* extracts (Leblond-Francilland *et al*, 1987). Since CMV RNA can be readily biotinylated using Photobiotin (Forster *et al*, 1985; results not shown), a CMV RNA-avidin-Sepharose column could be an effective purification step.

It is unlikely that gel filtration chromatography would be a useful step for large-scale purification of CMV-induced RdRPase, due to the relatively small sample volumes required and the low recoveries of enzyme activity obtained in preliminary experiments (results not shown). The elution of RdRPase in the void volume during Sephacryl S-200 and S-300 chromatography of Method A PEG-Enzyme (Table 4.1) suggests that the enzyme may be present as a high molecular weight complex. Dorssers *et al* (1983) observed that the CPMV RNA replication complex eluted in the void volume of a Sepharose 2B column and utilized this to obtain substantial purification of the complex from many contaminating proteins, including the host-coded RdRPase. It may therefore be possible to use a similar technique to obtain at least partially intact CMV RNA replicase to enable further characterization of its polypeptide composition.



**Future prospects for purification of CMV-induced RdRPase.**

Progress towards the large-scale purification of CMV RdRPase has been difficult and frustratingly slow. Problems of loss of activity and lack of reproducibility have been overcome or at least the factors affecting them are now more clearly understood (see above). Although the original goal of purifying the 100k polymerase for amino acid sequencing has not been realized, results obtained during this work and previous studies, and summarized in the foregoing discussion, suggest that this project may still be feasible. Using PEG precipitation and two column steps, Gill *et al* (1981) obtained a 53- to 63-fold purification of RdRPase present in the magnesium sulphate-solubilized supernatant to yield a preparation of similar specific activity to the 10,000-fold purified soluble enzyme (Kumarasamy and Symons, 1979a). Two additional columns (with unknown purification factors) were required to obtain the highly-purified fractions consisting mainly of the 100k protein (Gordon *et al*, 1982). A large-scale purification scheme will therefore probably require a several thousand-fold purification of PEG-Enzyme which is within the theoretical possibilities of an efficient 3 or 4 step procedure using the column techniques discussed above.

A suitable purification scheme would start with PEG-Enzyme prepared by Method B and could then utilize, for example, the following steps: batch adsorption and stepwise elution of diluted PEG-Enzyme from phosphocellulose; salt gradient elution from MONO Q or Q-Sepharose Fast Flow; Stepwise or affinity elution from a medium-scale column of BLUE-TRISACRYL M; and affinity chromatography on biotinylated CMV RNA-avidin-Sepharose.

According to earlier calculations, 2.5 kg of plants would be required for purification of 10  $\mu$ g of 100k protein, assuming 4 purification steps each with 80% recovery (see Introduction). However, unexpected losses or lower recovery during purification would rapidly increase the amount of CMV-infected leaves required. If, for example, each step gave 60% recovery, then nearly 8 kg of leaves would be needed to obtain 10  $\mu$ g of purified 100k protein. Growth, inoculation and extraction of plants and purification of enzyme on this scale becomes a major

logistical problem rather than a biochemical one.

Alternatively, techniques other than amino acid sequencing of purified protein could be used to obtain information about the 100k protein and other proteins involved in CMV RNA replication. Some possible alternative approaches are discussed in Chapter 5.

**CHAPTER FIVE.**

**CONCLUDING DISCUSSION.**

This work together with that of Davies and Symons (1988) has provided the complete primary sequence of the CMV genome and allowed the prediction of some secondary structural elements which may be important for its function. Studies on other Tricornaviridae, most notably BMV, have provided information on the role of conserved 5'- and 3'-terminal structures. We have suggested (Chapter 2) that the 5' termini of CMV RNAs are important not only for initiation of CMV (+)RNA synthesis but also for interaction with CMV satellite-RNA. However, more information is needed in order to understand the role of structural features in expression of the CMV genome. Two basic approaches could be taken to acquire this information.

Firstly, the sequencing of RNAs of different strains of CMV, particularly at their 5' and 3' termini, would be useful in determining those structures which are conserved between strains and therefore probably functionally significant. Limited sequence data available for the 3' termini of 4 different strains of CMV (Barker, Wilson and Symons, unpublished data, cited in Symons, 1985) have provided further evidence for the highly conserved structure of this region. However no data are available for the 5' termini of different CMV strains.

Secondly, the construction of full-length cDNA clones of CMV RNAs from which infectious transcripts can be produced, would allow manipulation of the CMV genome at the DNA level and investigation of the effects of these changes on replication of the virus and its interaction with sat-RNA. Infectious transcripts from cDNA clones of bacterial, animal, plant and insect viruses have been reported (for example, Taniguchi *et al* 1978; van der Werf *et al* 1986; Ahlquist *et al* 1984b; Dawson *et al* 1986, Dasmahaptra *et al* 1986) and will prove valuable for investigating the molecular biology of RNA viruses.

Full-length clones of CMV RNAs could also aid our understanding of the function of CMV-encoded proteins, since techniques for the expression of cloned genes in prokaryotic and eukaryotic cells are now well established. The application of expression systems to the study of RNA viruses is elegantly illustrated by the recent work of Morrow *et al* (1987) who achieved expression of

enzymatically active poliovirus RNA polymerase in *E. coli* cells.

Knowledge of the nucleotide sequence of the CMV genome has provided the amino acid sequences of the CMV-encoded proteins. Analysis of these sequences (Chapter 3) has indicated that CMV P1 and P2 are probably involved in viral RNA replication. However direct evidence to support these suggestions is very limited. Although non-structural proteins of all viruses are present at relatively low levels, those of CMV appear to be particularly scarce, making them very difficult to study. Other than the coat protein, no CMV-coded proteins have been detected in infected leaves or protoplasts, even using two-dimensional electrophoresis of proteins labelled during infection (Gonda and Symons, 1979; A.R. Gould, J. Haseloff and R.H.Symons, unpublished data). Treatment with actinomycin D or UV-irradiation of protoplasts did not enhance detection of CMV-encoded proteins (Gonda, 1979). In contrast, viral non-structural proteins synthesized in infected protoplasts (following suppression of host protein synthesis) have been detected in the case of BMV (Kiberstis *et al*, 1981), AMV (Samac *et al*, 1983; Joshi *et al*, 1984) and TMV (Siegel *et al*, 1978). Detection of TMV-encoded proteins can be further enhanced by using heat-shock to suppress host protein synthesis (Dawson and Boyd, 1987). TMV protein synthesis is not inhibited by heat-shock therefore allowing detection of viral proteins above a reduced background of host protein synthesis. Preliminary investigations of CMV-infected cucumber tissue using this technique resulted in the detection of proteins with electrophoretic mobilities corresponding to the 3a protein (Mr 30,000) and coat protein (Mr 24,000), but no detectable high molecular weight products (results not shown)

Antibodies have proved to be useful tools for the study of viral non-structural proteins. In particular, they have allowed the detection and *in situ* localization of all three non-structural proteins of AMV (Berna *et al*, 1984, 1985, 1986; Stussi-Garaud *et al*, 1987) and TMV (Ooshika *et al*, 1984; Tomenius *et al*, 1987; Saito *et al*, 1987). The subcellular location of a protein provides information relevant to its possible function. For example, immunocytochemistry was used to confirm that the viral-coded component of TYMV RNA replicase was

located on the chloroplast membrane, where viral RNA synthesis had been previously shown to occur (Garnier *et al* 1986). Antibodies against CMV-encoded proteins would be extremely valuable, especially for further characterization of possible viral-coded replicase components. However, synthetic peptides corresponding to regions of CMV P1 and P2 proteins with high predicted hydrophilicity, had low antigenicity in rabbits and failed to yield useful antisera (C. Davies, unpublished data). An alternative approach would be to raise antibodies against the *in vitro* translation products of CMV RNAs or to use the technique of *in vitro* immunization (Reading, 1982) to produce monoclonal antibodies against partially purified RdRPase proteins. This technique has been used successfully to generate antibodies from nanogram quantities of antigen (Borrebaeck, 1984) and is therefore ideally suited the study of CMV RNA polymerase. Antibodies against (even partially purified) RNA polymerase preparations could be used in antibody-linked polymerase assays (van der Meer *et al* 1983) to identify polypeptides with polymerase activity.

