This is to certify that the

thesis entitled

Cloning and Sequencing of Peach Rosette Mosaic Virus RNA1

presented by

Allan Henry Lammers

has been accepted towards fulfillment
of the requirements for

Master of Science degree in Botany and Plant Pathology

Richard F. Allison
Donald G. Ramsdell
Major professor

Date May 15, 1996

CLONING AND SEQUENCING OF PEACH ROSETTE MOSAIC VIRUS RNA1

By

Allan Henry Lammers

A THESIS

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

MASTER OF SCIENCE

Department of Botany and Plant Pathology

1996

ABSTRACT

CLONING AND SEQUENCING OF PEACH ROSETTE MOSAIC VIRUS RNA 1

By

Allan Henry Lammers

The complete nucleotide sequence of peach rosette mosaic nepovirus (PRMV) RNA1 has been determined. A Michigan grapevine isolate of PRMV was propagated, purified and cDNA clones representing 99.6% of the RNA1 were constructed. cDNA and direct RNA sequence analyses revealed a RNA species of 7977 nucleotides, excluding a 3' polyadenylated tail. The 5'- and 3'- untranslated regions are 52 and 1474 nucleotides, respectively. Computer analysis of PRMV RNA1 nucleotide sequence unveiled a single open reading frame of 6450 nucleotides encoding a 240 kD polyprotein. Analysis of predicted amino acid sequence of RNA1 uncovered amino acid motifs characteristic of a replicase, a proteinase, an NTP-binding protein and a proteinase cofactor. Order and identity of these putative proteins are consistent with other nepoviruses. This analysis of PRMV RNA1 further distinguishes the taxonomic subdivisions within the nepovirus group, confirms subgroup II status of PRMV and lays the groundwork for a pathogen-derived resistance strategy.

ACKNOWLEDGEMENTS

I would like to gratefully acknowledge my major professors Dr. Don Ramsdell and Dr. Richard Allison as well as my graduate committee member, Dr. Gus de Zoeten. I thank the members of their labs, including Bill Schneider, Ann Greene, Fang Gouwei, Dr. Jeanne Ohrnberger, Min Deng, Deb Rucker, Dr. Jihad Skaf, Dr. Steve Demler, Dr. Pat Traynor and Christiane Wobus for their friendship and scientific assistance. Thanks particularly to Jerri Gillett for her kind and helpful assistance, especially with virus purification. Thanks to Dr. John Halloin for the use of his photographic equipment and to our department office staff, especially Jacqueline Guyton, for their help. I am indebted to my family for their love and support, especially my sister Audrey for her skilled technical assistance. Lastly, I thank my wife and best friend Amy, without whom this work would not have been possible.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

LITERATURE REVIEW

Nepoviruses are considered a genus within the picornavirus-like supergroup of plant viruses which includes the potyviruses, comoviruses, and picornaviruses (Goldbach, *et al.*, 1987). Common features within this supergroup include genomic structure and organization, as well as nucleotide and amino acid sequence similarity. Nepoviruses have many unique features not found among other picornavirus-like members. Cadman (1963) recognized some of these unique aspects; he observed that the nepoviruses formed a natural group based on particle morphology and their ability to be transmitted by nematodes. In fact, his acronym nepovirus for a nematode-transmitted polyhedral virus was one of the first names that the International Committee on Taxonomy of Viruses (ICTV) approved for a group of plant viruses (Wildy, 1971). The number of definitive and possible nepoviruses has rapidly increased from eight in 1971 (Harrison *et al.*, 1971) to 26 in 1982 (Matthews, 1982) to the current 36 species (Goldbach *et. al*, 1995). The criteria of having a confirmed nematode vector is fulfilled by only 11 of the 36 nepoviruses. The remainder owe their present taxonomic assignment to possession of other nepovirus characteristics; for example, host range response (infecting vegetables, small fruits, or fruit trees), and physical and serological behavior (Martelli and Taylor, 1990)

Nepovirus taxonomy is unresolved. In the absence of genomic sequence data, taxonomic criteria have emphasized physical and serological characteristics. Most nepoviruses consist of three distinct particle types: a top (T) component consisting of empty polyhedral capsid proteins; and a middle component (M) and a bottom (B) component, each containing the identical capsid proteins plus single molecules of RNA2 and RNA1, respectively (Martelli and Taylor, 1990). Researchers have attempted to

subdivide nepoviruses using physical and serological criteria. Martelli (1975) proposed a four-part subdivision based on physical characteristics. Martelli's scheme depends on the sedimentation coefficients of RNA2 molecules, however, published particle sedimentation values and RNA molecular weights are prone to error (Francki et al., 1985); Francki *et al.* (1985) argued that the nepovirus group should be separated into two subgroups based on the distinct morphologies of RNA2. Subgroup I would consist of nepoviruses with RNA2 components smaller than 5.4 kb, while subgroup II members would have RNA2 components greater than 5.4 kb. Since nepoviruses are serologically unrelated, subgrouping them on this basis alone is insufficient (Francki, *et al.*, 1985).

Until recently, physical and serological data were the only criteria used to separate the nepoviruses. The nucleotide sequence of many nepoviruses is now available (see Appendix A). Subgroup I nepoviruses (RNA2 smaller than 5.4 kb) have been well characterized. In fact, complete nucleotide sequences are known for arabis mosaic virus (ArMV); grapevine chrome mosaic virus (GCMV); grapevine fanleaf virus (GFLV); olive latent ringspot virus (OLRSV); raspberry ringspot virus (RRSV); and tomato black ring virus (TBRV) (Loudes *et al.*, 1995; LeGall *et al.*, 1989 and Brault *et al.*, 1989; Serghini *et al.*, 1990 and Ritzenthaler *et al.*, 1991; Grieco *et al.*, 1995; Blok *et al.*, 1992; Greif *et al.*, 1988 and Meyer *et al.*, 1986, respectively). Partial nucleotide sequence is available for tobacco ringspot virus (TRSV) RNA2 (Buckley *et al.*, 1993). Until now, the only completely sequenced subgroup II nepovirus is tomato ringspot virus (TomRSV). Partial sequence analysis is available for blueberry leaf mottle nepovirus (BBLMV) RNA1 and RNA2 (Bacher *et al.*, 1994b). Further genomic analysis of subgroup II members is needed to confirm subdivision of the nepoviruses.

Based on physical and serological characteristics, PRMV is considered a subgroup II nepovirus (Ramsdell and Myers, 1974; Harrison and Murant, 1977; Ramsdell and Myers, 1978; Dias and Cation, 1980). Early molecular characterization of PRMV

focused on the physical properties of the virus. PRMV is unrelated to any nepovirus serologically, yet it shares many physical similarities with other members of the group. It consists of 28 nm isometric particles composed of 60 copies of a single capsid protein. A bipartite, single-stranded, positive-sense RNA genome is separately encapsidated and both nucleoprotein components are required for infection (Harrison and Murant, 1977). Mature PRMV proteins are released from two large polyprotein precursors corresponding to the translation products of RNA1 and RNA2, as demonstrated for the nepoviruses TomRSV, TBRV, GFLV, GCMV, as well as other picorna-like viruses including cowpea mosaic comovirus (Rott *et al.,* 1991, 1995; Demangeat *et al.,* 1990; Ritzenthaler *et al.,* 1991; Le Gall *et al.,* 1989; Lomonosoff and Shanks, 1983). Before discussing the nucleotide sequence analysis of PRMV it is necessary to understand the economic importance as well as some of the epidemiological aspects of the disease caused by PRMV.

The Disease

PRMV was first recognized as the cause of a disease of peaches (*Prunus persica* L.) in Michigan in 1917 (Klos *et al.*, 1976). Since then it has been reported to infect highbush blueberry (*Vaccinium corymbosum* L.) and many cultivars of grapevine (*Vitis labrusca* L.) (Ramsdell and Gillett, 1981; Ramsdell and Myers, 1974). PRMV infects a variety of weed species in Michigan; it was detected in 3 of 16 genera of weed species adjacent to diseased 'Concord' vines (Ramsdell and Gillett, 1981). Weeds infected included curly dock (*Rumex crispus* L.), Carolina horsenettle (*Solanum carolinense* L.), and common dandelion (*Taraxacum officinale* Weber). PRMV is seed-borne in dandelion at a low level (3.6%) (Ramsdell and Myers, 1978). Peaches, highbush blueberry, and grapevines grown where PRMV is endemic are susceptible to this disease.

PRMV has been reported most often in Michigan, occasionally in southwestern Ontario (Canada) and once in New York (Ramsdell and Myers, 1974; Stace-Smith and Ramsdell, 1987). Peach trees infected with PRMV exhibit delayed foliation, chlorotic mottling and distortion of the early formed leaves, and shortening of the internodes resulting in an overall rosette appearance to the plant. Typical symptoms in peach also include chlorosis of the leaves. Chlorotic areas are variable in color intensity and morphology. Twenty percent of Michigan's approximately 18,000 acres of highbush blueberries are produced where PRMV is endemic. Symptoms in blueberry appear as elongated, crescent-shaped mature leaves and spoon-shaped terminal leaves (Ramsdell and Gillett, 1981). Common symptoms of PRMV infection in 'Concord' grapevine include leaf deformation, extreme shortening of internodes, whorling of leaves, a typical umbrella-like growth habit of the vine, and sometimes death of the plant (Ramsdell and Myers, 1978). Infection in grapevine also results in delayed dormancy breaking and uneven bloom, small and uneven berry clusters, and a yield 50-fold lower than uninfected 'Concord' vines. Currently, PRMV infection has become a serious problem in over 100 Michigan vineyards (Ramsdell, *unpublished*). PRMV infection has just recently been detected in SW Ontario grapevines (Stobbs and Van Schagen, 1996).

PRMV Transmission

The predominant mode of natural transmission of PRMV is via nematodes. PRMV inoculum is spread by nematodes between vines and from weed hosts to adjacent vines (Ramsdell and Myers, 1978). Two nematode species, a dagger nematode, *Xiphinema americanum* (Cobb), and the needle nematode, *Longidorus diadecturus* (Eveleigh and Allen), have been reported as PRMV vectors (Klos *et al.*, 1967; Eveleigh and Allen, 1982). Another dagger nematode, *X. revisi*, is a suspected vector of PRMV in SW Ontario (Stobbs and Van Schagen, 1996). An Ontario population of *L. elongatus* (DeMan) transmitted PRMV at a low level (1 plant infected per 46 plants tested) but

investigators attributed this to non-specific retention of the virus (Allen and Ebsary, 1988). Occasional, non-specific transmission of nepoviruses by *Longidorus* spp. occurs (Allen, 1986) when unadsorbed ingested particles contaminate the stylet and are subsequently released into the transmission plant. The potential of *Criconemoides* sp. as a vector of PRMV has yet to be confirmed (Stace-Smith, R. and Ramsdell, D.C., 1987). Electron microscopy of thin sections of nematode vectors has identified virus retention sites within each of the vector genera. In *Longidorus* sp., viral particles of RRSV (raspberry ringspot virus) and TBRV adsorbed to the inner surface of the odontostyle. In *Xiphinema* sp., the particles of ArMV (arabis mosaic virus), SLRSV (strawberry latent ringspot virus) are associated with the cuticular lining of the odontophore, the esophagus, and the esophageal pump (Martelli and Taylor, 1990). The virus retention period within the Longidorid vector is approximately 3 weeks while PRMV can be retained for up to 11 months in *X. americanum.* The lengthy virus retention time in *X. americanum* provides an excellent over-wintering strategy but complicates efforts to control this disease. The dagger nematode appears to be the more important nematode vector of PRMV in Michigan while the Longidorid vector is more important in Ontario (Allen, 1986; Stace-Smith and Ramsdell, 1987). The preference of one vector species over another in these two locations is likely related to predominance of local nematode species rather than any physical differences between vector or viral populations.

Disease Control: 1. Control of Vector Populations

The nematode vectors of PRMV have been found beneath infected grape roots to a depth of 2.13 meters (Bird and Ramsdell, 1985). The depths at which these phytoparasitic nematodes persist present a challenge for chemical control of the vector. Long-term (10 yr.) fallowing of soil fails to prevent GFLV reinfection of grapevines by *X. index* (Raski, D.J. *et al.* 1965). Chemical treatment of the soil using a combined shallow (20 cm) plus deep (1 m) soil fumigation method provided good control of PRMV

vector populations over an 8-year study period in southwest Michigan (Ramsdell, D.C. and Gillett, J.M., 1983). Virus-free 'Concord' vines were introduced into treated soil in 1983 and, to date, have remained healthy. However, the future of chemical control of nematode-transmitted viral diseases appears ill-fated. The $5000/acre cost of the combined chemical control strategy described above is prohibitive. Additionally, all effective fumigants except for D-D (1,3-dichloropropane/dichloropropene mixture), have been decertified in the U.S. due to environmental concerns (Ramsdell *et al.*, 1995). This has led to a search for host resistance to PRMV as a suitable control strategy.

2. Resistance to PRMV among Grapevine Cultivars

The juice grape 'Concord' comprises 95% of Michigan's 11,000 acres of grapevine. 'Concord' is highly susceptible to PRMV infection, thus, two different studies have attempted to find highly resistant rootstock (Ramsdell and Gillett, 1985; Ramsdell *et al.*, 1995). In the first study, Ramsdell and Gillett (1985) tested the relative susceptibility of 28 cultivars of American, French hybrid, and European grapevine to PRMV infection. Groups of five test vines and a single 'Concord' control vine, were planted beneath a mature, PRMV-infected 'Concord' source vine. Over a 10-year period, leaf extracts from these vines were used to mechanically inoculate the herbaceous systemic host *Chenopodium quinoa* Willd. or tested by enzyme-linked immunosorbent assay (ELISA). Very low infection rates were reported for the American cvs. 'Delaware' and 'Niagara' (0.8% and 1.4%, respectively), while much higher infection rates were reported for cv. 'Concord' (35.4%). Ramsdell concluded that although cv. 'Delaware' exhibited the highest resistance level to PRMV, it's poor, spindly growth made it a less suitable rootstock than the more robust cv. 'Niagara'.

In a later study, Ramsdell *et al.* (1995) compared various scion and rootstock grapevine cultivars by measuring the effect of PRMV infection on the yield and the growth of vines. Cultivars tested included 'Concord' as well as those that did not show

significant PRMV infection in the earlier study. Results indicated that over a 4 yr period, the greatest reduction in yield and growth due to PRMV infection occurred in 'Concord' (42% and 64%, respectively). PRMV was detected in 5% of 'Chancellor' and 'Couderc 1616' vines, 7% of 'Couderc 1202' and 'Foch' vines, 18.2% of 'Niagara' and 'Delaware' vines, 20% of 'Teleki 5C' vines, and more than 50% of the vines of 'Vignoles', 'Teleki 5A', and 'Concord', respectively. The white wine grape cv. 'Seyval' remained uninfected during the study period but unfortunately is not used as a rootstock. Ramsdell *et al.* (1995) concluded that cv. 'C. 1616' would make a suitable rootstock for the valuable, yet susceptible, 'Concord' and 'Niagara' scions. Regrettably, however, very few rootstocks remain which are not PRMV-susceptible (Ramsdell *et al.*, 1985).

Conventional breeding will continue to have an important role in future production of PRMV-resistant grapevine. However, resistance is not always available in a closely associated interfertile relative and/or resistance genes may be tied to undesirable traits. Further, resistance may be multigenic and difficult to transfer (Grumet, 1995). Recent advancements in grapevine tissue transformation and genetic engineering of host resistance provide a plausible alternative for sustainable disease control.

3. Toward Engineering Resistance to PRMV

Molecular approaches have recently been developed for achieving high levels of virus resistance in a variety of crop systems (review Grumet, 1995). This has proven to be a highly successful strategy; up to 100% resistance has been obtained. In reports thus far, the source of genetically engineered virus resistance consists solely of pieces from the viral genome. Transgenic virus resistance is acquired by introducing a part of a plant virus genome into the host genome. Various portions of the viral genome have proven effective, including the capsid protein, movement protein and replicase genes, ribozymes, and anti-sense RNA. The effectiveness of movement protein, antisense and defective-interfering RNA, and ribozymes appear to differ among viruses and will not be discussed

further. Capsid protein (CP) and replicase genes have emerged as the most effective genes for conferring pathogen-derived resistance. Capsid protein-mediated resistance has been demonstrated for many viruses (Beachy *et al.*, 1990; Fitchen and Beachy, 1993 and Grumet, 1990, 1995); replicase-mediated resistance is discussed below. In general, CP-derived protection is limited; the transgenic host is protected from the virus from which the transgene was derived and a few closely related viruses. CP-derived virus resistance can be overcome by inoculation with large quantities of virus or naked RNA. The mechanism of capsid protein-mediated resistance is unknown and may vary among viruses and viral constructs (Grumet, 1995). Capsid protein-mediated resistance is usually ephemeral and probably not useful for woody plants.