Since antibodies against CMV-encoded proteins were not available and peptide mapping studies had proven inconclusive (Gordon *et al*, 1982, see Chapter 4 Introduction) a more direct approach was necessary to investigate the involvement of viral proteins in CMV-induced RdRPase. Hence, the studies presented in Chapter 4 were undertaken with the aim of obtaining amino acid sequence data for the 100k protein. Although this is the most unambiguous method for identification of this protein, it is also undoubtedly the most difficult due to the low amounts of virus-specific proteins in CMV-infected cells. Using ultrasensitive silver staining (Merril *et al* 1981), no virus-specific proteins could be detected in crude (PEG-Enzyme) RdRPase preparations or in phosphocellulose-purified fractions (results not shown) even though both procedures resulted in an increase in the specific activity of the enzyme. Further purification (Gill *et al* 1981; Gill, 1983) was necessary to detect differences between RdRPase preparations from CMV-infected and healthy plants. In contrast, crude solubilized RNA replicase preparations from BMV-infected barley contain stainable amounts

of a virus-specific protein which has an HPLC peptide map identical to that of BMV P1 (Bujarski *et al* 1982). There have been no reports of amino acid sequencing of plant viral non-structural proteins. However, Rice *et al* (1986) have utilized cell culture and immunoaffinity chromatography to obtain sufficient amounts of several alphavirus proteins for partial amino acid sequencing.

Although no amino acid sequence data was obtained for the CMV-induced 100k protein, the results of chromatographic trials (Chapter 4) suggested that large-scale purification is a feasible project. If the amount of purified protein obtained was not sufficient for sequencing, high-resolution peptide mapping using two-dimensional gel electrophoresis or HPLC to separate peptides would be an informative alternative. Currently available techniques require only submicrogram quantities of polypeptide to generate a peptide map (Lischwe and Ochs, 1982; Zingde *et al* 1986; Blackburn, 1986).

The ultimate aim of all these investigations is to understand the roles played by both virus- and host-coded proteins in CMV replication. This will require not only an understanding of individual proteins, but also their interactions within the membrane-bound replication complex. Initial attempts to obtain "intact" CMV RNA replicase by solubilization of the particulate fraction (PF) with 12-M, as has been done for BMV (Bujarski *et al* 1983) and AMV (Houwing and Jaspars, 1986), produced preparations which were template-dependent, but had low overall enzyme activity (results not shown). This technique was not pursued further since the attention of investigations was focused on purifying the 100k protein. The success of future studies on the CMV RNA replication complex will rely on the availability of suitable preparations of RNA replicase and on the application of techniques to study protein-protein and protein-nucleic acid interactions. For example, nucleotide photoaffinity analogues (Czarnecki *et al* 1979; Droms *et al* 1987) could be a valuable tool, since they would allow covalent cross-linking of RNA to neighbouring proteins in the replication complex. This would be especially effective if used in association with antibodies to identify cross-linked proteins.

It will be apparent from the foregoing discussion that there have been some

difficulties in the study of CMV replication. Initial work on CMV RNA replicase in our laboratory (Kumarasamy and Symons, 1979; Gill *et al* 1981; Gordon *et al* 1982, Jaspars *et al* 1985) was promising, however the further investigations reported here have shown that the low level of virus-induced proteins in CMV-infected cucumbers is a major problem for detailed study. The development of new techniques, and the refinement of existing ones, to allow the detection and characterization of ever-decreasing amounts of protein, will certainly facilitate our future understanding of the function of the CMV genome.



## APPENDIX A

### GROUPS AND FAMILIES OF VIRUSES INFECTING PLANTS

The 24 groups and 2 families of plant viruses infecting plants have been listed according to the nature and number of their genome components based on Matthews (1982) and Francki *et al* (1985). The data on the number of viruses in each group are from Matthews (1982). In some cases, more recent estimates are available (see for example Francki, 1983; Francki *et al*, 1985), but these have been ignored for simplicity. More detailed information on the tripartite (+)RNA viruses with polyhedral particles (the Tricornaviridae) is given in Chapter 1, Table 1.1.

In addition to these groups, Matthews (1982) lists a possible group of 4 viruses typified by velvet tobacco mottle virus. A meeting of the ICTV during the 6th International Congress of Virology in September, 1984, approved the formation of two additional groups named after their type members, rice stripe virus and maize rayado fino virus (Brown, 1986).

Genome nucleic acid	No. of genome segments	Particle morphology	Group or Family	Type member	Number of members
dsRNA <sup>a</sup>	1	isometric	Caulimovirus	Cauliflower mosaic virus	6
ssDNA <sup>a</sup>	2	geminatae	Geminivirus	Maize streak virus	5
dsRNA	10-12	isometric	Reoviridae <sup>b</sup> 1. <i>Phytoreovirus</i> 2. <i>Fijivirus</i>	Wound tumour virus Fiji Disease virus	2 8
ss(-)RNA	1	bacilliform	Rhabdoviridae <sup>c</sup> 1. Subgroup A 2. Subgroup B	Lettuce necrotic yellows virus Potato yellow dwarf virus	4 4
	4	spherical	Tomato spotted wilt virus group	Tomato spotted wilt virus*	1
ss(+)RNA	1	isometric	Maize chlorotic dwarf virus group	Maize chlorotic dwarf virus*	1
			Tymovirus	Turnip yellow mosaic virus	17
			Luteovirus	Barley yellow dwarf virus	14
			Tombusvirus	Tomato bushy stunt virus	7
			Sobemovirus	Southern Bean mosaic virus	2
			Tobacco necrosis <sup>d</sup> virus group	Tobacco necrosis virus*	1
		flexuous rod	Closterovirus	Beet yellows virus	11
			Carlavirus	Carrotin latent virus	23
			Potyvirus	Potato virus Y	48
			Potexvirus	Potato virus X	18
		rigid rod	Tobamovirus	Tobacco mosaic virus	10

Genome nucleic acid	No. of genome segments	Particle morphology	Group or Family	Type member	Number of members
ss(+)RNA	2	isometric	Dianthovirus	Camelion ringspot virus	3
			Comovirus	Cowpea mosaic virus	12
			Nepovirus	Tobacco ringspot virus	22
			Pea enation mosaic virus group	Pea enation mosaic virus*	1
	3	rigid rod	Tobravirus	Tobacco rattle virus	2
			Cucumovirus	Cucumber mosaic virus	3
		isometric	Bromovirus	Brome mosaic virus	3
			Ilarvirus	Tobacco streak virus	11
		bacilliform	Alfalfa mosaic virus group	Alfalfa mosaic virus*	1
			rigid rod	Hordeivirus	Barley stripe mosaic virus

- a. DNA of Caulimoviruses and Geminiviruses is circular. The nucleic acid of all other viruses listed is linear.
- b. Members of the family Reoviridae infecting plants have been divided into two genera - *Phytoreovirus* and *Fijivirus*.
- c. Members of the family Rhabdoviridae infecting plants have been divided into two subgroups A and B with lettuce necrotic yellows virus and potato yellow dwarf virus as their respective type members.
- d. The tobacco necrosis virus group has been given the name Necrovirus in the newly proposed taxonomic revisions by the ICTV (Brown, 1986).
- \* The type member is the sole member of this virus group.



**APPENDIX B**

**PUBLICATIONS**

**PUBLICATIONS.**

- Rezaian, M.A., Williams, R.H.V., Gordon, K.H.J., Gould, A.R. and Symons, R.H. (1984) Nucleotide sequence of cucumber mosaic virus RNA 2 reveals a translation product significantly homologous to corresponding proteins from other viruses. *Eur. J. Biochem.* **143**,277-284.
- Rezaian, M.A., Williams, R.H.V. and Symons, R.H. (1985) Nucleotide sequence of cucumber mosaic virus RNA 1. Presence of a sequence complementary to part of the viral satellite RNA and homologies with other viral RNAs. *Eur. J. Biochem.* **150**,331-339.

**ABSTRACTS.**

- Rezaian, M.A., Williams, R.H.V., Gordon, K.H.J., Gould, A.R. and Symons, R.H. (1984) Sequence and structure of the genome of cucumber mosaic virus., 6th Annual Conference on the Organization and Expression of the Eukaryotic Genome, Lorne, Australia.
- Gordon, K.H.J., Rezaian, M.A., Williams, R.H.V. and Symons, R.H. (1984) A multitude of sub-genomic RNAs formed during infection of cucumber by cucumber mosaic virus., *ibid.*
- Rezaian, M.A., Williams, R.H.V., Gordon, K.H.J. and Symons, R.H. (1984) The primary structure of the four RNAs of cucumber mosaic virus., 6th International Congress of Virology, Sendai, Japan.
- Rezaian, M.A., Williams, R.H.V.G. and Symons, R.H. (1985) The structure of the cucumber mosaic virus genome: Evolutionary relationship between plant viruses., 7th Annual Conference on the Organization and Expression of the Eukaryotic Genome, Lorne, Australia.

REFERENCES.

- Ahlquist, P., Dasgupta, R., Kaesberg, P. (1981a) Near identity of 3' RNA secondary structures in bromoviruses and cucumber mosaic virus. *Cell* **23**,183-189.
- Ahlquist, P., Dasgupta, R., Kaesberg, P. (1984a) Nucleotide sequence of the brome mosaic virus genome and its implications for viral replication. *J. Mol. Biol.* **172**,369-383.
- Ahlquist, P., French, R. and Bujarski, J.J. (1987) Molecular studies of brome mosaic virus using infectious transcripts from cloned DNA. *Adv. Virus Res.* **32**,215-242.
- Ahlquist, P., French, R., Janda, M. and Loesch-Fries, S. (1984b) Multicomponent RNA plant virus infection from cloned viral cDNA. *Proc. Natl. Acad. Sci. (USA)* **81**,7066-7070.
- Ahlquist, P., Luckow, V. and Kaesberg, P. (1981b) Complete nucleotide sequence of brome mosaic virus RNA 3. *J. Mol. Biol.* **153**,23-38.
- Ahlquist, P., Strauss, E.G., Rice, C.M., Strauss, J.H., Haseloff, J. and Zimmern, D. (1985) Sindbis virus proteins nsP1 and nsP2 contain homology to non-structural proteins from several RNA viruses. *J. Virol.* **53**,536-542.
- Allison, R., Johnston, R.E. and Dougherty, W.G. (1986) The nucleotide sequence of the coding region of tobacco etch virus genomic RNA: Evidence for the synthesis of a single polyprotein. *Virology* **154**,9-20.
- Andrews, N.C. and Baltimore, D. (1986) Purification of a terminal uridylyltransferase that acts as host factor in the *in vitro* poliovirus replicase reaction. *Proc. Natl. Acad. Sci. (USA)* **83**,221-225.
- Argos, P., Kamer, G., Nicklin, M.J.H. and Wimmer, E. (1984) Similarity in gene organization and homology between proteins of animal picornaviruses and a plant comovirus suggest common ancestry of these virus families. *Nuc. Acids Res.* **12**,7251-7267.
- Avila-Rincon, M.J., Collmer, C.W. and Kaper, J.M. (1986) *In vitro* translation of cucumoviral satellites. I. Purification and nucleotide sequence of cucumber mosaic virus-associated RNA 5 from cucumber mosaic virus strain S. *Virology* **152**,446-454.
- Barker, R.F., Jarvis, N.P., Thompson, D.V., Loesch-Fries, L.S. and Hall, T.C. (1983) Complete nucleotide sequence of alfalfa mosaic virus RNA 3. *Nuc. Acids Res.* **11**,2881-2891.
- Bennetzen, J.L. and Hall, B.D. (1982) Codon selection in yeast. *J. Biol. Chem.* **257**,3026-3031.
- Berna, A., Briand, J-P., Stussi-Garaud, C. and Godefroy-Colburn, T. (1986) Kinetics of accumulation of the three non-structural proteins of alfalfa mosaic virus in tobacco plants. *J. Gen. Virol.* **67**,1135-1147.
- Berna, A., Briand, J-P., Stussi-Garaud, C., Godefroy-Colburn, T. and Hirth, L. (1984) Immunodetection of a non-structural protein of alfalfa mosaic virus (P2) in infected tobacco plants. *Ann. Virol. (Inst. Pasteur)* **135E**,285-296.
- Berna, A., Godefroy-Colburn, T. and Stussi-Garaud, C. (1985) Preparation of an antiserum against an *in vitro* translation product of alfalfa mosaic virus RNA 3. *J. Gen. Virol.* **66**,1669-1678.

- Bienz, K., Egger, D. and Pasamontes, L. (1987) Association of polioviral proteins of the P2 genomic region with the viral replication complex and virus-induced membrane synthesis as visualized by electron microscopic immunocytochemistry and autoradiography. *Virology* **160**,220-226.
- Blackburn, S. (1986) (ed.) Techniques of peptide mapping. in "*CRC Handbook of Chromatography: Peptides*" vol I, pp 225-229.
- Blumenthal, T. and Carmichael, G.C. (1979) RNA replication: function and structure of Q $\beta$  replicase. *Ann. Rev. Biochem.* **48**,525-548.
- Bonnerjea, J., Oh, S., Hoare, M. and Dunhill, P. (1986) Protein purification: The right step at the right time. *Biotechnology* **4**,954-958.
- Borrebaeck, C.A.K. (1984) Dependence on T-cell-replacing factor and immunogenic dose for the production of monoclonal antibodies using the *in vitro* immunization technique. *Molec. Immunol.* **21**,841-845.
- Bouzoubaa, S., Quillet, L., Guilley, H., Jonard, G. and Richards, K. (1987) Nucleotide sequence of beet necrotic yellow vein virus RNA 1. *J. Gen Virol.* **68**,615-626.
- Bradford, M.M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein using the principle of protein-dye binding. *Anal. Biochem.* **72**,248-254.
- Bradley, D.W. and Zaitlin, M. (1971) Replication of tobacco mosaic virus. II. The *in vitro* synthesis of high molecular weight virus-specific RNAs. *Virology* **45**,192-199.
- Brederode, F.Th., Koper-Zwarthoff, E.C. and Bol, J.F. (1980) Complete nucleotide sequence of alfalfa mosaic virus RNA 4. *Nuc. Acids Res.* **8**,2213-2223.
- Brown, F. (1986) The classification and nomenclature of viruses: Summary of results of meetings of the International Committee on Taxonomy of viruses in Sendai, September 1984. *Intervirology* **25**,141-143.
- Bujarski, J.J., Ahlquist, P., Hall, T.C., Dreher, T.W. and Kaesberg, P. (1986) Modulation of replication, aminoacylation and adenylation *in vitro* and infectivity *in vivo* of BMV RNAs containing deletions within the multifunctional 3' end. *EMBO J.* **5**,1769-1774.
- Bujarski, J.J., Dreher, T.W. and Hall, T.C. (1985) Deletions in the 3' terminal tRNA-like structure of brome mosaic virus RNA differentially affect aminoacylation and replication *in vitro*. *Proc. Natl. Acad. Sci. (USA)* **82**,5636-5640.
- Bujarski, J.J., Hardy, S.F., Miller, W.A. and Hall, T.C. (1982) Use of dodecyl- $\beta$ -D-maltoside in the purification and stabilization of RNA polymerase from brome mosaic virus-infected barley. *Virology* **119**,465-473.
- Carmichael, G.G. (1979) Isolation of *Escherichia coli* 30S ribosomal protein S1 by poly(rC)-Cellulose chromatography. *Methods Enzymol.* **50**,456-459.
- Candresse, T., Mouches, C. and Bove, J.M. (1986) Characterization of the virus encoded subunit of turnip yellow mosaic virus RNA replicase. *Virology* **152**,322-330.



- Casino, A., Cipollaro, M., Guerrini, A.M., Mastrocinque, G., Spena, A. and Scarlato, V. (1981) Coding capacity of complementary DNA strands. *Nuc. Acids Res.* **9**,1499-1518.
- Chiffot, S., Sommer, P., Hartmann, D., Stussi-Garaud, C. and Hirth, L. (1980) Replication of alfalfa mosaic virus RNA: Evidence for a soluble replicase in healthy and infected tobacco leaves. *Virology* **100**,91-100.
- Clark, G.L., Peden, K.W.C. and Symons, R.H. (1974) Cucumber mosaic virus-induced RNA polymerase: Partial purification and properties of the template-free enzyme. *Virology* **62**,434-443.
- Clerx, C.M. and Bol, J.F. (1978) Properties of solubilized RNA-dependent RNA polymerase from alfalfa mosaic virus-infected and healthy tobacco plants. *Virology* **91**,453-463.
- Collmer, C.W. and Kaper, J.M. (1986) Infectious RNA transcripts from cloned cDNAs of cucumber mosaic virus satellites. *Biochem. Biophys. Res. Comm.* **135**,290-296.
- Collmer, C.W., Tousignant, M.E. and Kaper, J.M. (1983) Cucumber mosaic virus-associated RNA 5. X. The complete nucleotide sequence of a CARNA 5 incapable of inducing tomato necrosis. *Virology* **127**,230-234.
- Cornelissen, B.J.C., Brederode, F.Th., Moormann, R.J.M. and Bol, J.F. (1983a) Complete nucleotide sequence of alfalfa mosaic virus RNA 1. *Nuc. Acids Res.* **11**,1253-1265.
- Cornelissen, B.J.C., Brederode, F.Th., Veeneman, G.H., van Boom, J.H. and Bol, J.F. (1983b) Complete nucleotide sequence of alfalfa mosaic virus RNA 1. *Nuc. Acids Res.* **11**,3019-3025.
- Cornelissen, B.J.C., Janssen, H., Zuidema, D. and Bol, J.F. (1984) Complete nucleotide sequence of tobacco streak virus RNA 3. *Nuc. Acids Res.* **12**,2427-2437.
- Czarnecki, J., Geahlen, R. and Haley, B. (1979) Synthesis and use of azido photoaffinity analogs of adenine and guanine nucleotides. *Methods Enzymol.* **56**,642-653.
- Dasgupta, R. and Kaesberg, P. (1982) Complete nucleotide sequence of the coat protein messenger RNAs of brome mosaic virus and cowpea chlorotic mottle virus. *Nuc. Acids Res.* **10**,703-713.
- Dasmahaptra, B., Dasgupta, R., Saunders, K., Selling, B., Gallagher, T. and Kaesberg, P. (1986) Infectious RNA derived by transcription from cloned cDNA copies of the genomic RNA of an insect virus. *Proc. Natl. Acad. Sci. (USA)* **83**,63-66.
- Davies, C. and Symons, R.H. (1988) Further implications for the evolutionary relationships between plant viruses based on cucumber mosaic virus RNA 3., *Virology*, in press.
- Davies, J.W. and Hull, R. (1982) Genome expression of plant positive-strand RNA viruses. *J. Gen. Virol.* **61**,1-14.
- Dawson, W.O., Beck, D.L., Knorr, D.A. and Grantham, G.L. (1986) cDNA cloning of the complete genome of tobacco mosaic virus and production of infectious transcripts. *Proc. Natl. Acad. Sci.(USA)* **83**,1832-1836.