Replicase-mediated resistance appears to be a more attractive method of conferring resistance (review Fitchen and Beachy, 1993; Scholthof *et al.*, 1993). Although the spectrum of protection is narrow and similar to that conferred by the CP (limited to resistance against the source virus or its immediate relatives), resistance was not overcome by high quantities of inoculum or naked RNA. The level of resistance observed in replicase-expressing transgenic lines was greater than that for CP-mediated resistance (Grumet, 1995). Golemboski *et al.* (1990) reported that lines expressing a 54 kD replicase protein of tobacco mosaic tobamovirus (TMV) were 100% resistant to TMV infection. Perhaps more importantly, resistance was retained even at inoculum quantities up to 1000-fold higher than afforded by the TMV CP gene. Although the mechanism of replicase-mediated resistance is likewise unknown, its efficiency makes it the preferred system for pathogen-derived resistance.

Replicase-mediated resistance appears to be an attractive strategy for grapevine. Successful implementation of this strategy demands that a grapevine transformation system be available and useful genes be characterized. Various investigators have successfully transformed grapevine tissue using *Agrobacterium*-mediated or biolistic transformation methods (LeGall *et al.*, 1994; Mauro *et al.*, 1995; Krastanova *et al.*, 1995;

Lupo *et al.*, 1994; Nakano *et al.*, 1994; Martinelli and Mandolino, 1994; Bardonnet *et al.*, 1994; and Kikkert *et al.*, 1996). A transgene is placed under the direction of a constitutive promoter, often the 35S transcription promoter of cauliflower mosaic virus. The virus-derived nucleotide sequence is commonly nested within a cassette, which also contains one or more selectable marker genes that enables detection of transformed tissue. For example, if the neomycin-phosphotransferase (NPTII) marker gene is used, transformed embryonic tissue is selected for kanamycin resistance. Thus, the grapevine transformation system required for an engineered resistance strategy has been developed.

Protection against PRMV infection in grapevine with a molecular approach such as replicase-mediated resistance strategy also requires the molecular characterization of PRMV. Nepoviruses encode their replicase gene on RNA1 (Sanfaçon, 1995) and for this reason, we chose to sequence PRMV RNA1. Determination of the entire RNA1 sequence allows for isolation of the PRMV replicase gene. Once isolated, the replicase gene may be introduced into the grapevine genome to establish PRMV resistance.

4. Molecular Characterization of PRMV RNA1

We have determined the complete nucleotide sequence of PRMV RNA1. A grapevine isolate of PRMV from southwest Michigan was propagated and purified and cDNA clones representing 99.6% of RNA1 were obtained. cDNA sequence and direct RNA sequence analyses revealed an RNA species of 7977 nucleotides. The 5'- and 3'-untranslated regions consist of 52 and 1474 nucleotides, respectively. Computer analysis of the PRMV RNA1 nucleotide sequence unveiled a single long open reading frame of 6450 nucleotides capable of encoding a 240 kD polyprotein. Analysis of the predicted amino acid sequence of RNA1 revealed motifs characteristic of a replicase, a proteinase, an NTP-binding protein and a proteinase cofactor. The order and identity of these putative proteins are consistent with other nepoviruses.

The molecular characterization of PRMV RNA1 presented here is an essential step for developing a replicase-mediated resistance strategy. Portions of the RNA1 sequence, in conjunction with classical breeding and selection for resistant cultivars, may provide 'Concord' vineyards with PRMV resistance. Obtaining grapevines expressing a portion of the PRMV RNA1 genome will also further our attempt to understand the mechanism of pathogen-derived resistance.

The PRMV RNA1 sequence also serves an important taxonomic role. PRMV sequence data represents only the second complete sequence of a subgroup II nepovirus RNA. These RNA1 sequence data confirm the subgroup II status of PRMV. Further, analysis of the PRMV RNA1 genome supports Francki's (1985) bipartite subdivision of the nepovirus group along with TomRSV. This is the first attempt to separate the nepovirus group based on genomic sequence information. Sequence homology between PRMV and TomRSV, as well as similarities in genomic strategy, confirm the relationship of these subgroup II nepoviruses. Comparison of the PRMV genome to that of TomRSV and subgroup I nepoviruses highlights many interesting aspects which may add to the criteria used to distinguish the two nepovirus subgroups.

CHAPTER 2

CLONING AND SEQUENCING OF PEACH ROSETTE MOSAIC VIRUS RNA1

INTRODUCTION

The complete nucleotide sequence of peach rosette mosaic nepovirus (PRMV) RNA1 has been determined. PRMV has a bipartite, plus sense RNA genome which contains a 5'-VPg and a 3'-poly(A) tail at the termini. RNA1 is 7977 nucleotides excluding a 3'-polyadenylated tail. The 5'- and 3'- untranslated regions are 52 and 1474 nucleotides, respectively. The nucleotide sequence contains a single long open reading frame of 6450 nucleotides capable of encoding a 240 kD polyprotein. Analysis of the nucleotide sequence of RNA1 revealed motifs characteristic of a replicase, a proteinase, an NTP-binding protein and a proteinase cofactor. The order and identity of these putative proteins are consistent with other nepoviruses, especially subgroup II tomato ringspot nepovirus (TomRSV).

Francki *et al.* (1985) proposed taxonomic subdivision of the nepovirus group into two subgroups based on the size of RNA2. RNA2 is smaller than 5.4 kb in subgroup I and larger than 5.4 kb in subgroup II. Sequence analysis of the primary and secondary structure of PRMV RNA1 reveals several features characteristic of nepoviruses, such as nucleotide and amino acid sequence homology, polyprotein expression and overall genomic organization. Analysis confirms that PRMV belongs to a distinct subgroup of nepoviruses including TomRSV. With the availability of the complete nucleotide sequence of another subgroup II nepovirus in TomRSV, comparisons were made between individual RNA species of this subgroup. Comparison of the length of the 3'-untranslated region, putative polyprotein processing activity, and the level of sequence homology

between nepovirus RNA species strengthens our ability to distinguish between nepovirus subgroups. PRMV, like TomRSV and partially sequenced subgroup II nepoviruses BBLMV and CLRV, has a 3'-UTR approximately1.5 kb in length while subgroup I 3'-UTR is less than 0.5 kb. Polyprotein processing of the PRMV polyprotein appears to closely resemble that of TomRSV and picorna-like viruses such as CPMV, potyviruses and polio, distinct from subgroup I nepovirus. Overall amino acid sequence identity between PRMV and TomRSV further confirms PRMV subgroup II status and supports the subdivision of nepoviruses by Francki *et al.* (1985).

MATERIALS AND METHODS

Propagation and Purification of PRMV

Canes of PRMV-infected grapevine (*V. labrusca* cv. 'Concord') were harvested from a vineyard located at Michigan State University, E. Lansing, MI, in December, 1992. The tissue was ground by mortar and pestle in a 4°C solution of 0.01 M sodium phosphate buffer, pH 7.5 and rub-inoculated on primary leaves of the herbaceous host *Chenopodium quinoa* (Willd.) seedlings dusted with 600 mesh carborundum (Fisher Scientific, Pittsburgh, PA).

The following purification method for PRMV was adapted from Dias and Allen (1980). Fifty to 100 g of symptom-bearing *C. quinoa* leaves were harvested 10 to 14 d.p.i. All subsequent steps were performed at 4°C. The tissue was blended in a commercial Waring blender for 2 min. in 0.5 M boric acid buffer (12.5 mM sodium borate, 10 hydrate; 0.5 M boric acid; 0.5% (w/v) ascorbic acid; adjusted to pH 6.5 with 1.0 N sodium hydroxide). Approximately 2 ml of boric acid buffer were used per gram of tissue. Homogenized tissue was filtered through four-ply cheesecloth into a 500 ml beaker and chloroform was slowly added to the extract to a final volume of 8.5% and stirred for 3 min. The solution was centrifuged for 15 min. at 12,000 x g (J2-21 Centrifuge, Beckman Instruments, Inc., Palo Alto, CA) in a No. 30 rotor (Beckman) and

the pH of the supernatant was adjusted to 5.3 with 1N HCl. A 30 min. incubation on ice was followed by 15 min. (12,000 x g) centrifugation. Supernatant was collected in Beckman Quick-Seal centrifuge tubes (25 mm x 89 mm) and ultra-centrifuged (model L7-65, Beckman) at 105,000 x g for 4 hr in a Ti50.2 rotor (Beckman). Pellets were resuspended overnight in 1.8 ml 0.01 M potassium phosphate buffer, pH 7.0.

Linear-logarithmic 0-30% sucrose gradients were prepared and equilibrated at 4°C overnight. Resuspended virus solution was layered onto sucrose gradients, 0.3 ml per gradient, and centrifuged for 90 min. at 38K RPM (105,000 x g) in a swinging bucket rotor (Beckman SW 41). A density gradient fractionator (model 185, Instrument Specialties Co. (ISCO), Lincoln, NB) and absorbance monitor (model UA-5, ISCO) were used to separate the components of the sucrose column. The absorbance monitor UV lamp was adjusted to 254 nm wavelength for detection of the virus particles. Other absorbancy monitor settings included a chart speed of 60 cm/hr and a sensitivity of 2.0; the baseline was adjusted prior to fractionation. Fifty percent sucrose solution was used to push the gradient upward at a flow rate of 1.5 ml/min. Fractions corresponding to RNA1- and RNA2-containing absorbance peaks were collected in approximately 0.5 ml volumes. Pooled RNA1 and RNA2 fractions were diluted in 3 volumes of 0.01 M potassium phosphate buffer, pH 7.0, and centrifuged for 5 hr at 38K RPM (105,000 x g) in a Beckman 40 rotor. The pelleted virus was resuspended in 100 µl TE buffer (10 mM Tris-HCl, 1 mM ethyldiethanolamine (EDTA), pH 7.0) and transferred to sterile 1.5 ml Eppendorf tubes.

Viral RNA was extracted by adding 100 µl of Tris-saturated phenol (1:1 v/v) to the resuspended virus solution. The mixture was vortexed for 30 sec, and centrifuged in a bench top centrifuge (Eppendorf 5415C) at 14K RPM for 1 min. The RNA-containing aqueous layer was transferred to a new 1.5 ml Eppendorf tube and phenol extraction was repeated twice. Phenol extracted viral RNA was further purified by adding 100 µl of chloroform (1:1 v/v). The mixture was vortexed briefly and centrifuged at 14K RPM for

15 sec. and the upper aqueous layer containing the RNA was transferred to a new Eppendorf tube. The RNA was precipitated by adding 1/10 volume (10 µl) of 3M sodium acetate, pH 5.5, and 3 volumes of 100% ethanol (300 µl) and chilled at -80°C for 20 min. The solution was warmed to room temperature and centrifuged at 14K RPM in a bench top centrifuge for 15 min. to pellet the RNA. RNA was resuspended in 50 µl water and RNA concentration was quantified by diluting 5 µl of the RNA solution in 995 µl distilled water and measuring the optical density with a Beckman spectrophotometer (model DU-64, Beckman Instruments, Inc., Palo Alto, CA) at $A_{260\ nm}$ and $A_{280\ nm}$. An A_{260}/A_{280} ratio was determined and compared to the published value (Dias and Allen, 1980).

An aliquot (1 µg) of RNA, as determined spectrophotometrically, was electrophoresed on an agarose gel (0.8% agarose (w/v) dissolved in 1 x TBE: 10X=0.02 M EDTA, 1M Tris base, 1M boric acid, pH 8) to verify RNA concentration and to assess its quality. A horizontal mini-gel apparatus (Owl Scientific, Inc., Woburn, MA) with a running buffer consisting of 1 x TBE and 0.1 µg/ml ethidium bromide was used for all agarose gel electrophoresis experiments.

Synthesis and Cloning of PRMV RNA 1 cDNA

Synthesis and cloning of PRMV cDNA utilized a cDNA synthesis kit and a protocol adapted from manufacturer recommendations (Amersham Corp., Arlington Heights, IL). Using 1.0 µg purified PRMV RNA template and 1.2 µg oligo $(dT)_{12\text{-}18}$ primer, first strand cDNA synthesis was initiated by avian myeloblastosis virus (AMV) reverse transcriptase. First strand reaction mixture was incubated at 42°C for 1 hr, then placed on ice.

Primers for second strand synthesis were generated by nicking the viral RNA template with *E. coli* ribonuclease H and subsequent replacement of the RNA strand with dNTPs by DNA polymerase I. The second strand cDNA mixture (100 µl reaction volume) was incubated sequentially at 12°C for 1 hr and at 22°C for 1 hr. DNA

polymerase I was heat-inactivated at 70°C for 10 min. T4 DNA polymerase was added to 2.5 units per μg of original mRNA template, and the reaction was incubated for 10 min. at 37°C. Reaction was terminated by the addition of 4 μl of 0.25 M EDTA, pH 8.0, per 100 μl of second strand reaction mixture. Three prime to 5'-exonuclease activity of T4 DNA polymerase ensured blunt-ended termini of duplex DNA fragments which facilitated blunt end ligation of the cDNA fragments.

Double stranded cDNA product was purified using phenol/chloroform extraction and ethanol precipitation. The cDNA was extracted with an equal volume of phenol: chloroform: isoamyl alcohol (25:24:1 v/v/v). The aqueous phase was extracted once with an equal volume of chloroform (100 μl) and ethanol precipitated with one volume of 4M ammonium acetate, pH 5.5, and 2.5 volumes of -20°C ethanol. Following 15 min. precipitation at -80°C the mixture was centrifuged (14K RPM, 15 min.) to pellet the cDNA. The supernatant was vacuum aspirated and the resultant cDNA pellet was washed with 100 μl of 2M ammonium acetate, pH 5.5, and 200 μl of -20°C 70% ethanol by gentle agitation. Washed cDNA was pelleted (5 min. at 14K RPM); the supernatant was aspirated and the pellet dried for 2 min. in a vacuum. The cDNA pellet was resuspended in 50 μl of distilled water and purity and concentration of the double stranded cDNA product was estimated by electrophoresis on a 0.8% agarose gel. This product provided the 3'-terminal cDNA clone.

To complete cloning of RNA1, five additional cDNA clones were primed by oligonucleotides designed to complement the desired upstream sequence. The primer sequences were derived from the 5'-terminal nucleotide sequence of the appropriate 3'-cDNA clone and are listed in Appendix B. Primers were designed to produce overlaps between adjacent cDNA clones that contained a restriction endonuclease recognition site. Oligonucleotides were synthesized by the Macromolecular Structure Facility, Department of Biochemistry, Michigan State University, E. Lansing, MI. Synthesis of PRMV cDNA

with upstream primers was similar to that described for oligo (dT)-primed cDNA synthesis.

Plasmid vector Bluescript KS- (Stratagene, La Jolla, CA) was linearized at the polylinker EcoRV site. All restriction endonucleases and their appropriate incubation buffers used in subsequent steps were obtained from Boehringer Mannheim Corporation (Indianapolis, IN). Linearized plasmid (0.5 μg) was treated with 5 units of calf intestinal phosphatase (Boehringer Mannheim) to prevent self-ligation of vector termini (Tabor, 1987). All cDNA clones were ligated into the EcoRV site of KS- using a vector/cDNA ratio of 1:2 with 0.1 μg of vector. T4 DNA ligase (0.1U/μl), 1x T4 DNA ligase buffer (10x: Tris-HCl, 660 mM; $MgCl_2$, 50 mM; dithiothreitol, 10mM; ATP, 10mM; pH 7.5) and dATP (1mM) were added to the vector/insert mixture and incubated overnight at room temperature (22°C) in a 25 μl reaction volume. Ligation product was used to transform *E.coli* DH5α 'Max Efficiency' (Life Technologies, Gaithersburg, MD) calcium chloride competent cells; transformation mixture containing *E. coli* and the ligation product was incubated on ice for 30 min., followed by 2 min. of heat shock at 37°C (Hanahan, 1983). The entire transformation mixture (150 μl) was plated on solid 2xYT agar media (1.6% w/v tryptone, 1.0% w/v yeast extract, 0.5% w/v sodium chloride and 15% w/v agar). Agar plates were amended with ampicillin to aid in selection of plasmid-containing colonies (ampicillin, 50 μg/ml); blue/white colony screening of recombinant plasmids was enabled by the presence of X-gal (0.004% w/v 5-bromo-4-chloro-3-indolyl-β-D-galactoside, Boehringer Mannheim) and IPTG (20 μg/ml isopropyl-β-D-thiogalactopyranosid, Boehringer Mannheim) in the agar media.