- Dawson, W.O. and Boyd, C. (1987) TMV protein synthesis is not translationally regulated by heat shock. *Plant Mol. Biol.* **8**,145-149.
- Domier, L.L., Franklin, K.M., Shahabuddin, M., Hellmann, G.M., Overmeyer, H., Hiremath, S.T., Sian, M.F.E., Lomonosoff, G.P., Shaw, J.G. and Rhoads, R.E. (1986) The nucleotide sequence of tobacco vein mottling virus RNA. *Nuc Acids Res.* **14**,5417-5430
- Domier, L.L., Shaw, J.G. and Rhoads, R.E. (1987) Potyviral proteins share amino acid sequence homology with picorna-, como- and caulimoviral proteins. *Virology* **158**,20-27.
- Doolittle, R.F. (1981) Similar amino acid sequences: Chance or common ancestry? *Science* **214**,149-159.
- Dorssers, L., van der Meer, J., van Kammen, A. and Zabel, P. (1983) The cowpea mosaic virus RNA replication complex and the host-encoded RNA-dependent RNA polymerase-template complex are functionally different. *Virology* **125**,155-174.
- Dougherty, W.G. and Hiebert, E. (1985) Genome structure and gene expression of plant RNA viruses., in *"Molecular Plant Virology"* Vol. 2 (Davies, J.W., ed.) pp23-81, CRC Press, Florida.
- Dreher, T.W., Bujarski, J.J. and Hall, T.C. (1984) Mutant viral RNAs synthesized *in vitro* show altered aminoacylation and replicase template activities. *Nature (UK)* **311**,171-175.
- Droms, K.A., Haley, B.E. and Malkinson, A.M. (1987) Decreased incorporation of the photoaffinity probe 8N<sub>3</sub>-[ $\gamma$ -<sup>32</sup>P]-GTP into a 45 kD protein in lung tumours. *Biochem. Biophys. Res. Comm.* **144**,591-597.
- Efstratiadis, A., Vournakis, J.V., Donis-Keller, H., Chaconis, G., Dougall, D.K. and Kafatos, F.C. (1977) End labelling of enzymatically decapped mRNA. *Nuc. Acids Res.* **4**,4165-4174.
- Etchison, D. and Ehrenfeld, E. (1981) Comparison of replication complexes synthesizing poliovirus RNA. *Virology* **111**,33-46.
- Finch, J.T., Klug, A. and van Regenmortel, M.H.V. (1967) The structure of cucumber mosaic virus. *J. Mol. Biol.* **24**,303-305.
- Forster, A.C., McInnes, J.L., Skingle, D.C. and Symons, R.H. (1985) Non-radioactive hybridization probes prepared by the chemical labelling of DNA and RNA with a novel reagent, photobiotin. *Nuc. Acids Res.* **13**,745-761.
- Fraenkel-Conrat, H. (1986) RNA-directed RNA polymerases of plants. *CRC Critical Reviews in Plant Sciences* **4**,213-226.
- Francki, R.I.B. (1981) Plant Virus Taxonomy., in *"Handbook of plant virus infections: comparative diagnosis"* (Kurstak, E., ed.) pp 1-16, Elsevier/Horth-Holland, Amsterdam.
- Francki, R.I.B. (1983) Current problems in viral taxonomy., in *"A critical appraisal of viral taxonomy"* (Matthews, R.E.F., ed.) pp63-104, CRC Press, Florida.

- Francki, R.I.B. (1985a) ed. "The Plant Viruses. Vol. 1. Polyhedral virions with tripartite genomes" Plenum Press, New York.
- Francki, R.I.B. (1985b) Plant virus satellites. *Ann. Rev. Microbiol.* **39**,151-174.
- Francki, R.I.B., Milne, R.G., and Hatta, T. (1985) "Atlas of Plant Viruses" CRC Press, Florida.
- Francki, R.I.B., Mossop, D.W. and Hatta, T. (1979) Cucumber mosaic virus. *CMII/AAB Descriptions of plant viruses* No. 213.
- Francki, R.I.B., Randles, J.W., Chambers, T.C. and Wilson, S.B. (1966) Some properties of purified cucumber mosaic virus (Q strain) *Virology* **28**,729-741.
- Franssen, H., Leunissen, J., Goldbach, R., Lomonossoff, G. and Zimmern, D. (1984) Homologous sequences in non-structural proteins from cowpea mosaic virus and picornaviruses. *EMBO J.* **3**,855-861.
- French, R. and Ahlquist, P. (1987) Intercistronic as well as terminal sequences are required for efficient amplification of brome mosaic virus RNA 3. *J. Virol.* **61**,1457-1465.
- Friedman, R.M., Levin, J.G., Grimsley, P.M. and Berezesky, I.K. (1972) Membrane-associated replication complex in arbovirus infection. *J. Virol.* **10**,504-515.
- Furiuchi, Y. La Fiandra, A. and Shatkin, A.J. (1977) 5'-terminal structure and mRNA stability. *Nature (UK)* **266**,235-239.
- Gallie, D.R., Sleat, D.E., Watts, J.W., Turner, P.C. and Wilson, T.M.A. (1987a) The 5' leader sequence of tobacco mosaic virus RNA enhances the expression of foreign gene transcripts *in vitro* and *in vivo*. *Nuc Acids Res.* **15**,3257-3273.
- Gallie, D.R., Sleat, D.E., Watts, J.W., Turner, P.C. and Wilson, T.M.A. (1987b) A comparison of eukaryotic viral 5'-leader sequences as enhancers of mRNA expression *in vivo*. *Nuc Acids Res.* **15**,8693-8711.
- Garcia-Arenal, F., Zaitlin, M. and Palukaitis, P. (1987) Nucleotide sequence analysis of six satellite RNAs of cucumber mosaic virus: Primary sequence and secondary structure do not correlate with differences in pathogenicity. *Virology* **158**,339-347.
- Garnier, M., Candresse, T. and Bove, J.M. (1986) Immunocytochemical localization of TYMV-coded structural and non-structural proteins by the protein A-gold technique. *Virology* **151**,100-109.
- Gehrke, L. Auron, P.E., Quigley, G.J., Rich, A. and Sonenberg, N. (1983) 5'-conformation of capped alfalfa mosaic virus ribonucleic acid 4 may reflect its independence of the cap structure or of cap-binding protein for efficient translation. *Biochemistry* **22**,5157-5164.
- Gekko, K. and Timasheff, S.N. (1981) Thermodynamic and kinetic examination of protein stabilization by glycerol. *Biochemistry* **20**,4677-4686.
- Gill, D.S. (1983) Cucumber mosaic virus-induced particulate RNA replicase. Ph.D. Thesis, University of Adelaide, Adelaide, South Australia.

- Gill, D.S., Kumarasamy, R. and Symons, R.H. (1981) Cucumber mosaic virus-induced RNA replicase: Solubilization and partial purification of the particulate enzyme. *Virology* **113**,1-8.
- Gilliland, J.M. and Symons, R.H. (1968) Properties of a plant virus-induced RNA polymerase in cucumbers infected with cucumber mosaic virus. *Virology* **36**,232-240.
- Godefroy-Colburn, T., Thivent, C. and Pinck, L. (1985) Translational discrimination between the four RNAs of alfalfa mosaic virus. A quantitative evaluation. *Eur. J. Biochem.* **147**,541-548.
- Goelet, P., Lomonossoff, G.P., Butler, P.J.G., Akam, M.E., Gait, M.J. and Karn, J. (1982) Nucleotide sequence of tobacco mosaic virus RNA. *Proc. Natl. Acad. Sci. (USA)* **79**,5818-5822.
- Goldbach, R.W. (1986) Molecular evolution of plant viruses. *Ann. Rev. Phytopath.* **24**,289-310.
- Goldbach, R. and van Kammen, A. (1985) Structure, replication and expression of the bipartite genome of cowpea mosaic virus. in *"Molecular Plant Virology"* Vol 2 (Davies, J.W., ed.) pp 83-120, CRC Press, Florida.
- Gonda, T.J. (1979) Studies on cucumber mosaic virus and its genome: Replication in protoplasts and the use of molecular hybridization. Ph.D. Thesis, University of Adelaide, Adelaide, South Australia.
- Gonda, T.J. and Symons, R.H. (1979) Cucumber mosaic virus replication in cowpea protoplasts: Time course of virus, coat protein and RNA synthesis. *J. Gen Virol.* **45**,723-726.
- Gordon, K.H.J., Gill, D.S., and Symons, R.H. (1982) Highly purified cucumber mosaic virus-induced RNA-dependent RNA polymerase does not contain any of the full length translation products of the genomic RNAs. *Virology* **123**,284-295.
- Gordon, K.H.J. and Symons, R.H. (1983) Satellite RNA of cucumber mosaic virus forms a secondary structure with partial 3'-terminal homology to genomic RNAs. *Nuc. Acids Res.* **11**,947-960.
- Gould, A.R. and Symons, R.H. (1977) Determination of the sequence homology between the four RNA species of cucumber mosaic virus by hybridization analysis with complementary DNA. *Nuc. Acids Res.* **4**,3787-3802.
- Gould, A.R. and Symons, R.H. (1982) Cucumber mosaic virus RNA 3. Determination of the nucleotide sequence provides the amino acid sequences of the protein 3A and the viral coat protein. *Eur. J. Biochem.* **126**,217-226.
- Grantham, R., Gautier, C., Gouy, M. Jacobzone, M. and Mercier, R. (1981) Codon catalog usage is a genome strategy modulated for gene expressivity. *Nuc. Acids Res.* **9**,r43-r74.
- Guilley, H., Carrington, J.C., Balazs, E., Jonard, G., Richards, K. and Morris, T.J. (1985) Nucleotide sequence and genome organization of carnation mottle virus RNA. *Nuc. Acids Res.* **13**,6663-6677.

- Gunn, M.R. and Symons, R.H. (1980) Sequence homology at the 3'-termini of the four RNAs of alfalfa mosaic virus. *FEBS Lett.* **109**,145-150.
- Gustafson, G., Hunter, B., Hanau, R., Armour, S.L. and Jackson, A.O. (1987) Nucleotide sequence and genetic organization of barley stripe mosaic virus RNA  $\gamma$ . *Virology* **158**,394-406.
- Habeeb, A.F.S.A. (1972) Reaction of protein sulphhydryl groups with Ellman's reagent. *Methods Enzymol.* **25B**,457-464.
- Habili, N. and Francki, R.I.B. (1974a) Comparative studies on tomato aspermy and cucumber mosaic viruses. I. Physical and chemical properties. *Virology* **57**,392-401.
- Habili, N. and Francki, R.I.B. (1974b) Comparative studies on tomato aspermy and cucumber mosaic viruses. II. Virus stability. *Virology* **60**,29-36.
- Habili, N. and Francki, R.I.B. (1974c) Comparative studies on tomato aspermy and cucumber mosaic viruses. III. Further studies on relationship and construction of a virus from parts of the two viral genomes. *Virology* **61**,443-449.
- Hall, T.C. (1979) Transfer RNA-like structures in viral genomes. *Int. Rev. Cytol.* **60**,1-26.
- Hall, T.C., Miller, W.A. and Bujarski, J.J. (1982) Enzymes involved in the replication of plant viral RNAs., in "Advances in Plant Pathology" (Ingram, D. and Williams, P.H., eds.) vol I, pp170-211, Academic Press, New York.
- Hamilton, W.D.O., Boccara, M., Robinson, D.J. and Baulcombe, D.C. (1987) The nucleotide sequence of tobacco rattle virus RNA 1. *J. Gen. Virol.* **68**,2563-2575.
- Hardy, S.F., German, T.L., Loesch-Fries, L.S. and Hall, T.C. (1979) Highly active template-specific RNA-dependent RNA polymerase from barley leaves infected with brome mosaic virus. *Proc. Natl. Acad. Sci. (USA)* **76**,4956-4960.
- Harrison, S.C. (1984) Multiple modes of subunit association in the structures of simple spherical viruses. *Trends Biochem. Sci.* **9**,345-351.
- Haseloff, J., Goelet, P., Zimmern, D., Ahlquist, P., Dasgupta, R. and Kaesberg, P. (1984) Striking similarities in amino acid sequence among non-structural proteins encoded by RNA viruses that have dissimilar genomic organization. *Proc. Natl. Acad. Sci (USA)* **81**,4358-4362.
- Haseloff, J. and Symons, R.H. (1981) Crysanthemum stunt viroid: primary sequence and secondary structure. *Nuc. Acids Res.* **9**,2741-2752.
- Hidaka, S., Ishikawa, K., Takanami, Y., Kubo, S. and Miura, K. (1984) Complete nucleotide sequence of RNA 5 from cucumber mosaic virus (strain Y). *FEBS Lett.* **174**,38-42.
- Hidaka, S., Tsunasawa, S., Yoon, J-O., Narita, K., Takanami, Y., Kubo, S. and Miura, K-I. (1985) Messenger RNA structure participating in the initiation of synthesis of cucumber mosaic virus coat protein. *J. Biochem.* **97**,161-171.

- Hoopes, B.C. and McClure, W.R. (1981) Studies on the selectivity of DNA precipitation by spermine. *Nuc. Acids Res.* **9**,5493-5504.
- Houwing, C.J. and Jaspars, E.M.J. (1986) Coat protein blocks the *in vitro* transcription of virion RNAs of alfalfa mosaic virus. *FEBS Lett.* **209**,284-288.
- Huisman, M.J., Sarachu, A.N., Alblas, F. and Bol, J.F. (1985) Alfalfa mosaic virus temperature-sensitive mutants. II. Early functions encoded by RNAs 1 and 2. *Virology* **141**,23-29.
- Ikemura, T. (1985) Codon usage and tRNA content in unicellular and multicellular organisms. *Mol. Biol. Evol.* **2**,13-34.
- Jacrot, B., Chauvin, C. and Witz, J. (1977) Comparative neutron small-angle scattering study of small spherical viruses. *Nature (UK)* **266**,417-421.
- Jaspars, E.M.J. (1985) Interaction of alfalfa mosaic virus nucleic acid and protein. *in* "Molecular Plant Virology" Vol. 1 (Davies, J.W., ed.) pp 155-221, CRC Press, Florida.
- Jaspars, E.M.J., Gill, D.S. and Symons, R.H. (1985) Viral RNA synthesis by a particulate fraction from cucumber seedlings infected with cucumber mosaic virus. *Virology* **144**,410-425.
- Johnson, J.E. and Argos, P. (1985) Virus particle stability and structure. *in* "The Plant Viruses. Vol. 1. Polyhedral virions with tripartite genomes" (Francki, R.I.B., ed.) pp 19-56, Plenum Press, New York.
- Joshi, S. and Haenni, A-L. (1984) Plant RNA viruses: strategies of expression and regulation of viral genes. *FEBS Lett.* **177**,163-174.
- Joshi, R.L., Joshi, S., Chapeville, F. and Haenni, A-L. (1983a) tRNA-like structures of plant viral RNAs: conformational requirements for adenylation and aminoacylation. *EMBO J.* **2**,1123-1127.
- Joshi, S., Joshi, R.L., Haenni, A-L. and Chapeville, F. (1983b) tRNA-like structures in genomic RNAs of plant viruses. *Trends Biochem. Sci.* **8**,402-404.
- Joshi, S., Neeleman, L., Pleij, C.W.A., Haenni, A.-L., Chapeville, F., Bosch, L. and van Vloten-Doting, L. (1984) Non-structural alfalfa mosaic virus RNA-coded proteins present in tobacco leaf tissue. *Virology* **139**,231-242.
- Joshi, R.L., Ravel, J.M. and Haenni, A-L. (1986) Interaction of turnip yellow mosaic virus val-RNA with eukaryotic elongation factor EF-1  $\alpha$ . Search for a function. *EMBO J.* **5**,1143-1148.
- Kamer, G. and Argos, P. (1984) Primary structural comparison of RNA-dependent polymerases from plant, animal and bacterial viruses. *Nuc. Acids Res.* **12**,7269-7282.
- Kaper, J.M. (1982) Plant disease regulation by virus-dependent satellitelike replicating RNAs. *in* "Control of virus diseases" (Kurstak, E. and Marusyk, R.G., eds.) pp317-343, Dekker, New York.