White *E. coli* colonies were selected from 2xYT plates and grown in 2 ml of 2xYT liquid with shaking at 37°C (350 RPM). Overnight cultures were collected in 1.5 ml Eppendorf tubes and centrifuged at 14K RPM in a bench top centrifuge to pellet cells (Lee and Rasheed, 1990). The supernatant was aspirated and resuspended with 100 μl of solution I (25mM Tris-Cl, pH 8.0; 10 mM EDTA; 50 mM glucose), vortexed thoroughly,

and the completely resuspended pellet was held at room temperature (22°C) for 5 min. Next, 200 µl of freshly made solution II (0.2 N NaOH, pH 5.5; 1% v/v sodium dodecyl sulfate) was mixed into the solution by inverting the tube and the mixture was incubated in an ice-water bath for 5 min. A 150 µl aliquot of 7.5 N ammonium acetate, pH 5.5, was added and the mixture was placed on ice for 5 min. and centrifuged for 5 min. at 14K RPM. The plasmid-containing supernatant was transferred to a fresh Eppendorf tube containing 0.6 volume isopropanol and incubated for 10 min. at room temperature. This solution was centrifuged at 14K RPM for 10 min. and the supernatant was aspirated. The pellet was resuspended in 100 µl 2 N ammonium acetate, pH 5.5, and placed on ice for 5 min. Following centrifugation at 14K RPM for 5 min., the supernatant was transferred to another Eppendorf tube containing 100 µl of isopropanol and held at room temperature for 10 min. A final 10 min. centrifugation (14K RPM) pelleted the purified plasmid DNA and the supernatant was aspirated. The pellet was dried in vacuum for 3 min. and resuspended in 50 µl water. Resuspended, purified plasmid DNA containing cDNA inserts were size selected by restriction endonuclease analysis.

To determine the size of each cDNA insert, 0.1 µg of recombinant plasmid was linearized with a restriction endonuclease (EcoRI, XbaI or XhoI) with a unique recognition site within the vector polylinker. In a reaction volume of 20 µl, including 2 µl manufacturer-supplied 10X incubation buffer, 0.1 µg DNA, 1.0 unit of restriction endonuclease and water, the DNA template was digested for 3 hr at 37°C. Following restriction digests, 2 µl of loading dye (30% glycerol, 0.25% bromophenol blue in water) was added to the digestion mixture. The entire reaction was electrophoresed in a 1X TBE running buffer with 0.1 µg/ml ethidium bromide in a 0.8% agarose gel at 100 volts (EC452 power supply, E-C Apparatus Corp., St. Petersburg, FL; Horizontal Electrophoresis System, #A1 Owl Scientific, Inc., Cambridge, MA) for approximately 3 hr. The gel was photographed and plasmid sizes were compared with 1 kilobase (kb)

double-stranded DNA ladder (Life Technologies, Gaithersburg, MD) and linearized KS-plasmid. The cDNA clones with the largest inserts were selected for further analysis.

RNA1 Origin of cDNA Clones Verified

The RNA1 origin of each cDNA clone was confirmed by probing a PRMV northern blot with a nucleotide sequence unique to each cDNA clone. One μg of purified PRMV RNA was electrophoresed in a non-denaturing agarose gel (0.8% w/v) and transferred to nylon membrane (Fisher Scientific, Pittsburgh, PA) according to Brown (1993). One hundred pmol of oligonucleotide primer RA42 (5'-AAATCATCATCGATCTCAAC-3'), complementary to position 4838-4857 near the 5'-terminus of the 3'-most cDNA clone, was labeled with digoxigenin-11-dUTP according to manufacturer's recommendations (3'-Oligonucleotide Tailing Kit, Genius System version 2.0, Boehringer Mannheim). The RNA1 origin of each upstream clone was confirmed by probing a PRMV RNA blot with the digoxigenin-labeled synthetic oligonucleotide primer used for its synthesis. The product of the labeling reaction was an oligonucleotide with a 3'-tail containing multiple digoxigenin-11-dUTP residues which was then diluted in 10 ml northern pre-hybridization solution (5X SSC, 50% formamide 0.02% sodium dodecyl sulfate, 2% (w/v) blocking reagent (Boehringer Mannheim), and 20 mM sodium maleate, pH 7.5) to a concentration of 10 pmol probe/ml. The northern blot was incubated for 2 hr at 37°C in northern pre-hybridization solution in a sealed plastic bag. Northern pre-hybridization solution was discarded from the bag and replaced with the dilute probe solution. At the end of a 6 hr hybridization at 37°C, dilute probe was decanted and stored at -20°C. The treated membrane was washed twice, 5 min. per wash, in 2X wash solution (2X SSC containing 0.1% SDS) and subsequently washed twice with 0.5X wash (0.5X SSC containing 0.1% SDS), 15 min. per wash. All membrane washes were performed at room temperature (approximately 22°C). Following washing, the membrane was incubated in 50 ml maleate buffer (100 mM maleic acid, 150

mM NaCl, pH 7.5) for 1 min. Maleate buffer was replaced with 50 ml of northern blocking solution (10 mM sodium maleate containing 1 % (w/v) blocking reagent) in which the membrane was incubated for 30 min. One µl (0.75 U) of anti-digoxigenin alkaline phosphatase Fab fragments (Boehringer Mannheim) was added and the membrane was incubated in this solution for 30 min. at room temperature. Treated membrane was washed twice, 15 min. per wash, with 200 ml maleate buffer at room temperature and equilibrated in 50 ml Genius buffer 3 (100 mM Tris-HCl, 100 mM NaCl, 50 mM $MgCl_2$, pH 9.5). The membrane was removed from this solution and placed on blotter paper (roughly 2X membrane size), covered with approximately 1 ml Lumi-Phos 530 (Boehringer Mannheim), and wrapped in Film Wrap (Gordon Food Service, Grand Rapids, MI). Treated membranes were exposed to X-OMAT autoradiograph film (Eastman Kodak Inc., Rochester, NY) for approximately 30 min. Autoradiographs were developed (model M7B RP X-OMAT Processor, Kodak) and analyzed.

Confirmation of the RNA1 origin of each cDNA clone was followed by nucleotide sequence analysis; the largest of the cDNA clones obtained from cDNA synthesis with upstream oligonucleotide primers was sequenced to verify its 3'-overlap with the preceding clone. The location and nucleotide sequence of primers used for cloning and northern analysis of each cDNA clone are listed in Appendix B. Oligonucleotide probe preparation and northern analysis of the five upstream cDNA clones was similar to that described above.

Exonuclease III Deletion Analysis

Exonuclease III (Boehringer Mannheim, Indianapolis, IN) was used to create a series of deletions originating at each end of the seven full length cDNA inserts (Henikoff, 1984). Varying the duration of exonuclease III (exo III) treatment and subsequent treatment with mung bean nuclease (Life Technologies, Gaithersburg, MD) generated a series of progressively smaller subclones (Maniatis *et al.*, 1982).

Restriction endonucleases were used to generate 3'-overhangs adjacent to the sequencing primer (either universal or reverse) site and 5'-overhangs adjacent to the insert cDNA. For exo III deletion of the KS-/insert cDNA plus strand, ApaI and XhoI were used to create 3-' and 5'-overhangs respectively; SacI and XhoI were used for similar treatment of the opposite strand. Exo III treatment produced unidirectional deletions from 5'-overhangs. The linearized DNA was extracted once with buffered phenol, once with phenol/chloroform/isoamyl alcohol (25/24/1 v/v/v), and ethanol precipitated (see above). DNA was washed with -20°C 70% ethanol, dried in a vacuum and resuspended in water for a final concentration of 0.1 μg/μl.

The temperature and the time of exo III incubation were regulated to obtain a deletion series on both strands of each cDNA clone (Henikoff, 1984). Exo III activity was terminated by transferring equal volume aliquots of the digestion mixture to 1X mung bean nuclease buffer (10X concentration: 0.3M sodium acetate, pH 5.0; 0.5 M sodium chloride; 10 mM zinc chloride in 50% glycerol) and heating the mixture to 68°C for 15 min. Subsequent treatment with mung bean nuclease ensured blunt-ended termini of the nested deletion mutants. Exo III-deleted plasmids were re-circularized overnight by ligation with T4 DNA ligase at 14°C (Slatko *et al.*, 1994). Competent DH5α *E.coli* cells were transformed with the ligation product and screened on 2xYT plates amended with ampicillin, X-Gal, and IPTG, as described earlier.

Individual *E. coli* colonies containing exo III-deleted plasmids were selected from 2xYT selection plates and grown overnight in 2xYT with agitation (350 RPM) at 37°C. Purification of plasmid DNA was similar to that described earlier. Three to 5 μl of the DNA solution was linearized by restriction enzyme digestion. Restriction endonuclease ClaI was used to linearize the series of subclones with exo III deletions generated from the 5'-overhang of XbaI. Endonuclease XbaI was used to linearize exoIII mutants generated from the 5'-overhang of XhoI and reaction conditions for restriction endonuclease digestion were described earlier. Linearized plasmids were electrophoresed

on an agarose gel 0.8% (w/v) and a series of consecutive deletions each differing by 150-200 bases was selected for sequence analysis. Exo III clones were selected in this manner to bring the entire cDNA fragment into the sequencing range of either universal or reverse primers (Boehringer Mannheim) whose respective recognition sites are located within the Bluescript KS- vector polylinker.

Nucleotide Sequence Analysis of cDNA

The cDNA clones of PRMV RNA1 were completely sequenced in both directions by the dideoxynucleotide chain-termination method of Sanger *et al.* (1977). DNA templates sequenced included full length cDNA fragments as well as exo III deletion subclones. DNA sequencing reactions were primed using either the universal or reverse primers on the pBluescript KS- vector (Stratagene, La Jolla, CA). DNA sequencing protocol with Sequenase (Sequenase Version 2.0, United States Biochemical, Cleveland, OH) was adapted from manufacturer's recommendations.

For each double stranded DNA template, 2 μg of DNA was combined with 0.1 volume of 2N NaOH and 0.2 volume of 1 mM EDTA, pH 8, in a 30 μl reaction volume. The reaction was incubated for 30 min. at 37°C and ethanol precipitated using 0.1 volume sodium acetate and 3 volumes -20°C ethanol. Following a 15 min. incubation at -80°C, the mixture was centrifuged at 14K RPM for 10 min. The pellet was washed in 70% ethanol, dried and resuspended in 7 μl water.

One μl (0.5 pmol) of either the reverse or forward oligonucleotide primer, as well as 2μl of 5X Sequenase buffer (0.2M Tris-HCl, pH 7.5, 0.1 M $MgCl_2$, 0.25 M NaCl) were combined with the denatured DNA solution. The tube containing the primer/DNA solution was heated to 65°C for 2 min. and slowly cooled to room temperature in a water bath (approx. 45 min.). Eppendorf tubes containing the annealed mixture were placed on ice.

Sequenase was added to the DNA/primer hybrid solution in the presence of deoxynucleotides (dNTPs), dithiothreitol (DTT) and ^{35}S dATP (Dupont, Boston , MA),

for DNA polymerization. Polymerization reaction mixture contained the manufacturer's recommended volume of 5X Sequenase labeling mixture, 7.5 μM dGTP, 7.5 μM dCTP, and 7.5 μM dTTP, 5 μCi ^{35}S-dATP and Sequenase diluted 1:10 (v/v) in accompanying enzyme dilution buffer. Solution was mixed and incubated at room temperature (22°C) for 5 min.

Termination reactions were performed in a V-bottom 96-well microwell plate (VWR Scientific, Batavia, IL). Prior to the addition of the product of the labeled reaction to 2.5 μl of the termination mixture, supplied dideoxynucleotide termination mixtures were pre-warmed at 37°C for 2 min. Three and one half μl of the labeling mixture were added to each of the four wells in the microwell plate and incubated for 5 min. at 37°C. Four μl of stop solution (95% formamide, 20 mM EDTA, 0.05% bromophenol blue, and 0.05% Xylene Cyanol) were added to each microwell and samples were denatured at 80° C for 2 min. prior to electrophoresis.

Sequence reactions were electrophoresed on 8% acrylamide gels (41.5 cm x 37.0 cm). An 8% acrylamide gel contained 30 ml 20% acrylamide solution (96.5 g acrylamide, 3.35 g methylene-bis-acrylamide, 233.5 g urea in 500 ml 1X TBE); 45 ml 8 M urea; 75 μl 25% ammonium persulfate; and 75 μl N,N,N',N'-tetramethylethylenediamine (TEMED). The 8% acrylamide gel polymerized between 0.25 mm spacers (approx.1hr) and pre-run in 1X TBE running buffer at 65 watts (W) for 30 min. (Fisher FB650 power supply, Pittsburgh, PA; SE1500 Sequencer apparatus, Hoeffer, San Francisco, CA). Two and one half μl of each sequence reaction sample was added per lane between the teeth of a shark tooth comb (Hoeffer, San Francisco, CA) and electrophoresed at 65 W for approximately 2.5 hr.

The gel was removed from the sequence apparatus, immersed in fixer solution (15% methanol, 5% acetic acid) for 20 min., transferred to 41.5 cm x 32 cm x 3 mm chromatography paper (Whatman, Hillsboro, OR) and covered with film wrap. The sequence gel was dried under vacuum at 80°C for 1hr (Vapor Trap, Vacuum Pump,

BioRad, San Francisco CA; Slab Gel Dryer, Hoeffer Scientific, San Francisco, CA) and exposed to 43 cm x 35 cm autoradiograph film (Kodak X-OMAT AR) overnight in an autoradiograph cassette (Fisher FBXC 1417). Autoradiographs were developed in an automatic film developer (Kodak RP X-OMAT Processor, model M7B) and analyzed.

RNA Sequence Analysis

Primer extension was used to determine the nucleotide sequence of the 5' terminus of RNA1 after the method of Fang *et al.* (1995) using viral RNA template and synthetic oligonucleotide RA75 (5'-GACCAAATATTCCATCAC-3') complementary to RNA1 nucleotide position 50-67. The 5'-terminal nucleotide of RNA1 was verified using terminal deoxynucleotidyl transferase (Allison *et al.*, 1988).

Computer-Assisted Genome Analysis

Sequence data were analyzed using Genbank databases and a Genetics Computer Group (GCG) software package (version 8.1) available through Silicon Graphics, Inc. computer services at the Department of Biochemistry, Michigan State University, E. Lansing, MI. The sequence data were assembled and manipulated through the SEQED program. Restriction endonuclease recognition sites were verified using the MAP, MAPPLOT, and MAPSORT programs (Devereaux *et al.*, 1984). Sequence comparisons utilized BESTFIT and GAP GCG programs (Devereaux *et al.*, 1984). Parameters for BESTFIT and GAP included a gap creation penalty of 5.0 and a gap extension penalty of 0.3. Statistical significance of alignments was assessed by including a randomization program (RAN) with GAP and BESTFIT. Ten randomized comparisons were made for each pair of sequences by repeatedly shuffling one of the sequences and aligning it with the non-randomized sequence. Similarity between sequences was deemed significant if it exceeded the mean randomized similarity plus three standard deviations (Doolitle, 1981). Viral sequences used for comparison to the PRMV RNA1 sequence were obtained from Genbank. Genbank accession numbers for viral sequences are included in Appendix A.

Multiple sequence alignments utilized the PILEUP programs (GCG) with a gap weight of 3.0 and a length weight of 0.1. Consensus sequences were generated using files created by the PILEUP program, followed by analysis with PRETTY (plurality of either 4.0 or 6.0; vote weight and threshold of 1.0, each)(Devereaux *et al.*, 1984). Secondary structure predictions of proteins used PLOTSTRUCTURE from GCG (Devereaux *et al.*, 1984). Optimal secondary structures for the 3'-UTR of RNA 1 were predicted with GCG FOLDRNA (Jaeger *et al.*, 1989). Output files from FOLDRNA were used to plot RNA secondary structures with GCG SQUIGGLES (Devereaux *et al.*, 1984).