- Kaper, J.M. and Geelen, J.L.M.C. (1971) Studies on the stabilizing forces of simple RNA viruses. II. Stability, dissociation and reassembly of cucumber mosaic virus. *J. Mol. Biol.* **56**,277-294.
- Kaper, J.M. and Tousignant, M.E. (1977) Cucumber mosaic virus-associated RNA 5. I. Role of host plant and helper strain in determining amount of associated RNA 5 with virions. *Virology* **80**,186-195.
- Kaper, J.M., Tousignant, M.E., Diaz-Ruiz, J.R. and Tolin, S.A. (1978) Peanut stunt virus-associated RNA 5: Second tripartite virus with an associated satellite-like replicating RNA. *Virology* **88**,166-170.
- Kaper, J.M. and Waterworth, H.E. (1981) Cucumoviruses., in *"Handbook of Plant Virus Infections: Comparative Diagnosis"* (Kurstak, E., ed.) pp 257-332, Elsevier/North-Holland, Amsterdam.
- Kemper, B. (1976) Inactivation of parathyroid hormone mRNA by treatment with periodate and aniline. *Nature (UK)* **262**,321-323.
- Khan, Z.A., Hiriyanna, K.T., Chavez, F., and Fraenkel-Conrat, H. (1986) RNA-directed RNA polymerases from healthy and from virus-infected cucumber. *Proc. Natl. Acad. Sci.(USA)* **83**,2383-2386.
- Kiberstis, P.A., Loesch-Fries, L.S. and Hall, T.C. (1981) Viral protein synthesis in barley protoplasts inoculated with native and fractionated brome mosaic virus RNA. *Virology* **112**,804-808.
- Kohanski, R.A., and Lane, M.D. (1985) Receptor affinity chromatography., in *"Biotin"* (Dakshinamurti, K. and Bhagavan, H.N., eds.) *Ann. N.Y. Acad. Sci.* **447**,373-385.
- Kohl, R.J. and Hall, T.C. (1974) Aminoacylation of RNA from several viruses: amino acid specificity and differential activity of plant, yeast and bacterial synthetases. *J. Gen. Virol.* **25**,257-261.
- Koper-Zwarthoff, E.C. and Bol, J.F. (1980) Nucleotide sequence of the putative recognition site for coat protein in the RNAs of alfalfa mosaic and tobacco streak virus. *Nuc. Acids Res.* **8**,3307-3318.
- Koper-Zwarthoff, E.C., Brederode, F.Th., Walstra, P. and Bol, J.F. (1979) Nucleotide sequence of the 3' non-coding region of alfalfa mosaic virus RNA 4 and its homology with the genomic RNAs. *Nuc. Acids Res.* **7**,1887-1900.
- Kozak, M. (1980a) Influence of mRNA secondary structure on binding and migration of 40S ribosomal subunits. *Cell* **19**,79-90.
- Kozak, M. (1980b) Evaluation of the "scanning model" for initiation of protein synthesis in eucaryotes. *Cell* **22**,7-8.
- Kozak, M. (1984) Compilation and analysis of sequences upstream from the translational start site in eukaryotic mRNAs. *Nuc. Acids Res.* **12**,857-872.

- Kozak, M. (1986a) Point mutations define a sequence flanking the initiator codon that modulates translation by eukaryotic ribosomes. *Cell* **44**,283-292.
- Kozak, M. (1986b) Influences of mRNA secondary structure on initiation by eukaryotic ribosomes. *Proc. Natl. Acad. Sci. (USA)* **83**,2850-2854.
- Kozak, M. (1987a) An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. *Nuc. Acids Res.* **15**,8125-8148.
- Kozak, M. (1987b) At least six nucleotides preceding the AUG initiator codon enhance translation in mammalian cells. *J. Mol. Biol.* **196**,947-950.
- Kumarasamy, R. (1980) Purification and characterisation of the cucumber mosaic virus-induced RNA replicase. Ph.D. Thesis, University of Adelaide, Adelaide, South Australia.
- Kumarasamy, R. and Symons, R.H. (1979a) Extensive purification of the cucumber mosaic virus-induced RNA replicase. *Virology* **96**,622-632.
- Kumarasamy, R. and Symons, R.H. (1979b) The tritium labelling of small amounts of protein for analysis by electrophoresis on sodium dodecyl sulfate-polyacrylamide slab gels. *Anal. Biochem.* **95**,359-363.
- Kurath, G. and Palukaitis, P. (1987) Biological activity of T7 transcripts of a prototype clone and a sequence variant clone of a satellite RNA of cucumber mosaic virus. *Virology* **159**,199-208.
- Lafleche, D., Bove, C., Dupont, G., Mouches, C., Astier, T., Garnier, M. and Bove, J.M. (1972) Site of viral RNA replication in the cells of higher plants: TYMV-RNA synthesis on the chloroplast outer membrane system. Proc. 8th FEBS meeting, Amsterdam, *RNA Viruses/Ribosomes*, **27**,43-71., North Holland, Amsterdam.
- Langereis, K., Mugnier, M-A., Cornelissen, B.J.C., Pinck, L. and Bol, J.F. (1986) Variable repeats and poly(A)-stretches in the leader sequence of alfalfa mosaic virus RNA 3. *Virology* **154**,409-414.
- Lazarus, L.H. and Barzilai, R. (1974) Association of foot-and-mouth disease virus replicase with RNA template and cytoplasmic membranes. *J. Gen. Virol.* **23**,213-218.
- Leblond-Francillard, M., Dreyfus, M. and Rougeon, F. (1987) Isolation of DNA-protein complexes based on streptavidin and biotin interaction. *Eur. J. Biochem.* **166**,351-355.
- LeRoy, C., Stussi-Garaud, C. and Hirth, L. (1977) RNA-dependent RNA polymerases in uninfected and in alfalfa mosaic virus-infected tobacco plants. *Virology* **82**,48-62.
- Lischwe, M.A. and Ochs, D. (1982) A new method for partial peptide mapping using N-chlorosuccinimide/urea and peptide silver staining in sodium sulfate-polyacrylamide gels. *Anal. Biochem.* **127**,453-457.
- Loesch-Fries, L.S. and Hall, T.C. (1982) *In vitro* aminoacylation of brome mosaic and barley stripe mosaic virus RNAs. *Nature (UK)* **298**,771-773.



- Lomonosoff, G.P. and Wilson, T.M.A. (1985) Structure and in vitro assembly of tobacco mosaic virus. in *"Molecular Plant Virology"* (Davies, J.W., ed.) CRC Press, Florida, vol. 1, pp 43-83.
- Lot, H. and Kaper, J.M. (1976a) Physical and chemical differentiation of three strains of cucumber mosaic virus and peanut stunt virus. *Virology* **74**,209-222.
- Lot, H. and Kaper, J.M. (1976b) Further studies on the RNA component distribution among the nucleoproteins of cucumber mosaic virus. *Virology* **74**,223-226.
- Lot, H., Marchoux, G., Marrou, J., Kaper, J.M., West, C.K., van Vloten-Doting, L. and Hull, R. (1974) Evidence for three functional RNA species in several strains of cucumber mosaic virus. *J. Gen. Virol.* **22**,81-93.
- Lubinski, J.M., Ransone, L.J. and Dasgupta, A. (1987) Primer-dependent synthesis of covalently linked dimeric RNA molecules by poliovirus replicase. *J. Virol.* **61**,2997-3003.
- Lutcke, H.A., Chow, K.C., Mickel, F.S., Moss, K.A., Kern, H.F. and Scheele, G.A. (1987) Selection of AUG initiation codons differs in plants and animals. *EMBO J.* **6**,43-48.
- Maruyama, T., Gojobori, T. Aota, S. and Ikemura, T. (1986) Codon usage tabulated from the GenBank genetic sequence data. *Nuc. Acids Res.* **14**,r151-r197.
- Matthews, R.E.F. (1982) Classification and nomenclature of viruses: Fourth report of the International Committee on Taxonomy of Viruses. *Intervirology* **17**,4-199.
- Matthews, R.E.F. (1983) *"A critical appraisal of viral taxonomy"* CRC Press, Florida.
- Maxam, A.M. and Gilbert, W. (1980) Sequencing end-labelled DNA with base-specific chemical cleavages. *Methods Enzymol.* **65**,499-560.
- May, J.T., Gilliland, J.M. and Symons, R.H. (1969) Plant virus-induced RNA polymerase: Properties of the enzyme partly purified from cucumber cotyledons infected with cucumber mosaic virus. *Virology* **39**,54-65.
- May, J.T., Gilliland, J.M. and Symons, R.H. (1970) Properties of a plant virus-induced RNA polymerase in particulate fractions of cucumbers infected with cucumber mosaic virus. *Virology* **41**,653-664.
- May, J.T. and Symons, R.H. (1971) Specificity of the cucumber mosaic virus-induced RNA polymerase for RNA and polynucleotide templates. *Virology* **44**,517-526.
- McLachlan, A.D. (1971) Tests for comparing related amino-acid sequences. Cytochrome c and cytochrome c551. *J. Mol. Biol.* **61**,409-424.
- Merril, C.R., Goldman, D., Sedman, S.A. and Ebert, M.H. (1981) Ultrasensitive stain for proteins in polyacrylamide gels shows regional variation in cerebrospinal fluid proteins. *Science* **211**,1437-1438.