RESULTS AND DISCUSSION

Propagation and Purification of PRMV

Purification of PRMV yielded 0.2-0.3 mg virus/100g infected *C. quinoa*. A tracing of a typical sucrose density gradient fractionation for PRMV is shown (Fig.1a) with top (T), middle (M), and bottom (B) components present in varying quantities. In similar tracings, Dias and Allen (1980) observed that B component, RNA1, frequently was present in higher proportion than M component, RNA2. An absorbence ratio E_{260}/E_{280} of 1.8 for pooled M and B components compared well to published values by Dias and Allen (1980) which ranged from 1.7 to 1.9.

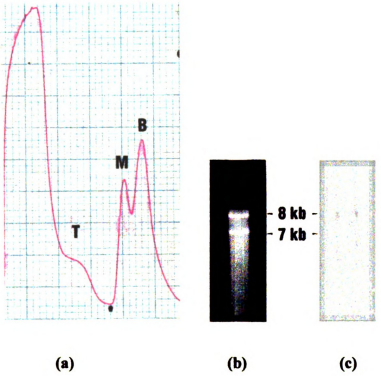


Fig 1. Purification of PRMV and confirmation of the RNA1 origin of an oligo (dT)-derived cDNA clone. The ultraviolet absorbance scanning pattern following 0-30% linear logarithmic sucrose density gradient centrifugation of purified PRMV is shown (a). The middle (M) and bottom (B) components were collected separately from top (T) component (empty capsid protein) and further purified by phenol extraction to liberate RNA1 and RNA2 from components B and M, respectively. In (b), 1 μg of purified PRMV RNA was electrophoresed on a non-denaturing 0.8% (w/v) agarose gel, stained with ethidium bromide and photographed. The RNA gel from (b) was northern blotted to nylon membrane and probed with digoxigenin-dUTP-labeled oligonucleotide primer RA42, the same primer used for PRMV cDNA synthesis of the 3'-terminal cDNA clone. Probe construction and northern analysis is detailed in the text. A photograph of the digoxigenin/RA42-probed northern blot is shown (c).

Purified PRMV RNA separated electrophoretically as two distinct bands corresponding to RNA1 and RNA2 with estimated sizes of 8 and 7 kb respectively (Fig.1b) and matched RNA1 and 2 sizes predicted by Dias and Allen (1980). PRMV RNA1 and RNA2 sizes are comparable to those of subgroup II nepovirus TomRSV (8214 and 7273 nucleotides, respectively). The RNA2 was substantially larger than the 5.4 kb cutoff for a subgroup I nepovirus (Sanfaçon, 1995). The RNA1 origin of each cDNA clone was confirmed by probing a PRMV RNA northern blot with the digoxigenin-labeled oligonucleotide primer used for its synthesis. For example, the oligonucleotide complementary to the 5'-terminus of clone 5-9 was used to prime synthesis of cDNA clone 2-1. Hybridization of this oligonucleotide to only RNA1 (Fig. 1c) evidenced the RNA1 origin of both clones.

cDNA Synthesis and Sequencing of PRMV RNA1

A series of six overlapping cDNA clones was selected for sequencing PRMV RNA1. The molecular cloning strategy is illustrated in Fig. 2. The cDNA clones are designated 5-9, 2.1, 50-3.9, 52-4.2, 68-2.90, and 70.20 and contained inserts of the following sizes, respectively: 3120, 1202, 1408, 1131, 882, 190. Optimal cDNA synthesis (i.e. largest cDNA product) occurred when the RNA template was heated at 70°C for 1 min, then placed immediately on ice prior to addition of Amersham's first strand cDNA synthesis reaction components. A series of exoIII deletion mutants was created for both directions in all six full length cDNA clones. At 37°C, nucleotide digestion rates at susceptible 5'-ends were approximately 250 bases/min. The deletion series for clone 5-9 illustrates the technique (Fig.3). Each exo III subclone chosen for nucleotide sequencing was 150-300 nucleotides shorter than the preceding clone and together spanned the entire

cDNA insert. The cDNA sequence analysis and assembly indicated that collectively, these inserts represent 99.6% of the PRMV RNA1 genome. The sequence of the 5'-terminal region of RNA1 was determined by direct dideoxynucleotide chain termination sequencing of the genomic RNA template using AMV reverse transcriptase. Direct RNA sequencing indicated that the 5'-terminal cDNA clone contained all but the 5'-terminal 44 nucleotides of RNA1. The 5'-terminal nucleotide (U) was identified with TdTase treatment. The complete unique nucleotide sequence of the cDNA of PRMV RNA1 is 7977 nucleotides (Fig.4). A polyadenylated tract of 30-60 ATP residues is located at the 3'-terminus of the RNA1 nucleotide sequence.

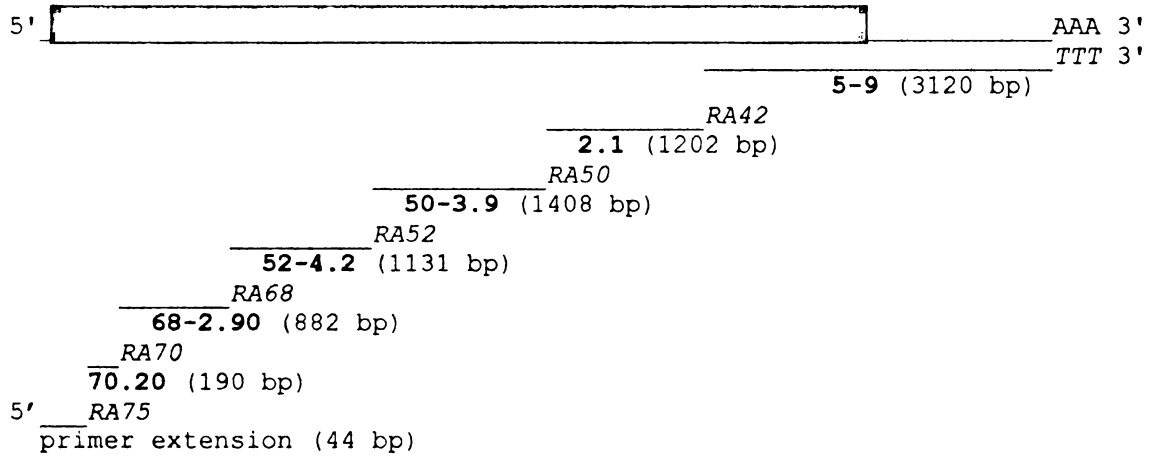


Fig.2. Cloning strategy for PRMV RNA1. The red rectangle above RNA1 represents the major ORF present in the virion sense. Oligonucleotide primers oligo (dT), RA42, RA50, RA52, RA68, and RA70 (italicized) were used to generate a consecutive series of RNA1 cDNA clones; respective primer nucleotide sequences are identified in Appendix B. The primer RA75 was used in primer extension analysis to determine the 5'-terminal 44 nucleotides of RNA1. Overlapping cDNA clones (boldface) are shown with their respective lengths in base pairs (bp); cDNA clones are positioned relative to RNA1.

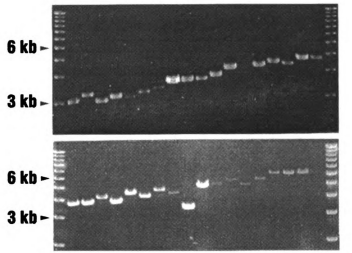


Fig.3. Exonuclease III (exoIII) digestion of PRMV RNA1 cDNA clone 5-9. The 5-9 cDNA contains 3120 unique nucleotides. ExoIII was used to create a series of single-stranded nested deletions from full-length 5-9 cDNA insert. Mung bean nuclease treatment degraded the remaining single strand and ensured blunt-ended termini which were ligated to re-circularize exo III-treated 5-9 deletion plasmids. Plasmids were linearized with 5 units of XbaI and subjected to electrophoresis in 0.8% agarose gel. Linearized deletion mutants are arranged in size on the agarose gels relative to the 1 kb ladder (on both sides of each gel); the exo III deletion series continues from the upper gel to the lower gel. Sizes of plasmids ranged from approximately 3 kilobase-pair (kb) fragments (little or none of the 5-9 insert remaining) to vector (Bluescript KS-) plus complete insert (approximately 6 kb, total). From these clones, a series of nested exoIII deletion mutants (approximately 100-300 nucleotides apart in length) was selected for sequence analysis of the entire 5-9 cDNA.

```
1     TATGAAAAATCACTAATCTATTACCTTCTTAACTATTGCTGTTTCTTTTGTGATGGAATA
                                                        M  E  Y    3

61    TTTGGACTATCTTCCTGCGCGAACAAAATGGGTGGCCATAGTGCCAAAAGCTGTCCTGGA
       L  D  Y  L  P  A  R  T  K  W  V  A  I  V  P  K  A  V  L  E   23

121   AGCCACCAGGATAGCTAATGTCCTGCTAGCAAAGCCTGCCAACTTTGCTATTTCTTTTTT
       A  T  R  I  A  N  V  L  L  A  K  P  A  N  F  A  I  S  F  L   43

181   GGCTCAGGGTGCCTCCCTGAAGCCACGTTCTGTAGCTCTGGCGGTTGCAATGGGTTATTG
       A  Q  G  A  S  L  K  P  R  S  V  A  L  A  V  A  M  G  Y  C   63

241   CCACTGGCCCAGAGTTCTGCATCTATACTCCGAAGGAGTTCCCCTAACTTGGGGAGATGC
       H  W  P  R  V  L  H  L  Y  S  E  G  V  P  L  T  W  G  D  A   83

301   ACCACCGGTGCCCCTTTTATTAAGGGCCCTGGCTAAGATGGAATCTGGGCTATATGCCGA
       P  P  V  P  L  L  L  R  A  L  A  K  M  E  S  G  L  Y  A  D  103

361   TGGGAGAGGAACTGGCTTTTTGCCAGTTCAAGAGGCAAGTGCCTCACCTGCGGGCCGCCA
       G  R  G  T  G  F  L  P  V  Q  E  A  S  A  S  P  A  G  R  Q  123

421   GCAAGCCGTCGAAGAGAAAAAGGCTCTTTACAGAGCCAAAGGTGCTGCAGCAACAGCATC
       Q  A  V  E  E  K  K  A  L  Y  R  A  K  G  A  A  A  T  A  S  143

481   GAAAAAGGCTGCTGCTAGAGCAGCCTTGGAAGCCCGCCGTTCCTGTGGCGGACAAGGAAG
       K  K  A  A  A  R  A  A  L  E  A  R  R  S  C  G  G  Q  G  R  163

541   AGCGCCTAAAGTACTGAAAAAGAAGGCCACCAAGCGGGTGGTCACTGCTGCACTGGCAAC
       A  P  K  V  L  K  K  K  A  T  K  R  V  V  T  A  A  L  A  T  183

601   AGTCAAAGAGAGCCAACGCTTGGCTCTATTTTTCCTTTTTCCTCTTCTCTCTTTTCCTCT
       V  K  E  S  Q  R  L  A  L  F  F  L  F  P  L  L  S  F  P  L  203

661   CCCCCTCTCCTCCGTGAAAAGGGGGTTCCTTTTAATCCTCCTCAACGGGAGGATTTTCTT
       P  L  S  S  V  K  R  G  F  L  L  I  L  L  N  G  R  I  F  F  223

721   TCCTCTCCTCCTCCTCCTTTGGTGGCTTTGTAAAAGCCCACTTTCTTATGGGTCCTATTG
       P  L  L  L  L  L  W  W  L  C  K  S  P  L  S  Y  G  S  Y  C  243

781   TGGACCTTGGGCCTCTCTTGGCCCTATTTTAGAAACTGGAGCTCCAGGAGCTCAACGGGC
       G  P  W  A  S  L  G  P  I  L  E  T  G  A  P  G  A  Q  R  A  263

841   ACTTTTTGCCGCTATTCGAAAACTCCCTCTCTCTACTTTTCACGAGAGAGTTCTCTTCCG
       L  F  A  A  I  R  K  L  P  L  S  T  F  H  E  R  V  L  F  R  283

901   GGATACTCAAGTTGCAGTGTCCCAACTTTTCGTTTTGTATCCCTCTGTACATATACTTGG
       D  T  Q  V  A  V  S  Q  L  F  V  L  Y  P  S  V  H  I  L  G  303

961   GGATCTTAATTCTTTTTTCCTTCAGGATTGCCGTGGCATGCGTTTAGCACTGGAAAGTGC
       D  L  N  S  F  F  L  Q  D  C  R  G  M  R  L  A  L  E  S  A  323

1021  TCGACGTATTGCAGATGGTATTTCCTCCATTCTTCCTCAGCATCGGGTTGTACATACTTT
       R  R  I  A  D  G  I  S  S  I  L  P  Q  H  R  V  V  H  T  F  343
```

Fig. 4 (cont'd).

```
1081  TCTTGATGCAGTGAAGAAGGTTGGTTCCTATATTTCAGGAGCTGCCTCTGCAGTTAAAAG
       L  D  A  V  K  K  V  G  S  Y  I  S  G  A  A  S  A  V  K  S  363

1141  TAAAGTTTCTAACTTTACCTCTTCACTCTTTGATTCTATTTTGGACAAATGTAAATATTG
       K  V  S  N  F  T  S  S  L  F  D  S  I  L  D  K  C  K  Y  C  383

1201  TTTCATGTCCACTTTTTCTCCCTTCTTGGCTTCTCTGCAATCAGCCAAAGCTGAAATTGA
       F  M  S  T  F  S  P  F  L  A  S  L  Q  S  A  K  A  E  I  E  403

1261  AAAATTTTGGCAAA3ATTGCATGAGTTGGGCTAGGAACTTGTGGAGTAAGGCTCACCTTGC
       K  F  W  Q  N  C  M  S  W  A  R  N  L  W  S  K  A  H  L  A  423

1321  TCTACAAGCTCTTGGCCTTTATGCCATTTGGGCTTTAGTGTTGACAATCCTTTGTGGGAT
       L  Q  A  L  G  L  Y  A  I  W  A  L  V  L  T  I  L  C  G  I  443

1381  TGTTTATTTATTAGAATCTCTTTTTATTACTGCGGGGGTAATAGGCTCCCATGGTATTAT
       V  Y  L  L  E  S  L  F  I  T  A  G  V  I  G  S  H  G  I  I  463

1441  TCTCTCTATTTTTCTTTCCGTGGTTATGGCTGCAGCTGGATTCACTATCTTTACCGTTGG
       L  S  I  F  L  S  V  V  M  A  A  A  G  F  T  I  F  T  V  G  483

1501  TAAAGAAAGTGCTCAAATGATTCGGACAATGCGCGAGGGTATTCTCATGATGGTAATACC
       K  E  S  A  Q  M  I  R  T  M  R  E  G  I  L  M  M  V  I  P  503

1561  CGATGATGCCGCTAAGTCGATTGGAGGTAGAACCAGGTACCCAACAGTGCATAGTCTTTT
       D  D  A  A  K  S  I  G  G  R  T  R  Y  P  T  V  H  S  L  F  523

1621  TGATTTGGCTATGGCACCTGTAAATTTTTTGGAGTCCATTGCTAGTGGACTTTCTCTTTT
       D  L  A  M  A  P  V  N  F  L  E  S  I  A  S  G  L  S  L  F  543

1681  TTCCACCTCCTCAATTACAGTTTTAGGTAAATTGGGGAATTCTTTGGAAGGTATTCGGAA
       S  T  S  S  I  T  V  L  G  K  L  G  N  S  L  E//G  I  R  K  563

1741  AGGCTATAATTGCCTGACCGATTTTATTTCCATTTTCTTTGAGAAGATGGGAGGTCTATG
       G  Y  N  C  L  T  D  F  I  S  I  F  F  E  K  M  G  G  L  W  583

1801  GGAAGGTATTTCTGGTAAGCAGACCACCTTCTTTCGAGATCTCACCACGGCTGTTAAGAT
       E  G  I  S  G  K  Q  T  T  F  F  R  D  L  T  T  A  V  K  I  603

1861  TAATATCAGTTCGTGGACCCAGGATGCTCGTCGGTTAATTGAATACCACGAGATGGCTGG
       N  I  S  S  W  T  Q  D  A  R  R  L  I  E  Y  H  E  M  A  G  623

1921  TACCCTGGATAAGTTCGAGTACGAGAAAGTTCGCCTCTTATTTATCAAGGGAAGAATAGT
       T  L  D  K  F  E  Y  E  K  V  R  L  L  F  I  K  G  R  I  V  643

1981  CGATACTGCCAATAAGGGCAGGCAATCCCATACCAGCAACCAATTTTTGAGAGTTGTTGG
       D  T  A  N  K  G  R  Q  S  H  T  S  N  Q  F  L  R  V  V  G  663

2041  TTCTTTGTTGACAGATTTGAGGGAGGTGCGTGCTAAGTGCGCTCGTTCCCTCCGTTTTGA
       S  L  L  T  D  L  R  E  V  R  A  K  C  A  R  S  L  R  F  D  683
```

Fig. 4 (cont'd).

```
2101  TGGTTGGCGTCGTCAACCTTTTTGGGTTTATATTTTCGGTGCATCACAGTGTGGTAAGTC
       G  W  R  R  Q  P  F  W  V  Y  I  F  G  A  S  Q  C  G  K  S  703

2161  CACTTTAGCCAACTATTTGTGCCCCCTTTTATTGGCACATATGGGTTGGGATGCTCATGA
       T  L  A  N  Y  L  C  P  L  L  L  A  H  M  G  W  D  A  H  D  723

2221  CGTCTACTCCAAGGATCCCACAGAAGGATACTGGAGTGGATACTACCAGCAGAAATGTTT
       V  Y  S  K  D  P  T  E  G  Y  W  S  G  Y  Y  Q  Q  K  C  L  743

2281  AAAGATGGATGATCTTTCTGCGGTAGTGCCTAAGCAGGTATCTCCTCTTGAGCAACAGCT
       K  M  D  D  L  S  A  V  V  P  K  Q  V  S  P  L  E  Q  Q  L  763

2341  CATTCCCCTTATTTCTACGGAGGAGAAAATGGTATCTGCAGCTGAGATTAATGGCAAGGG
       I  P  L  I  S  T  E  E  K  M  V  S  A  A  E  I  N  G  K  G  783

2401  AATTCAGTTTTTATCTGAATTGGTCATATCCAGCTCGAATGTGAATGATGCACCTACATC
       I  Q  F  L  S  E  L  V  I  S  S  S  N  V  N  D  A  P  T  S  803

2461  GTGTGAGATTCTTGATCCTGAAGCATATCGCCTAAGGAGAAAGGTTCTCTTACGCTGTAG
       C  E  I  L  D  P  E  A  Y  R  L  R  R  K  V  L  L  R  C  R  823

2521  ACGTGCAGCGACTTACCAGCATGATGAAGCTGGGAACAAGACTGAGGTAGTTGATGCTGA
       R  A  A  T  Y  Q  H  D  E  A  G  N  K  T  E  V  V  D  A  E  843

2581  GGGAAATATTGTGTGTCGACAATATGATCCCAGTGATGCATTAGCTTGCACTGAAGTCAG
       G  N  I  V  C  R  Q  Y  D  P  S  D  A  L  A  C  T  E  V  S  863

2641  TGGCTACATGCCAATTCTTGTACTCGGTTCCAGGACCAGCAGGACTGTGGCACCCGCCCA
       G  Y  M  P  I  L  V  L  G  S  R  T  S  R  T  V  A  P  A  H  883

2701  CTCCACCATTCCTCTCATTAAGGATGCCATGGATGCGCATTTCTTAGTAGAGGATGCCAA
       S  T  I  P  L  I  K  D  A  M  D  A  H  F  L  V  E  D  A  K  903

2761  AAGAGAAGCGTGGGTGCAACAAACAAATATGCACTCGCGAACTGGAGCTGAGGTCTCCAG
       R  E  A  W  V  Q  Q  T  N  M  H  S  R  T  G  A  E  V  S  S  923

2821  CTATTTGCAATCCTTAGTGTGTGCACTGGGCTCTTATAAAGCCATTCAGCGCTCTTCCGA
       Y  L  Q  S  L  V  C  A  L  G  S  Y  K  A  I  Q  R  S  S  D  943

2881  CGTTTCAGATGCGGGGGAGCGTAAATTTTTGGTAGCTGTTGATGGAACTATTTATTCCAT
       V  S  D  A  G  E  R  K  F  L  V  A  V  D  G  T  I  Y  S  I  963

2941  CGATTCTTTAGGTAGGGCGACCAAGGAAGCGGCAGACGCGTACGACAATGTTGAGGCATT
       D  S  L  G  R  A  T  K  E  A  A  D  A  Y  D  N  V  E  A  L  983

3001  GGAGTCCACTACCCTTCTGCAATATCGCCTTGATTTTCGACAGGTTAGGGAACATTCCCT
       E  S  T  T  L  L  Q  Y  R  L  D  F  R  Q  V  R  E  H  S  L 1003

3061  CTTAACCAATGATGGTAGTTTCCATTCCTCTATGGTGAGGGATCTACTAAGGATATCTTG
       L  T  N  D  G  S  F  H  S  S  M  V  R  D  L  L  R  I  S  C 1023
```

Fig. 4 (cont'd).

```
3121  TGAAGAAGCTTGTGTGGTCTCTGTTGATAAAATCAGTAGGGATTCCAAACAACTTCACAG
       E  E  A  C  V  V  S  V  D  K  I  S  R  D  S  K  Q  L  H  R 1043

3181  GGACTTGTGGAGTGAGTTAAAGCTTGCGAACGATTTTTTTCCGCGTTTCTCAAAAGCTCT
       D  L  W  S  E  L  K  L  A  N  D  F  F  P  R  F  S  K  A  L 1063

3241  TAACCAACTGCGCGACCAACCACATTTTAAGGTTGATGTGCAGTCAGTTTCCTTCAGCAT
       N  Q  L  R  D  Q  P  H  F  K  V  D  V  Q  S  V  S  F  S  I 1083

3301  ATGGCTGATTTTAGAGATGCCATTGTTGATAATAGGCAAAAATTCTTCTTTTTTTCAGAG
       W  L  I  L  E  M  P  L  L  I  I  G  K  N  S  S  F  F  Q  S 1103

3361  CTATCTTTTGGTGGGGGCTTGCATCATGGAGTTTTTTGTCCTTGATAAAACCTTCCTTAG
       Y  L  L  V  G  A  C  I  M  E  F  F  V  L  D  K  T  F  L  S 1123

3421  TGGATCTGTGGGATTTGGGAGTGCTTTGGCTCTCAAAAACCAATTGGATGTACATAGCTC
       G  S  V  G  F  G  S  A  L  A  L  K  N  Q  L  D  V  H  S  S 1143

3481  TGTTGCTTCTTCTGGGTCTATTGCAACTCAGTCATATGCACGGAGCATACCAATTGTATG
       V  A  S  S  G  S  I  A  T  Q//S  Y  A  R  S  I  P  I  V  W 1163

3541  GGCAAAAGTAGCTCGCTATGCCAATGTCCATTCACAGGTTGAGGAGTCGAGTCATTTCAA
       A  K  V  A  R  Y  A  N  V  H  S  Q  V  E  E//S  S  H  F  N 1183

3601  TTTTTTTGAAGATGGCCTGGCGCACCTTTTAGTTAGATTGGTGGGTACTAGTGGTCTTTG
       F  F  E  D  G  L  A  H  L  L  V  R  L  V  G  T  S  G  L  C 1203

3661  TGAGACTGCTATTTTGTTTGGTTCCAGAGCTATTGCTCTGTGTGCCCATCAGATACGCAT
       E  T  A  I  L  F  G  S  R  A  I  A  L  C  A  H  Q  I  R  M 1223

3721  GTTCCCAGATCACGACCGGGTTACTGTGCATTATTTGGACAAAGCCCGGATTGCAAAGTG
       F  P  D  H  D  R  V  T  V  H  Y  L  D  K  A  R  I  A  K  C 1243

3781  CTTTCCTATGACATGGCATTGGGTAAATGCTATTGAGGAAAAAGATACGGAGGTGTGCGT
       F  P  M  T  W  H  W  V  N  A  I  E  E  K  D  T  E  V  C  V 1263

3841  TTATAGGGACGACCAATTAACGCCTCTCCCTGTCTATCCAGATTCCATTTATCTTAAGGG
       Y  R  D  D  Q  L  T  P  L  P  V  Y  P  D  S  I  Y  L  K  G 1283

3901  TGAGACACAATTACCGTCTGCAGTTAATATAAATCGAGTTTCCATAAAGAAGCGAAGATA
       E  T  Q  L  P  S  A  V  N  I  N  R  V  S  I  K  K  R  R  Y 1303

3961  TTATGAGGACGCTTCTTTGACGCCTGATGAACGATTACTGGATGGTGAAAGTCCAATTAT
       Y  E  D  A  S  L  T  P  D  E  R  L  L  D  G  E  S  P  I  I 1323

4021  ACGTTCGTGGAGTAACGTCGCTGCCTTGAGTACTAGTGTGCAAACAATTTCAAACCCTGC
       R  S  W  S  N  V  A  A  L  S  T  S  V  Q  T  I  S  N  P  A 1343

4081  ACCTGGTATTGCATACAAGCGTGATTTAAATCGCTACCTGACATCCTCGTATGCTGCGGG
       P  G  I  A  Y  K  R  D  L  N  R  Y  L  T  S  S  Y  A  A  G 1363
```

Fig. 4 (cont'd).

```
4141  GGTGCATGATTGTGGTGGTTTAATATCCATTTTGCACCAAGGACGACGCAAGGTTGTGGG
       V  H  D  C  G  G  L  I  S  I  L  H  Q  G  R  R  K  V  V  G 1383

4201  GTTGCACGTAGCAGGAACTAGAGTTGGACATCTTTTTTCGTCCACTATTAGTTTCTTGCC
       L  H  V  A  G  T  R  V  G  H  L  F  S  S  T  I  S  F  L  P 1403

4261  ACACGGCAATTTTTCCGATGTTCATTCTCAGGGAGATTTTTTTATACCTGAGGTAGGTGA
       H  G  N  F  S  D  V  H  S  Q//G  D  F  F  I  P  E  V  G  D 1423

4321  TCGAGAGGCTGGTTATGAGAAAATAGGATTTATTGATAATTCAGCCAAAGCCCACATACT
       R  E  A  G  Y  E  K  I  G  F  I  D  N  S  A  K  A  H  I  L 1443

4381  AGTACCACTACCCAATTGGGCAGGGTACCTACTAATTTTGAAACCCCTTCAACTTTTGAT
       V  P  L  P  N  W  A  G  Y  L  L  I  L  K  P  L  Q  L  L  M 1463

4441  GAGGAGGAGGAAAGAAAATTTCGTCGATGCTGGTGAAACATTTGAAATAAAAGAGCCAGC
       R  R  R  K  E  N  F  V  D  A  G  E  T  F  E  I  K  E  P  A 1483

4501  AATTCTTTCAAAAAAAGATCCTCGTCTTGAGGATCCTGATTCTTTTGACCCATTGCGGAC
       I  L  S  K  K  D  P  R  L  E  D  P  D  S  F  D  P  L  R  T 1503

4561  TGGGATGAGCAAATTTGCAAATCCTATGTCTGTACTTGATGAAGCTTTGTTGGAAGCAGT
       G  M  S  K  F  A  N  P  M  S  V  L  D  E  A  L  L  E  A  V 1523

4621  TTGTGAGGACATTTTTACCACTTGGTATGATGCCCTCCCAGCTGTTACTGACAACCAGGG
       C  E  D  I  F  T  T  W  Y  D  A  L  P  A  V  T  D  N  Q  G 1543

4681  GAATGTTTCTCGTATTTTATTAGAGAAAACTTCTTTAGATATAGCATTGAATGGAGTTCC
       N  V  S  R  I  L  L  E  K  T  S  L  D  I  A  L  N  G  V  P 1563

4741  AGGAGATGCTTATCTTGAGCCAATGAAACTTGACACTTCTGAGGGTTATCCCCATTGTGT
       G  D  A  Y  L  E  P  M  K  L  D  T  S  E  G  Y  P  H  C  V 1583

4801  CAGGCGAGGTCCTGGTGAGAGTGGAAAGCGTCGATTTGTTGAGATCGATGATGATTTCCA
       R  R  G  P  G  E  S  G  K  R  R  F  V  E  I  D  D  D  F  H 1603

4861  TTTTTCTTTGAAGCCTGATACCGATGTTTTTAAAAACTATCAGGCGCTTTCTGGGACTAT
       F  S  L  K  P  D  T  D  V  F  K  N  Y  Q  A  L  S  G  T  I 1623

4921  TTCTCAACAAGTCCCAGTCCTCAATTGCGTAGAGTGCTTGAAAGATGAATGTCTCAAGAA
       S  Q  Q  V  P  V  L  N  C  V  E  C  L  K  D  E  C  L  K  K 1643

4981  AAGGAAAGTGGCTACCCCACGCCTTTTTGATGTGATGCCTTTTGAGCACAATATTCTCTT
       R  K  V  A  T  P  R  L  F  D  V  M  P  F  E  H  N  I  L  L 1663

5041  GCGGGAATATTTTTTGAATTTTTCCGCTTTTATTCAGGCTAACCGGATTTATCTTTCCGC
       R  E  Y  F  L  N  F  S  A  F  I  Q  A  N  R  I  Y  L  S  A 1683

5101  TTGTGTTGGAACCAATCCTTATTCTCGAGAGTGGACTACACTCTATGATAGATTAGCAGA
       C  V  G  T  N  P  Y  S  R  E  W  T  T  L  Y  D  R  L  A  E 1703
```

Fig. 4 (cont'd).

```
5161  GTATTCCGATACTGGCTTGAACTGTGATTATTCCAAATTTGATGGTTTAATTTCCCATCA
       Y  S  D  T  G  L  N  C  D  Y  S  K  F  D  G  L  I  S  H  Q 1723

5221  AATATCTCGTGGATGGCTGCAACCATCAACCGTGTTTTTAGAGACGGTGAGGAAGCAAAT
       I  S  R  G  W  L  Q  P  S  T  V  F  L  E  T  V  R  K  Q  I 1743

5281  TCTGCGCGTAGGAAATCTCCTACTCATGTTCATTGGTCGCCGCTCTATTTGTGGTAGACA
       L  R  V  G  N  L  L  L  M  F  I  G  R  R  S  I  C  G  R  Q 1763

5341  AGTGTATATGGTTAGGGGCGGTATGCCTTCTGGCTGTGCTTTGACAGTCGTTATAAATAG
       V  Y  M  V  R  G  G  M  P  S  G  C  A  L  T  V  V  I  N  S 1783

5401  TATTTTTAATGAAATTTTAATTAGGTATGTTTATAGGAAGGTTACACCCGCACCTGCTTG
       I  F  N  E  I  L  I  R  Y  V  Y  R  K  V  T  P  A  P  A  C 1803

5461  TAATTTTTTTAACAAGTATGTGCGCCTCATGGTGTACGGTGACGACAATCTTCTCACCAT
       N  F  F  N  K  Y  V  R  L  M  V  Y  G  D  D  N  L  L  T  I 1823

5521  TAAAGAGGAGGTAATTCCTTTCTTTGATGGTCCAGTGATCAAGAGGGAGATGGCTAGTGT
       K  E  E  V  I  P  F  F  D  G  P  V  I  K  R  E  M  A  S  V 1843

5581  TGGTATCACCATTACGGATGGCACTGACAAGAGTTCATTGACTCTTGAGAGGAAACCTCT
       G  I  T  I  T  D  G  T  D  K  S  S  L  T  L  E  R  K  P  L 1863

5641  AGCATCTCTTGAATTTTTGAAGAGAGGTTTTAGAGTGCAGGAGAATGGGCTTGTTGTTGC
       A  S  L  E  F  L  K  R  G  F  R  V  Q  E  N  G  L  V  V  A 1883

5701  CCCTTTAGATAAGACTTCAATGTACACGCGGCTTTTTTATCTACCGCTGGCATTGATGGC
       P  L  D  K  T  S  M  Y  T  R  L  F  Y  L  P  L  A  L  M  A 1903

5761  ATTTATCCCTGGATATTTTTCGAAGGGAAATGTCAAGAGTTTTTTGGAGGAGATTGTTTT
       F  I  P  G  Y  F  S  K  G  N  V  K  S  F  L  E  E  I  V  L 1923

5821  GCACCCCAATCACCGCCGAGAATTTTACCGGGTGCGTAATTTTTATGTGAGCAAGGCCCC
       H  P  N  H  R  R  E  F  Y  R  V  R  N  F  Y  V  S  K  A  P 1943

5881  ACATTGGGGGATATCTTGCCTACATATGGCGCTGCTGTTGATTTTCATTATCGGCAGCAG
       H  W  G  I  S  C  L  H  M  A  L  L  L  I  F  I  I  G  S  R 1963

5941  ACGACCAATACCCCCTACCAGACGCAACGGCTTTTTGAACGCGTCACATGGAGGGGAACA
       R  P  I  P  P  T  R  R  N  G  F  L  N  A  S  H  G  G  E  H 1983

6001  TAAAATGATGGCTGGACAGGATTGCCAGACCAGACCATTTGGGGTAACAAGTCGTCTAGC
       K  M  M  A  G  Q  D  C  Q  T  R  P  F  G  V  T  S  R  L  A 2003

6061  TATCTTGGTAGTAGAACCCAAGTTCCAAGGGGTAGTCAACACTTTATTGTGGCGTGCGGT
       I  L  V  V  E  P  K  F  Q  G  V  V  N  T  L  L  W  R  A  V 2023

6121  TTCGTCCCTTCGTGGGGGTGAGCGTGGCATTGCATTAAAGTGGAGACTGCCTCTGGAACG
       S  S  L  R  G  G  E  R  G  I  A  L  K  W  R  L  P  L  E  R 2043
```