- Meshi, T., Ishikawa, M., Motoyoshi, F., Semba, K. and Okada, Y. (1986) *In vitro* transcription of infectious RNAs from full-length cDNAs of tobacco mosaic virus. *Proc. Natl. Acad. Sci (USA)* **83**,5043-5047.
- Messing, J. (1979) A multipurpose cloning system based on the single-stranded DNA bacteriophage M13. *Recomb. DNA Tech. Bull.* **2**,43-48.
- Messing, J., Crea, R. and Seeburg, P.H. (1981) A system for shotgun DNA sequencing. *Nuc. Acids Res.* **9**,309-321.
- Messing, J. and Vieira, J. (1982) A new pair of M13 vectors for selecting either DNA strand of double digest restriction fragments. *Gene* **19**,269-276.
- Meyer, F., Weber, H. and Weissman, C. (1981) Interactions of Q $\beta$  replicase with Q $\beta$  RNA. *J. Mol. Biol.* **153**,631-660.
- Miller, W.A., Bujarski, J.J., Dreher, T.W. and Hall, T.C. (1986) Minus-strand initiation by brome mosaic virus replicase within the 3' tRNA-like structure of native and modified RNA templates. *J. Mol. Biol.* **187**,537-546.
- Miller, W.A., Dreher, T.W. and Hall, T.C. (1985) Synthesis of brome mosaic virus subgenomic RNA *in vitro* by internal initiation of (-) -sense genomic RNA. *Nature (UK)* **313**,68-70.
- Miller, W.A. and Hall, T.C. (1983) Use of micrococcal nuclease in the purification of highly template dependent RNA-dependent RNA polymerase from brome mosaic virus-infected barley. *Virology* **125**,236-241.
- Mills, D.R. and Kramer, F.R. (1979) Structure-independent nucleotide sequence analysis. *Proc. Natl. Acad. Sci.(USA)* **76**,2232-2235.
- Morrow, C.D., Warren, B. and Lentz, M.R. (1987) Expression of enzymatically active poliovirus RNA-dependent RNA polymerase in *Escherichia coli*. *Proc. Natl. Acad. Sci. (USA)* **84**,6050-6054.
- Mossop, D.W. and Francki, R.I.B. (1979) Comparative studies on two satellite RNAs of cucumber mosaic virus. *Virology* **95**,395-404.
- Mouches, C., Candresse, T. and Bove, J.M. (1984) Turnip yellow mosaic virus RNA-replicase contains host and virus-encoded subunits. *Virology* **134**,78-90.
- Murthy, M.R.N. (1983) Comparison of the nucleotide sequences of cucumber mosaic virus and brome mosaic virus. *J. Mol. Biol.* **168**,469-475.
- Nassuth, A., Alblas, F. and Bol, J.F. (1981) Localization of genetic information involved in the replication of alfalfa mosaic virus. *J. Gen. Virol.* **53**,207-214.
- Nassuth, A. and Bol, J.F. (1983) Altered balance of the synthesis of plus- and minus-strand RNAs induced by RNAs 1 and 2 of alfalfa mosaic virus in the absence of RNA 3. *Virology* **124**,75-85.
- Nelson, S.E. and Loesch-Fries, L.S. (1987) RNA complementary to alfalfa mosaic virus RNA 4 is not translated *in vitro*. *Intervirology* **27**,172-176.

- Nozaki, Y. and Tanford, C. (1971) The solubility of amino acids in aqueous ethanol and dioxane solutions. Establishment of a hydrophobicity scale. *J. Biol. Chem.* **246**,2211-2217.
- Ochoa, J.-L. (1978) Hydrophobic (interaction) chromatography. *Biochimie* **60**,1-15.
- Ohno, T., Aoyagi, M., Yamanashi, Y., Saito, H., Ikawa, S., Meshi, T. and Okada, Y. (1984) Nucleotide sequence of the tobacco mosaic virus (tomato strain) genome and comparison with the common strain genome. *J. Biochem.* **96**,1915-1923.
- Ooshika, I., Watanabe, Y., Meshi, T., Okada, Y., Igano, K., Inouye, K. and Yoshida, N. (1984) Identification of the 30k protein of TMV by immunoprecipitation with antibodies directed against a synthetic peptide. *Virology* **132**,71-78.
- Peden, K.W.C., May, J.T. and Symons, R.H. (1972) A comparison of two plant virus-induced RNA polymerases. *Virology* **47**,498-501.
- Peden, K.W.C. and Symons, R.H. (1973) Cucumber mosaic virus contains a functionally divided genome. *Virology* **53**,487-492.
- Rao, A.L.N. and Francki, R.I.B. (1981) Comparative studies on tomato aspermy and cucumber mosaic viruses. VI. Partial compatibility of genome segments from the two viruses. *Virology* **114**,573-575.
- Ravelonandro, M., Godefroy-Colburn, T. and Pinck, L. (1983) Structure of the 5'-terminal untranslated region of the genomic RNAs from two strains of alfalfa mosaic virus. *Nuc. Acids Res.* **11**,2815-2826.
- Ravelonandro, M., Pinck, M. and Pinck, L. (1984) Complete nucleotide sequence of RNA 3 from alfalfa mosaic virus, strain S. *Biochimie* **66**,395-402.
- Reading, C.L. (1982) Theory and methods for immunization in culture and monoclonal antibody production. *J. Immunol. Meth.* **33**,261-291.
- Reisner, A.H. and Bucholtz, C.A. (1983) Apparent relatedness of the main component of ovine 1.714 satellite DNA to bovine 1.715 satellite DNA. *EMBO J.* **2**,1145-1149.
- Rezaian, M.A. and Symons, R.H. (1986) Anti-sense regions in satellite RNA of cucumber mosaic virus form stable complexes with the viral coat protein gene. *Nuc. Acids Res.* **14**,3229-3239.
- Rezaian, M.A., Williams, R.H.V., Gordon, K.H.J., Gould, A.R. and Symons, R.H. (1984) Nucleotide sequence of cucumber mosaic virus RNA 2 reveals a translation product significantly homologous to corresponding proteins of other viruses. *Eur. J. Biochem.* **143**,277-284.
- Rezaian, M.A., Williams, R.H.V. and Symons, R.H. (1985) Nucleotide sequence of cucumber mosaic virus RNA 1. Presence of a sequence complementary to part of the viral satellite and homologies with other viral RNAs. *Eur. J. Biochem.* **150**,331-339.