Fig. 4 (cont'd).

```
6181  GGTGTCTTACCTTAACTCAAACGTGGTTAATAGTTTCAGCCTTCACCACGAAACGAGCGA
       V  S  Y  L  N  S  N  V  V  N  S  F  S  L  H  H  E  T  S  D 2063

6241  CTCTTTTTTGAAGGACTTACATGAGGGATGTCACTTGTATTTAGGTTCGAGATGTACCCT
       S  F  L  K  D  L  H  E  G  C  H  L  Y  L  G  S  R  C  T  L 2083

6301  TATTACATGGGTGGTGTGCATTGCAGCAGAATTTGCTAAGGCCCAGGGGTTGAGCACATC
       I  T  W  V  V  C  I  A  A  E  F  A  K  A  Q  G  L  S  T  S 2103

6361  CAGTGTTATAGCTCTGTTTGAGGAGTATAAACCCAGAAAAGGGGGGGATATAGCTCCCCT
       S  V  I  A  L  F  E  E  Y  K  P  R  K  G  G  D  I  A  P  L 2123

6421  TTTAGCTGAGCGCTCCTATAAGAGGTTCGCTCAAAGACCAATATTTGATATGTCAAGTAT
       L  A  E  R  S  Y  K  R  F  A  Q  R  P  I  F  D  M  S  S  I 2143

6481  TAAGCAGCATCTAGCTGCTTCCTAAGCGCAGGGGGTCTCTTAGCGCCAGTTTCTAGTCCT
       K  Q  H  L  A  A  S  *                                    2150

6541  GTAGGCTAGAGGTCTTGTGGGCCTAACCCACATCCAAGAGGTTGTCATCAATTAGCATTT

6601  TACCTTCGGGTTGAAGATGTGAATGGAAGAGTGATGCCCTTCCAGACCTCTCCTTTGGAG

6661  AACCATGAGTCAACACATGGTCTTGGAGGTCACAGTTCCGATTCTAACTGTGTGCTTTTA

6721  CCAATTTTAAAGAAATGGAAGAGTAGGAGATGCTCTTGTGTGATGAGTGTGTAGATACCT

6781  TCATGTTGCTCATTACAACACATTAATGAATTCATTAATAGTTATGTGTTTGTGGTGGCA

6841  TGTTGGGTGTGTTTATCTATACATGATTTGAAAATCTCAAATGACTAGGGAGAAAGATCC

6901  TGTAGGTGTGGAAATCACCCGCTTTGTTGGAGAGCCAATTCCAACTCTTTGCTACCTTCA

6961  AGAAAGGAGATTGTACTGGTGAAATTCCAGTCCTTATATTTATTGCTTTCTAGGACTTGA

7021  GTCTTTTAGTTTTGCAATCTTGCAGAGTTGCTTTAGTAGATCTGCACGTGAAGTGCGTCA

7081  ACGTTATGGCGTAATAGTGTGTTGTGTCTCCCACACAATAAGTAATGAGACAACGCTGGG

7141  TTAGATCCCGGGAGGGTGGTTCCCTCTGACAACATTTGTGCTTTAGTAGATAAGCACCCT

7201  TTTCTTCCAGTCTTACTGAGGCAGGATATCAAAAGTAGGCTTGCAGATTATAGATTTGTG

7261  GTTAACTGATTAGACTTTGAGTAATTGTAAGAACTATCCATAAGATTATCTTGGATTGTT

7321  TAATACTCTCATGCTTATCAGCTCTTTCCATGAATACTACTGCGATACCGCTGGCGTATT

7381  CTAGTTTTAAAGACGGTATGCTGCTTCCAGCATATAAAAGCAGATATAGTAGCCATAAGC

7441  ATGATGGTTAAGCTAAATTCACCGATGAGTCGGAGGAGCCATCATGTGTACAATAGGGGG

7501  AAGCCCCTATGGCAAATTATCTGTATAGGAGCCCTTTGCTGGGGTTAAAAGCTTAAGGTT
```

```
7561  TAGTGTAACACAACATTGGGTGTACTCAAGAGCGTGTGGGGTGGCACCCACGTGCTTGGA

7621  TGAGGTCCGGAAATGAATACCGGGGGATAATTAATCCCAGCTCAGGCACTAAGCTGACTT

7681  TCATGGAAGTGTCCATGACGCATTTTAAGGTAGGTTTTAGACATAACCTCCCGGGATGGA

7741  AGTGATTACCATTTCGTTATTCGTTATTAGTTTCTTGCAACTATGATGAGGGGACCACAT

7801  CTTAAGCGATGTTGCTGCATTGCGTACCTATGGTCATCTGGTTAGTTGTCGTATTTTCTT

7861  TTAGCTTTTGTGGCGACAGATGAGGTTTGACTCCTTTTCCTTGACTCTTGACCTAAGTTG

7921  GACACAAAAATATGGTCTTTTGACTTTCAATAGAGTCGATGAAAATGTCTGCATCAC-poly(A)
```

Fig. 4. Nucleotide sequence of the cDNA of PRMV genomic RNA1. The predicted amino acid sequence of the large ORF of the plus sense (virion sense) RNA is shown below the nucleotide sequence. Nucleotide and amino acid positions are numbered to the left and right of the nucleotide sequence, respectively. The termination codon at the 3' end of the RNA1 ORF is marked with an asterisk (*). Binding sites for oligonucleotides used in cDNA synthesis of RNA1 are highlighted. The predicted polyprotein sequence was searched for dipeptides E/S, E/G, Q/G, Q/M, and Q/S, which are common proteinase cleavage recognition sites within como-, poty-, picorna-, and tomato ringspot nepovirus (TomRSV) polyproteins. By analogy with confirmed dipeptide sites in cowpea mosaic comovirus B component and putative sites in TomRSV RNA1, putative peptide cleavage sites of PRMV RNA1-encoded polyprotein are identified in gray.

Nepoviruses contain a high U content in their untranslated regions and PRMV RNA1 shares this characteristic (5'-UTR, 46%; 3'-UTR, 32%). These values are more similar to those reported for TomRSV (44.2% and 31.2%. respectively), than to those of subgroup I nepoviruses. Subgroup I nepoviruses TBRV, GCMV, GFLV, TRSV, as well as comovirus CPMV (Lomonosoff and Shanks, 1983), have U content ranging from 40-48% for both the 5'-UTR and the 3'-UTR (Rott *et al.*, 1991). Interestingly, the 3'-UTR U content for TomRSV (Sanfaçon, 1995) and PRMV approaches the subgroup I level if only their extreme 3'-termini are considered (3'-110 bp, 44.2% U for TomRSV; 3'-150 bp, 38.4% U for PRMV). Dias and Allen (1980) reported a ribonucleotide composition (mole percentage) for PRMV RNA1 of 23.6 (G), 24.1 (A), 30.9 (U) and 19.9 (C) and nucleotide sequence analysis of RNA1 revealed similar values: 23.8 (G), 24.6 (A), 31.3

(U), 20.3 (C). The M_r of RNA1 as calculated from the nucleotide sequence is 2.6×10^6 as estimated by PAGE (Dias and Allen, 1980).

Computer analysis of both the plus and minus strands of the genomic RNA1 nucleotide sequence identified several putative open reading frames (ORFs) including a single large ORF containing 6450 nucleotides. An initiation codon (AUG) was identified beginning at position 53 and a termination codon at position 6503. Analysis of this ORF indicated that it is capable of encoding a polypeptide of 2150 amino acids with a predicted molecular weight of 240 kD (Fig.4). Analysis of the remaining two reading frames of the plus strand and the three reading frames of the minus strand revealed ORFs of less than 400 nucleotides.

Analysis of Untranslated Regions

PRMV RNA1 5'- and 3'-UTRs are 53 and 1474 nucleotides, respectively. Computer prediction of RNA1 3'-UTR secondary structures of the 500 3'-terminal nucleotides revealed extensive secondary structure including stemloops, bulges, interior and bifurcation loops (Fig.5). Comparison of nepovirus 3'-UTR nucleotide sequences by pairwise alignment reveals a low and statistically insignificant nucleotide sequence identity in this region with a few notable exceptions as follows. Abbreviations of virus names and references to sequence numeration are identified in Appendix A.

(1) 5'-UUUCUUUU-3' octamer: This octamer was detected at positions 42, 171, and 7855 of PRMV RNA1. Serghini *et al.* (1990) found this octamer at variable distances from the poly (A) tail in the 3'-UTR of the RNA2 of GCMV, GFLV-F13, and TBRV. This sequence was also shown to be present once in the RNA2 5'-UTR of GFLV and SLRSV, respectively, twice in that of TomRSV, and four times in that of TBRV (Kreiah *et al.*, 1994). Kreiah *et al.* (1994) also reported that this sequence was present at two locations in the coding region of RRSV RNA2 (positions 2458 and 3478) but not in the untranslated regions.

(2) 5'-GAAAA(A)U-3': This sequence was first identified by Fuchs *et al.* (1989) for GFLV, TBRV, and GCMV, and occupies identical positions at the 5'-terminus of nepovirus genomic RNAs. All nepovirus RNAs whose entire sequence has been determined, except for RRSV RNA2, contain this sequence initiating within 6 nucleotides of the 5' terminus. The sequence was found in PRMV RNA1 at position 4; TomRSV RNA1 and RNA2 (nt 6); TBRV RNA1 and RNA2 (nt 3); ArMV RNA2 (nt 3); GCMV RNA1 and RNA2 (nt 4); GFLV RNA1 and RNA2 (nt3). Satellite RNA may be associated with nepovirus infection (Sanfaçon 1995). A search for the GAAAA(A)U sequence in nepovirus satellite RNA revealed that this sequence is located at the 5'-terminus of the large (>1kb) satRNA but not in that of the small (<0.5kb) satRNA: ArMV 1104 bp satRNA (lilac isolate) at position 4; CYMV 1165 bp sat RNA at nt 3; TBRV 1375 bp satRNA (nt 3) and GFLV satRNA (F13 strain) at position 4. It is noteworthy that the large satRNA molecules have predicted coding regions unlike the small satRNA. The GAAAAU sequence was also found in variable locations within the coding regions (CR) and/or the 3'-UTR of several nepoviruses including PRMV (four times in CR at positions 1259, 2365, 4339, and 4454; twice in the 3'-UTR at position 6870 and 7961), TomRSV RNA1 (six times in CR at positions 1571, 2479, 3464, 3698, 4143, and 5363), TomRSV RNA2 (twice in CR at positions 1467 and 5701), TBRV RNA1 (10 times in CR at positions 1182, 1497, 2008, 3216, 3654, 3726, 4980, 5908, and 6640), TBRV RNA2 (four times in CR at positions 603, 1686, 2335 and 4203), ArMV RNA2U (three times in CR at positions 2022, 2188, and 3584), GCMV RNA1 (9 times in CR at positions 724, 3279, 3555, 5028, 5111, 5603, 5809, 6542, and 6652), GCMV RNA2 (twice in CR at positions 944 and 4067), GFLV RNA1 (seven times in CR at positions 1722, 4218, 5543, 5690, 5813, 5961 and 7064), GFLV RNA2 (once in CR at position 3194). The subgroup II nepoviruses PRMV, BBLMV, CYMV, and CLRV are distinct from subgroup I nepoviruses in having the GAAAAU sequence located within the 3'-UTR; the GAAAAU sequence is present twice in the BBLMV 3'-UTR of both RNA1 and RNA2 (positions

2418, 2488 in RNA2 and positions 1243 and 1314 of RNA1), and once each in the 3'-UTR RNA1 and RNA2 of CLRV (positions 694 and 720, respectively). PRMV RNA1 and TomRSV RNA1 each contain this sequence in seven locations, as identified above. Although the biological function of this sequence is unknown, the frequency of its occurrence in the nepovirus genome far exceeds the random probability of its appearance ($1/4^6$=1 in 4096 chances for GAAAAU).

(3) A 17-nucleotide consensus sequence (5'-GGACACAAAAAGAUUUU-3') was identified near (but not at) the 3'-UTR of nepoviruses by Fuchs *et al.* (1989). Serghini *et al.* (1990) noted the presence of this sequence in TBRV, GCMV, and GFLV and Buckley *et al.* (1993) added ArMV to list of nepoviruses with this sequence. This sequence was not found in TomRSV, TRSV, or RRSV RNAs (Buckley *et al.*, 1993), however, a similar sequence was identified near the 3'-termini of PRMV RNA1 (14/17 nucleotides conserved) starting at position 7920, and also near the 3'-termini of BBLMV (15/17 nucleotides conserved) as reported by Bacher *et al.*(1994).

(4) 5'-AAAAGC-3' or 5'-AAAAAGC-3' immediately preceding the 3'-poly (A) tail of nepovirus genomic RNAs was first identified by Rott *et al.* (1991) for TomRSV, TBRV and GCMV. This sequence was identified in SLRSV-H RNA2 at a position 3 bases removed from the 3'-terminus and also in RRSV RNA2 commencing at position 3574 (Kreiah *et al.*, 1994). BBLMV RNA1 and RNA2 and PRMV RNA1 may now be added to the list of nepoviruses whose 3'-UTR contains this sequence (BBLMV RNA1, position 901; BBLMV RNA2, position 2076). This sequence is found in two locations in the 3'-UTR of PRMV RNA1 commencing at nucleotide 7416 and 7547, respectively, and positioned 561 and 430 nucleotides from the 3'-terminus, respectively. In the genomic RNAs of TomRSV, TBRV, GCMV, and SLRSV- H (RNA2 only) this sequence occurs at the extreme 3'-terminus; as with PRMV RNA1, the AAAAGC sequence in RRSV (RNA2 only), BBLMV RNA1 and RNA2 occurred at variable distances from the 3'-terminus (354, 1007, 1006, respectively).

Three other 3'-UTR nucleotide sequences are conserved among some nepoviruses but were not detected within the PRMV RNA1 sequence. These sequences include a 35 nucleotide region reported for TRSV, TomRSV and RRSV (Buckley *et al*, 1993); a stretch of 14 nucleotides identified in SLRSV-H RNA2, GFLV, and ArMV (Kreiah *et al.*,1994) and a stretch of 30 nucleotides shared by SLRSV-H RNA2 and TBRV (Kreiah *et al,* 1994).

The biological significance of these nucleotide consensus sequences is unknown. However, it is possible that these nucleotide sequences may be involved in polymerase recognition or packaging signal functions (Buckley *et al,* 1993).