- Rice, C.M., Aebersold, R., Teplow, D.B., Pata, J., Bell, J.R., Vorndam, A.V., Trent, D.W., Brandriss, M.W., Schlesinger, J.J. and Strauss, J.H. (1986) Partial N-terminal amino acid sequences of three nonstructural proteins of two alphaviruses. *Virology* **151**,1-9.
- Richards, K.E., Jonard, G., Jacquemond, M. and Lot, H. (1978) Nucleotide sequence of cucumber mosaic virus-associated RNA 5. *Virology* **89**,395-408.
- Rietveld, K., van Poelgeest, R., Pleij, C.W.A., van Boom, J.H. and Bosch, L. (1982) The tRNA-like structure at the 3' terminus of turnip yellow mosaic virus RNA. Differences and similarities with canonical tRNA. *Nuc. Acids Res.* **10**,1929-1946.
- Rietveld, K., Pleij, C.W.A. and Bosch, L. (1983) Three-dimensional models of the tRNA-like 3' termini of some plant viral RNAs. *EMBO J.* **2**,1079-1085.
- Roberts, P.L. and Wood, K.R. (1981) Comparison of protein synthesis in tobacco systemically infected with a severe or a mild strain of cucumber mosaic virus. *Phytopath. Z.* **102**,257-265.
- Romaine, C.P. and Zaitlin, M. (1978) RNA-dependent RNA polymerases in uninfected and tobacco mosaic virus-infected tobacco leaves: Viral-induced stimulation of a host polymerase. *Virology* **86**,241-253.
- Rylatt, D.B. and Parish, C.R. (1982) Protein determination on an automatic spectrophotometer. *Anal. Biochem.* **121**,213-214.
- Saito, T., Hosokawa, D., Meshi, T. and Okada, Y. (1987) Immunocytochemical localization of the 130k and 180k (putative replicase components) of tobacco mosaic virus. *Virology* **160**,477-481.
- Samac, D.A., Nelson, S.E. and Loesch-Fries, L.S. (1983) Virus protein synthesis in alfalfa mosaic virus infected alfalfa protoplasts. *Virology* **131**,455-462.
- Sanger, F., Coulson, A.R., Barrell, B.G., Smith, A.J.H. and Roe, B.A. (1980) Cloning in single-stranded bacteriophage as an aid to rapid DNA sequencing. *J. Mol. Biol.* **143**,161-178.
- Sarachu, A.N., Huisman, M.J., van Vloten-Doting, L. and Bol, J.F. (1985) Alfalfa mosaic virus temperature-sensitive mutants. I. Mutants defective in viral RNA and protein synthesis. *Virology* **141**,14-22.
- Savithri, H.S. and Murthy, M.R.N. (1983) Evolutionary relationship of alfalfa mosaic virus with cucumber mosaic virus and brome mosaic virus. *J. Biosci.* **5**,183-187.
- Schofield, P. and Williams, K.R. (1977) Purification and some properties of *Escherichia coli* tRNA nucleotidyltransferase. *J. Biol. Chem.* **252**,5584-5588.
- Schwinghamer, M.W. and Symons, R.H. (1975) Fractionation of cucumber mosaic virus RNA and its translation in a wheat embryo cell-free system. *Virology* **63**,252-262.
- Schwinghamer, M.W. and Symons, R.H. (1977) Translation of the four major RNA species of cucumber mosaic virus in plant and animal cell-free systems and in toad oocytes. *Virology* **79**,88-108.

- Scopes, R.K. (1982) Protein purification. Principles and practice. Springer-Verlag, New York.
- Siegel, A., Hari, V. and Kolacz, K. (1978) The effect of tobacco mosaic virus infection on host and virus-specific protein synthesis in protoplasts. *Virology* **85**,494-503.
- Silberklang, M., Gillum, A.M. and RajBhandary, U.L. Use of <sup>32</sup>P labelling in the sequence analysis of non-radioactive tRNAs. *Methods Enzymol.* **59**,58-109.
- Staden, R. (1982a) Automation of the computer handling of gel reading data produced by the shotgun method of DNA sequencing. *Nuc. Acids Res.* **10**,4731-4751.
- Staden, R. (1982b) An interactive graphics program for comparing and aligning nucleic acid or amino acid sequences. *Nuc Acids Res.* **10**,2951-2961.
- Staden R. (1984) Computer methods to locate signals in nucleic acid sequences. *Nuc. Acids Res.* **12**,505-519.
- Stussi-Garaud, C., Garaud, J.-C., Berna, A. and Godefroy-Colburn, T. (1987) *In situ* localization of an alfalfa mosaic virus non-structural protein in plant cell walls: correlation with virus transport. *J. Gen. Virol.* **68**,1779-1784.
- Strauss, E.G. and Strauss, J.H. (1983) Replication strategies of the single-stranded RNA viruses of eukaryotes. *Curr. Top. Microbiol. Immunol.* **105**,1-98.
- Symons, R.H. (1975) Cucumber mosaic virus contains 7-methyl guanosine at the 5'-terminus of all four RNA species. *Mol. Biol. Rep.* **2**,277-285.
- Symons, R.H. (1978) The two-step purification of ribosomal RNA and plant viral RNA by polyacrylamide slab gel electrophoresis. *Aust. J. Biol. Sci.* **31**,25-37.
- Symons, R.H. (1979) Extensive sequence homology at the 3'-termini of the four RNAs of cucumber mosaic virus. *Nuc. Acids Res.* **7**,825-837.
- Symons, R.H. (1985) Viral genome structure., in "The Plant Viruses. Vol. I. Polyhedral virions with tripartite genomes" (Francki, R.I.B., ed.) pp 57-81, Plenum Press, New York,
- Symons, R.H., Gill, D.S., Gordon, K.H.J. and Gould, A.R. (1982) Gene content and expression of the four RNAs of cucumber mosaic virus., in "Manipulation and Expression of Genes in Eukaryotes" (Nagley, P., Linnane, A.W., Peacock, W.J. and Pateman, J.A., eds.) pp 373-380, Academic Press, Sydney.
- Takanami, Y. and Fraenkel-Conrat, H. (1982) Comparative studies on ribonucleic acid dependent RNA polymerases in cucumber mosaic virus infected cucumber and tobacco and uninfected tobacco plants. *Biochemistry* **21**,3161-3167.
- Taniguchi, T., Palmieri, M. and Weissman, C. (1978) Q $\beta$  DNA-containing hybrid plasmids giving rise to Q $\beta$  phage formation in the bacterial host. *Nature (UK)* **274**,223-228.
- Tomenius, K., Clapham, D. and Meshi, T. (1987) Localization by immunogold cytochemistry of the virus-coded 30k protein in plasmodesmata of leaves infected with tobacco mosaic. *Virology* **160**,363-371.

- Traub, A., Diskin, B., Rosenberg, H. and Kalmar, E. (1976) Isolation and properties of the replicase of encephalomyocarditis virus. *J. Virol.* **18**,375-382.
- van Belkum, A., Bingkun, J., Rietveld, K., Pleij, C.W.A. and Bosch, L. (1987) Structural similarities among valine-accepting tRNA-like structures in tymoviral RNAs and elongator tRNAs. *Biochemistry* **26**,1144-1151.
- van der Meer, J., Dorssers, L. and Zabel, P. (1983) Antibody-linked polymerase assay on protein blots: a novel method for identifying polymerases following SDS-polyacrylamide gel electrophoresis. *EMBO J.* **2**,233-237.
- van der Werf, S., Bradley, J., Wimmer, E., Studier, F.W. and Dunn, J.T. (1986) Synthesis of infectious poliovirus RNA by purified T7 RNA polymerase. *Proc. Natl. Acad. Sci. (USA)* **83**,2330-2334.
- van Vloten-Doting, L. (1985) Virus Genetics. in *"The Plant Viruses. Vol. I. Polyhedral virions with tripartite genomes"* (Francki, R.I.B., ed.) pp 117-161, Plenum Press, New York.
- van Vloten-Doting, L., Dubelaar, M. and Bol, J.F. (1983) Open reading frame in the minus strand of two plus type RNA viruses. *Plant Mol. Biol.* **1**,155-158.
- van Vloten-Doting, L., Francki, R.I.B., Fulton, R.W., Kaper, J.M. and Lane, L.C. (1981) Tricornaviridae - A proposed family of plant viruses with tripartite, single-stranded RNA genomes. *Intervirology* **15**,198-203.
- Watanabe, Y. and Okada, Y. (1986) *In vitro* viral RNA synthesis by a subcellular fraction of TMV-inoculated tobacco protoplasts. *Virology* **149**,64-73.
- Weening, C.J. and Bol, J.F. (1975) Viral RNA replication in extracts of alfalfa mosaic virus-infected *Vicia Faba*. *Virology* **63**,77-83.
- Wilson, P.A. and Symons, R.H. (1981) The RNAs of cucumoviruses: 3'-terminal sequence analysis of two strains of tomato aspermy virus. *Virology* **112**,342-345.
- Winter, G. and Fields, S. (1980) Cloning of influenza cDNA into M13: The sequence of the RNA fragment encoding the A/PR/8/38 matrix protein. *Nuc. Acids Res.* **8**,1965-1974.
- Young, N.D. and Zaitlin, M. (1986) An analysis of tobacco mosaic virus replicative structures synthesized *in vitro*. *Plant Mol. Biol.* **6**,455-465.
- Zaitlin, M., Duda, C.T. and Petti, M.A. (1973) Replication of tobacco mosaic virus. V. Properties of the bound and solubilized replicase. *Virology* **53**,300-311.
- Zimmern, D. (1977) The nucleotide sequence of the origin of assembly on tobacco mosaic virus RNA. *Cell* **11**,463-482.
- Zimmern, D. (1983) An extended secondary structural model for the TMV assembly origin and its correlation with protection studies and an assembly defective mutant. *EMBO J.* **2**,1901-1907.

- Zimmern, D. and Kaesberg, P. (1978) 3'-Terminal nucleotide sequence of encephalomyocarditis virus RNA determined by reverse transcriptase and chain-terminating inhibitors. *Proc. Natl. Acad. Sci. (USA)* **75**,4257-4261.
- Zingde, S.M., Shirsat, N.V. and Gothoskar, B.P. (1986) Peptide mapping of proteins in gel bands after partial cleavage with acidic cyanogen bromide vapors. *Anal. Biochem.* **155**,10-13.