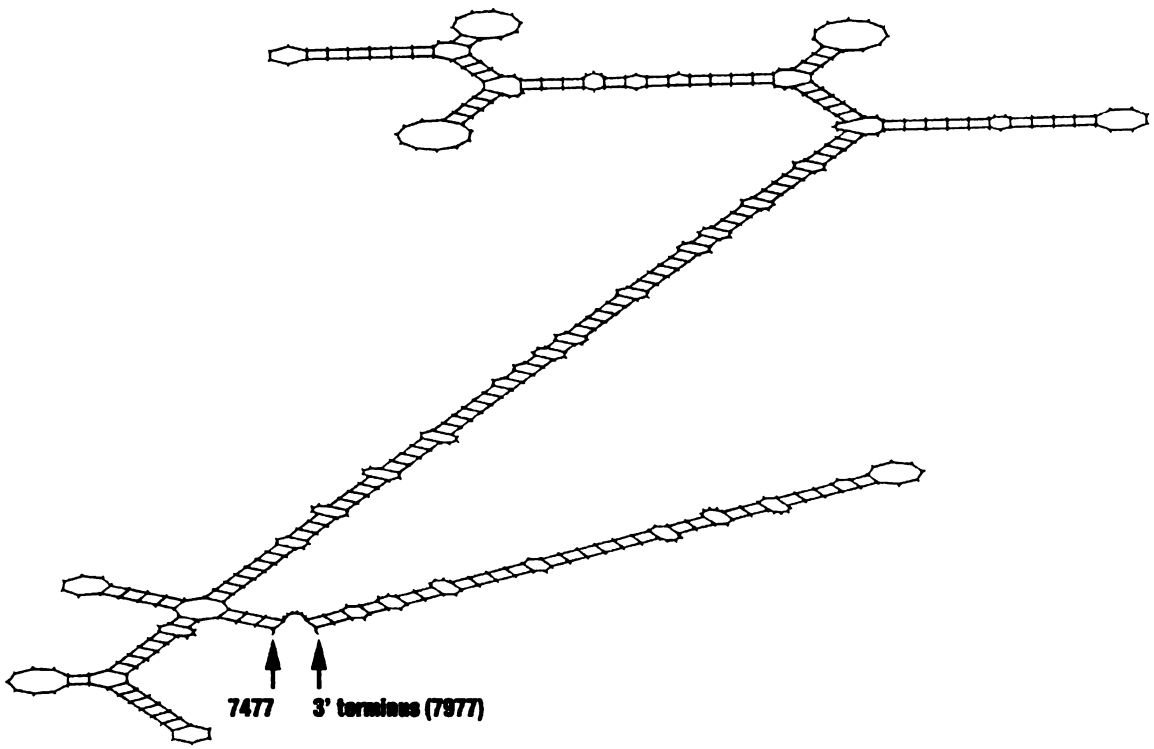


Fig. 5. Computer-predicted 3' UTR secondary structure of the 500 3'-terminal nucleotides of PRMV RNA1 commencing at nucleotide position 7477. Optimal secondary structures for the 3'-UTR of RNA 1 were predicted with the GCG FOLDRNA program (Zuker, 1989). Output files from FOLDRNA were used to plot RNA secondary structures with the GCG SQUIGGLES program.

PRMV RNA1 Polyprotein Analysis

As expected, the predicted amino acid sequence of the PRMV RNA1 polyprotein shares highest identity with that of subgroup II nepovirus TomRSV (29.8%) and to a lesser, yet significant, degree with subgroup I nepoviruses (26.7% TBRV, 27.4% GCMV, and 27.9% GFLV), CPMV B (24.9%). PRMV RNA1 amino acid identity with that of tobacco etch potyvirus (TEV), which also produces a polyprotein is insignificant (Table 1). Predicted RNA1 polyprotein sequence was examined for motifs characteristic of a proteinase cofactor (ProCF), an NTP-binding protein, a viral proteinase and an RNA-dependent RNA polymerase (RdRp). Alignment of motifs within the polyprotein of PRMV RNA1, TomRSV, GCMV, GFLV, CPMV is shown in Fig. 6. Processing of the PRMV polyprotein will be described later.

Proteinase Cofactor

A conserved amino acid sequence, F-x_{27}-W-x_{11}-L-x_{21}-L-x-E (x_n refers to the number of amino acid residues between conserved residues), is located near the N-terminus of the PRMV RNA1 polyprotein sequence beginning at amino acid residue 384 (Fig.6a). This region of conserved amino acid residues was previously identified in other nepovirus and comovirus polyprotein sequences and a proteinase cofactor function was suggested (Ritzenthaler *et al.*, 1991; Rott *et al.*, 1995). The N-terminal 32K protein of the CPMV B polyprotein contains this sequence and has been demonstrated to function as a cofactor for the CPMV 24K proteinase (Vos *et al.*, 1988; Peters *et al.*, 1992). PRMV ProCF amino acid sequence resembles that of TomRSV (24.9% identity) more so than subgroup I nepoviruses (20.9-22.8%) or CPMV B (16.4%) (Table 1). An N-terminal consensus sequence detected for TomRSV, TBRV, and GCMV (Rott *et al.*, 1995) was not found in the predicted amino acid sequence of PRMV RNA1.

NTP-Binding Protein

An amino acid motif characteristic of NTP-binding proteins is located downstream of the PRMV putative proteinase cofactor protein, beginning at amino acid

residue 696 (Fig.6b). The highly conserved 'A' and 'B' sites typical of the NTP-binding protein are G-x_4-GKS/T and DD/E, respectively (Gorbalenya and Koonin, 1989; Gorbalenya *et al.*, 1989). These two sites are thought to be important for anchoring the replication complex to the lipid membrane (Rott, *et al.,* 1995). PRMV shares the highest level of amino acid identity with subgroup II TomRSV (27.5%) and lesser identity with subgroup I nepoviruses (23.0% to 25.4%), CPMV B, and TEV, 21.9% and 14.2%, respectively (Table 1).

VPg

Nepoviruses, as with many other plant and animal viruses, contain a genome-linked protein (VPg) covalently linked to the 5'-terminus of the genomic RNA (Harrison and Barker, 1978; Matthews, 1992). Picornaviral VPg molecules are thought to play a role as a primer for the replication of both plus and minus strand RNA (Matthews, 1992). A covalently linked VPg at the 5'-ends of PRMV RNAs 1 and 2 was reported (Martelli, 1975). The VPg amino acid sequence is located between the NTP-binding protein and the proteinase for CPMV and GFLV, 4 kD and 2.9 kD, respectively (Goldbach and Rezelman, 1983; Pinck *et al.*, 1991). Tentative location and size of the VPg of TomRSV (2.7 kD), TBRV (2.3 kD), and GFLV (2.9 kD) have been reported (Rott *et al.*, 1995; Greif *et al.*, 1988; Ritzenhaler *et al.*, 1991). The putative cleavage pattern of the PRMV RNA1 polyprotein (see below) suggests a 2.9 kD VPg between amino acid positions 1154 and 1179. As with all other nepoviruses, the PRMV VPg is on RNA1 between the NTP-binding protein coding region and the proteinase. Comparison between nepovirus VPg amino acid sequences revealed no significant identity with the exception of TBRV, which compared very well with the corresponding VPg sequence in GCMV (76.5%).

Proteinase

Viruses that utilize a polyprotein expression strategy encode a proteolytic enzyme. A motif characteristic of cysteine proteinases is found in a region beginning at amino acid

residue 1219, *H*-x_{40}-*E*-x_{106}-*C*G-x_8-G-x_5-G-x-**H**-x_2-G. The residues H, E, and C (italicized) form the putative catalytic triad of the proteinase shown in Fig. 6 (Bazan and Fletterick, 1989; Gorbalenya *et al.*, 1989; Hammerle *et al.*, 1991; Dessens and Lomonosoff, 1992; Margis and Pinck, 1992). The histidine residue (bold face H), is conserved among PRMV, TomRSV, como-, poty-, and picornavirus proteinases but is replaced by a leucine in proteases of subgroup I nepoviruses (Rott *et al.*, 1995). Referring to this position as the "substrate-binding pocket" of the polio 3C proteinase, Bazan and Fletterick (1988) suggested that the His residue at this position may recognize and hydrogen-bond to the amino acid residue immediately upstream (-1 position) of the dipeptide cleavage site before cleaving the polyprotein. The cleavage site specificity of nepovirus subgroup I proteinases differs from that of picorna-, potyviruses, como-, and subgroup II nepoviruses TomRSV (Sanfaçon *et al.*, 1995), and PRMV. The difference between subgroup I and subgroup II nepovirus proteinase cleavage site specificity may be due to the replacement of the His residue with a Leu at the substrate-binding pocket (Bazan and Fletterick, 1988; Demangeat *et al.*, 1992; Ritzenthaler *et al.*,1991, Rott *et al.*, 1995). Presence of the His residue in the PRMV polyprotein suggests that the cleavage sites for maturation of the PRMV polyprotein may be similar to those of picorna-, poty-, and comoviruses (see below). Comparison between putative proteinase of PRMV and other members of the picornavirus superfamily revealed that PRMV shares a low yet significant level of amino acid sequence identity with the proteinase of subgroup II TomRSV (27.7%). Proteinase amino acid sequence of subgroup I nepoviruses and other picorna-like viruses compared less favorably (19.1-24.3%) (Table 1).

RNA-dependent RNA polymerase

A conserved GDD amino acid motif is characteristic of RNA dependent RNA polymerases (RdRp) (Argos, 1988). This motif was located in the polyprotein of PRMV RNA1 at amino acid position 1816 (Fig.4). The GDD motif as well as the flanking amino acids (1710-1821) share extensive sequence identity with other members of the picorna-

like species (Fig.6d). Sequence identity between PRMV RdRp amino acid sequence and other nepoviruses (33%-36%) was higher than that of TEV (22.6%) (Table 1). The putative active processing site of the RdRp includes a hydrophobic region of 15 amino acid residues flanking the GDD sequence (Argos, 1988) which is also found in the putative RdRp of PRMV.

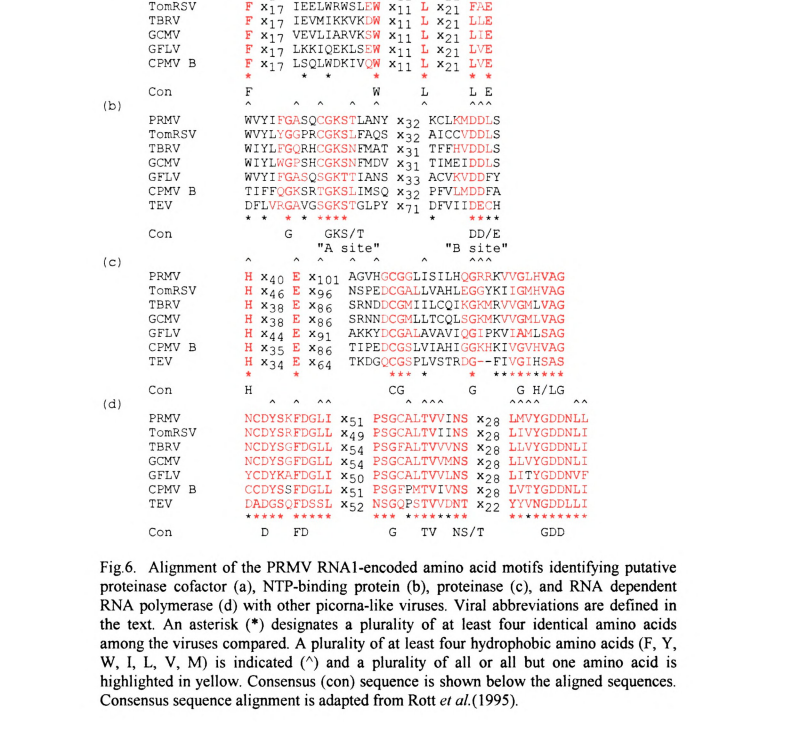


Fig.6. Alignment of the PRMV RNA1-encoded amino acid motifs identifying putative proteinase cofactor (a), NTP-binding protein (b), proteinase (c), and RNA dependent RNA polymerase (d) with other picorna-like viruses. Viral abbreviations are defined in the text. An asterisk (*) designates a plurality of at least four identical amino acids among the viruses compared. A plurality of at least four hydrophobic amino acids (F, Y, W, I, L, V, M) is indicated (^) and a plurality of all or all but one amino acid is highlighted in yellow. Consensus (con) sequence is shown below the aligned sequences. Consensus sequence alignment is adapted from Rott *et al.*(1995).

Table 1. Comparison of RNA1-encoded amino acid sequences for seven members of the picorna-virus supergroup including the nepoviruses PRMV, TomRSV, TBRV, GCMV, GFLV, and cowpea mosaic comovirus and tobacco etch potyvirus. These viruses employ a polyprotein strategy for genome expression and mature proteins are post-translationally cleaved. From the polyprotein precursor (A), mature products include a proteinase cofactor (B), a nucleotide (NTP)-binding protein (C), a genome-linked protein, or VPg[a], a proteinase (D), and a polymerase (E). Amino acid sequences of individual proteins, as well as complete polyproteins of each virus were compared using GCG Bestfit or Gap programs (Devereaux *et al.*, 1984). The amino acid location of each protein relative to the polyprotein N-terminus is shown at the right of the table. Values are expressed as percentage amino acid homology and amino acid identity to the right and to the left of the darkened cells, respectively.

A) POLY	PRMV	TomRSV	TBRV	GCMV	GFLV	CPMV	TEV	Position
PRMV		51.8	50.1	50.8	50.9	48.9	43.0*	1-2150
TomRSV	29.8		50.6	50.9	52.9	51.3	45.2	1-2197
TBRV	26.7	28.9		77.4	50.6	50.8	45.6*	1-2266
GCMV	27.4	30.2	64		51.1	52.5	45.1*	1-2253
GFLV	27.9	30.9	28.4	28.3		50.3	44.2*	1-2284
CPMV	24.9	27.6	28	28.8	27		44.2	1-1866
TEV	17.9*	20.4	20.1*	19.5*	20.3*	19.2		1-2791
B) PCF								
PRMV		47	46.7	42.6*	44.5*	45.3*		1-559
TomRSV	24.9		44.4	42.6*	43.4	47.5		1-620
TBRV	22.8	21.8		66.7	40.7*	41.5*		1-565
GCMV	18.3*	21.4*	49.4		38*	46.7*		1-460
GFLV	20.9*	24.7	16.8*	16.7*		43.5*		1-417
CPMV	16.4*	24.4	17.6*	19.4*	18.5*			1-326
TEV**								**
C) NTP								
PRMV		50.8	52.2	53.1	49.4	46	39.9*	560-1153
TomRSV	27.5		49.9	49.2	50.3	50.9	43.5*	621-1212
TBRV	23	29.1		78	51.2	49.1	42.9*	566-1211
GCMV	25.4	29.7	63.5		51.4	50.1	42.7*	461-1182
GFLV	24.3	27.3	27.1	27.5		43.9	43.6*	417-1217
CPMV	21.9	23.5	24.2	23.2	20.2		39.5*	327-919
TEV	14.2*	16.3*	15.3*	18*	20*	15.8*		1163-1796
D) PRO								
PRMV		47	44.3	48.8	46.5	46	43.8*	1179-1413
TomRSV	27.7		48.3	47.7	51.2	51.5	46.9*	1237-1465
TBRV	22.9	23.4		84	48.3	49	44.1*	1233-1440
GCMV	22.9	25.1	71.4		49.3	49.2	44.9*	1219-1428
GFLV	24.3	31.6	23.2	25.9		46.4	45.3*	1241-1460
CPMV	20	26.3	21.4	23.4	24.2		51.2*	948-1155
TEV	19.1*	23.3*	18.6*	18.4*	18.9*	17.4*		1850-2279

Table 1 (cont'd).

E) POL	PRMV	TomRSV	TBRV	GCMV	GFLV	CPMV	TEV	Position
PRMV		56.8	52.4	53.3	58.5	54.3	46.4*	1414-2150
TomRSV	36		56.6	58.8	60.4	53.4	44.5	1466-2197
TBRV	33.3	37.8		81.4	56.9	56.1	47.3	1441-2266
GCMV	33.6	39.3	70.7		57.2	59.3	43.5*	1429-2253
GFLV	36.3	39	36.7	36.8		57.5	42.8*	1461-2284
CPMV	33.6	33.3	35.9	39.1	36.5		43.8	1156-1866
TEV	22.6*	22	24.9	20.8*	21.2*	21.1		2280-2791

[a] Comparisons between VPg amino acid sequences are not shown; only TBRV and GCMV VPg showed significant amino acid identity (76.5%).
* insignificant sequence homology or identity. Significance was assessed by shuffling one of the pair of sequences being compared repeatedly (10 times) and aligning it with the non-randomized sequence using GCG (GAP or BESTFIT with randomizing parameter, Devereaux *et al.*, 1984). Values were deemed significant if they exceeded the mean randomized comparison plus 3 standard deviations (Doolittle, 1981).
** not present

Processing of the PRMV RNA1 Polyprotein

As mentioned in the proteinase section, the histidine residue is conserved in the putative active site of the proteinase of como-, poty-, picornaviruses and the subgroup II nepoviruses PRMV and TomRSV but is replaced by a leucine in proteinases of subgroup I nepoviruses. This suggests that the dipeptide cleavage site specificity of PRMV proteinase is more similar to the aforementioned viruses than subgroup I viruses TBRV, GCMV, and GFLV (Hans and Sanfaçon, 1995; Grief *et al.*, 1988; LeGall *et al.*, 1989; Ritzenhaler *et al.*, 1991; Margis *et al.*, 1991, 1994; Hemmer *et al.*, 1995). The known dipeptide cleavage sites for maturation of polyproteins of como-, poty- and picornaviruses is E/G, E/S, Q/G, Q/S, and Q/M. (Wellink *et al.*, 1986; Wellink and Van Kammen, 1986; Hellen *et al.*, 1989; Palmenberg, 1990). A search for these sites in the

PRMV RNA1-encoded polyprotein and subsequent alignment of the proposed cleavage products to TomRSV and CPMV B revealed a conservation in both order and size of the putative translation products. Until direct protein sequencing of the mature PRMV proteins is accomplished, assignment of cleavage sites are tentative. The proposed genomic strategy for PRMV RNA1 is shown (Fig.7).

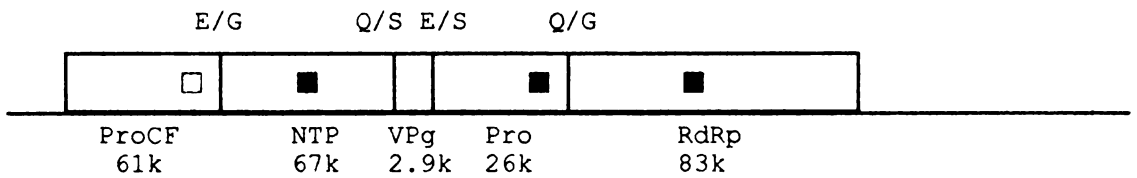


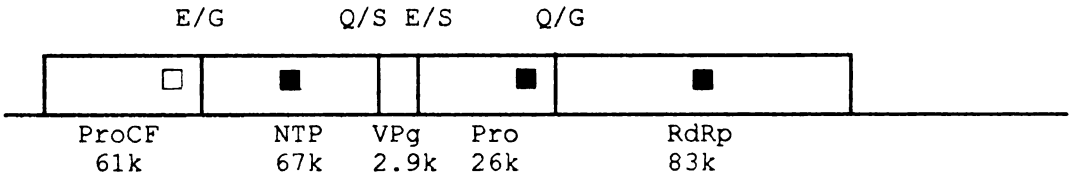
Fig. 7. Genomic organization of PRMV RNA1. Large rectangles represent the polyprotein expressed from the major open reading frame of plus sense RNA1. Conserved amino acid motifs are indicated by uniquely colored boxes positioned at the relative location of the motif within each protein. Putative proteins encoded by PRMV RNA1 are abbreviated as follows: putative proteinase cofactor (ProCF); NTP-binding protein (NTP); proteinase (Pro); RNA-dependent RNA polymerase (RdRp) and genome-linked protein (VPg). Known cleavage sites of CPMV B and putative sites in TomRSV RNA1 aided in identification of potential cleavage sites in PRMV. Picorna-like proteinase recognition sites are E/G, E/S, Q/G, Q/S, and Q/M.

Putative RNA1 cleavage sites include an E/G dipeptide at positions 559-560 between the putative N-terminal proteinase cofactor and NTP-binding protein. A Q/S and an E/S site are located between the NTP-binding and putative proteinase at amino acid positions 1153-1154 and 1178-1179, respectively. The region between these two cleavage sites is 25 amino acids in length, comparing to 24 or 27 amino acids for TomRSV (two sites are proposed by Rott *et al.* (1995) for potential cleavage at the C-terminus of TomRSV VPg) and 28 amino acids corresponding to the CPMV B-encoded VPg. A possible cleavage site between the PRMV-encoded RdRp and proteinase is Q/G

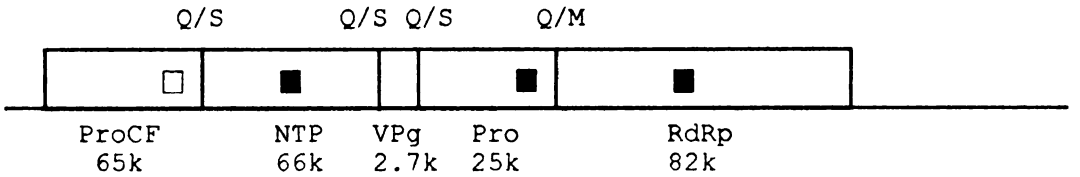
acid position 1413-1414, which aligns well with the Q/G site in CPMV and the Q/M site in TomRSV. Corresponding sites in TBRV, GCMV and GFLV are Q/S, Q/I, and R/G, respectively.

The proposed cleavage sites for the PRMV RNA1-encoded polyprotein result in mature polypeptide cleavage products which are comparable in size with those proposed for other sequenced nepoviruses (Fig.8). Sequentially from the N-terminus of the polyprotein, putative PRMV protein products from the RNA1 polyprotein include a 61 kD proteinase cofactor; 67 kD NTP-binding protein; 2.9 kD VPg; 26kD proteinase and an 83 kD polymerase. Comparable putative proteins in the TomRSV RNA1-encoded polyprotein are 65 kD, 66kD, 2.7kD, 25 kD, and 82 kD in size, respectively (Rott *et al.*, 1995). A comparison of the genomic strategies of PRMV RNA1 with TomRSV, GFLV and CPMV, including the location and sizes of mature polypeptides is shown in Fig.8. Putative cleavage sites for PRMV, TomRSV and GFLV are also included. The known location and identity of dipeptide cleavage sites for CPMV are shown.

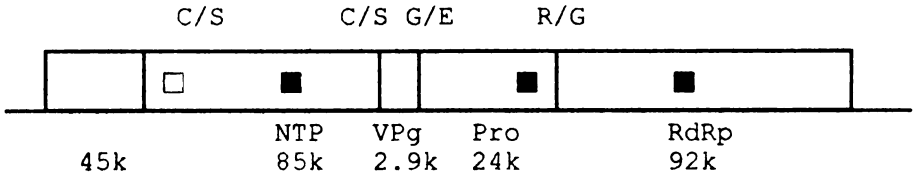
PRMV RNA1



TomRSV RNA1



GFLV RNA1



CPMV B RNA

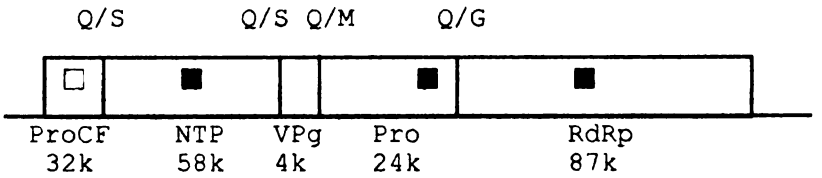


Fig.8. Comparison of the genomic strategy of PRMV RNA1 with other members of the picornavirus superfamily. Large rectangles represent polyproteins expressed from the major open reading frame of the plus sense RNA and noncoding sequences are represented by a horizontal line. Conserved amino acid motifs are indicated by uniquely colored boxes positioned in the relative location of the motif within each protein. Putative proteins encoded by each genome are abbreviated as follows: putative proteinase cofactor (ProCF); NTP-binding protein (NTP); proteinase (Pro); RNA-dependent RNA polymerase (RdRp); and the genome-linked protein (VPg). The known dipeptide cleavage sites of CPMV B are shown and were used in identifying potential cleavage sites in TomRSV and in PRMV. Proteinase recognition sites common to poty-, como-, and picornaviruses include E/G, E/S, Q/G, Q/S, and Q/M.

Summary and Conclusions

The complete nucleotide sequence of PRMV RNA1 isolated from Michigan 'Concord' grapevine has been determined. cDNA clones representing 99.6% of RNA1 were obtained and the cDNA sequence, as well as direct RNA sequencing analysis of the remaining RNA sequence revealed an RNA species of 7977 nucleotides, excluding a 3'-polyadenylated tail of variable length. RNA1 3'-and 5'-untranslated regions are 52 and 1474 nucleotides, respectively. The nucleotide sequence of the genomic RNA1 and RNA2 of all nepoviruses contains a single long open reading frame ORF (Sanfaçon, 1995) and is also found in PRMV RNA1. Analysis revealed a single ORF of 6450 nucleotides initiating at nucleotide 53 and terminating at nucleotide 6503. This coding region encodes 2150 amino acid residues with a coding capacity of 240 kD.

Analysis of the untranslated regions of RNA1 revealed several features common among nepoviruses. PRMV is similar to other subgroup II nepoviruses TomRSV, CLRV, and BBLMV in having a very long 3'-UTR sequence. The 3'-UTR for these viruses are 1474, 1543, 1500, and 1392 nucleotides, respectively, contrasted with the subgroup I nepoviruses 3'-UTR which ranges from 198 nucleotides (ArMV) to 301 nucleotides (TBRV) to 583 nucleotides (TRSV). Evidence demonstrating the precise role of the long 3'-UTR in subgroup II nepoviruses is lacking. Nucleotide sequence identity between nepovirus 5'- and 3'- UTRs is limited to a few short consensus sequences and it is likely that some of these conserved sequences have biological significance such as replicase recognition or packaging signals as they occur at a far higher frequency than that predicted by random probability.

Comparison of the predicted PRMV RNA1 amino acid sequence to that of other nepoviruses revealed, as expected, that PRMV was most similar to subgroup II TomRSV, and less so with subgroup I nepoviruses such as TBRV, GFLV, GCMV. Therefore, bipartite subdivision of the nepovirus group by Francki *et al.* (1985) is supported by direct sequence comparison. Analysis of the predicted amino acid sequence of PRMV

RNA1 uncovered motifs characteristic of a replicase, a proteinase, an NTP-binding protein and a proteinase cofactor. The order and identity of these motifs within the PRMV polyprotein aligns well with that of subgroup I nepoviruses and CPMV B, and especially with that of subgroup II TomRSV (Fig.8). In addition, putative mature proteins of PRMV share the highest level of amino acid sequence identity with their counterparts in TomRSV (Table 1).

The cleavage site specificity of subgroup I proteinases differs from that of subgroup II nepoviruses, perhaps due to a replacement of a Leu residue with a His in the putative substrate-binding pockets of subgroup II proteinases. Como-, poty-, and animal picornaviruses, whose well-characterized cysteine proteinases resemble subgroup II nepovirus proteinases, including those of TomRSV and PRMV, in having a His at this position, cleave the dipeptides Glu/Gly, Glu/Ser, Gln/Gly, Gln/Ser, and Gln/Met. This cleavage pattern is distinct from that of subgroup I nepovirus proteinases, whose targets include Cys/Ala, Cys/Ser, Gly/Glu, Arg/Ala, Arg/Gly and.Lys/Ala (Sanfaçon 1995). It is likely, therefore, that the PRMV proteinase is more related to that of subgroup II TomRSV and other picorna-like viruses than to that of nepovirus subgroup I.

Sequence analysis of RNA1 confirms PRMV as a member of the subgroup of nepoviruses including TomRSV and further distinguishes the nepovirus subgroups. Portions of the RNA1 sequence, in conjunction with classical breeding and selection for resistant cultivars, may provide 'Concord' vineyards with PRMV resistance. Obtaining genetically engineered grapevines expressing a portion of the PRMV RNA1 genome will further our attempt to understand the mechanism of pathogen-derived resistance.

APPENDICES

APPENDIX A

Appendix A

Species of the Nepovirus genus (family *Comoviridae*) both confirmed (A) and tentative (B). Viral names and abbreviations are compiled from Goldbach *et. al* (1995). Nucleotide sequence information, including Genebank accession numbers, was assembled using the UW GCG Stringsearch program (Devereaux, J., Haeberli, P., and Smithies, O., 1984). Nepovirus subgrouping was adapted from Francki *et al.* (1985).

A. Confirmed Nepovirus Species 1. Subgroup I

Virus	Abbreviation	Sequence Description	GCG Access Code
Arabis mosaic virus	**ArMV (4 sequences)**	**RNA2 3' terminal region (2406 bp) 5/92**	**D10086**
		capsid protein gene (1515 bp) 1/91	**X55460**
		polyprotein P2-U (3852 bp) 3/95	**X81814**
		polyprotein P2-L (3712 bp) 3/95	**X81815**
Arabis mosaic virus satellite RNA	**sArMV (2 sequences)**	**complete satellite genome (300 bp) 7/89**	**M21212**
		satellite RNA (1104 bp) 3/91	**D00664**
Arracacha virus A	**AVA**	**na**	**na**
Artichoke Italian latent virus	**AILV**	**na**	**na**
Cassava American latent virus	**CsALV**	**na**	**na**
Cacao necrosis virus	**CNV**	**na**	**na**
Crimson clover latent virus	**CCLV**	**na**	**na**
Cycas necrotic stunt virus	**CNSV**	**na**	**na**
Grapevine chrome mosaic virus	**GCMV (2 sequences)**	**RNA2 (4441 bp) 9/93**	**X15163**
		RNA1 (7212 bp) 9/93	**X15346**
Grapevine fanleaf virus	**GFLV (6 sequences)**	**RNA1 (7342 bp) 4/94**	**D00915**
		RNA2 (3774 bp) 9/93	**X16907**
		VPg (84 bp) 1/94	**S38553**
		RNA2 deletion mutant (501 bp) 10/94	**U11770**
		capsid protein, partial (1515 bp) 10/94	**U11768**
		capsid protein, complete (2305 bp) 10/91	**X60775**
Grapevine fanleaf virus satellite RNA	**sGFLV (1 sequence)**	**complete sequence (1114 bp) 3/91**	**D00442**

Appendix A (cont'd).

Grapevine Tunisian ringspot virus	**GTRSV**	**na**	**na**
Mulberry ringspot virus	**MRV**	**na**	**na**
Olive latent ringspot virus	**OLRSV (2 sequences)**	**RNA3 (2438 bp) 3/95**	**X76993**
		RNA4 (2078 bp) 3/95	**X77115**
Potato black ringspot virus	**PBRSV**	**na**	**na**
Raspberry ringspot virus	**RRSV (1 sequence)**	**RNA2 (3928 bp) 2/93**	**S46011**
Tobacco ringspot virus	**TRSV (4 sequences)**	**self-cleavage consensus region (52 bp) 10/94**	**A13898**
		mutant D-51 self-cleavage consensus region (63 bp) 10/94	**A13899**
		complete capsid protein gene (2018 bp) 9/94	**L09205**
		mRNA (360 bp) 12/90	**M17439**
Tobacco ringspot virus satellite RNA	**sTRSV (14 sequences)**	**satellite RNA (359 bp) 7/89**	**M14879**
		satellite autolytic junction (71 bp) 12/90	**M31515**
		various satellite genomic RNA mutations	**S63883; S63888; S63895; S63896; S63897; S63901; S63903; S63904; S63907; S63908; S63910; S63911**
Tomato black ring virus	**TBRV (8 sequences)**	**RNA1 (7362 bp) 10/94**	**D00322**
		RNA2 strain S (4662 bp) 9/93	**X04062**
		RNA1 strain C 3' terminus (166 bp) 7/89	**X05304**
		RNA2 strain C 3' terminus (151 bp) 7/89	**X05305**
		RNA 2 strain G 3' terminus (100 bp) 7/89	**X05306**
		RNA1 strain A 3' terminus (163 bp) 7/89	**X05307**
		RNA2 strain A 3' terminus (127 bp) 11/87	**X05308**
		RNA2 strain ED (4618 bp) 8/94	**X80831**
Tomato black ring virus satellite RNA	**sTBRV (7 sequences)**	**sRNA (1375 bp) 9/93**	**X00978**
		isolate C sRNA (1374 bp) 2/91; 7/91	**D00142; X05689**

Appendix A (cont'd).

Tomato black ring satellite RNA (cont'd)		**isolate E sRNA (1372 bp) 2/91; 7/91**	**D00143; X05688**
		isolate L sRNA (1376 bp) 2/91; 7/91	**D00144; X05687**

Subgroup II Nepoviruses

Artichoke yellow ringspot virus	**AYRV**	**na**	**na**
Blueberry leaf mottle virus	**BBLMV (2 sequences)**	**RNA2 3' terminus (3082 bp) 5/95**	**U20621**
		RNA1 3' terminus (1908 bp) 5/95	**U20622**
Cassava green mottle virus	**CGMV**	**na**	**na**
Cherry leaf roll virus	**CLRV (6 sequences)**	**RNA2 birch isolate I2 3' terminus (1920 bp) 7/91**	**S63537**
		RNA1 3'terminus (1743 bp) 1/94	**S84124**
		RNA2 3'terminus (1805 bp) 1/94	**S84125**
		R25 3' terminus (1182 bp) 1/94	**S84126**
		RNA2 3' terminus (1565 bp) 5/95	**U24694**
		genomic RNA walnut isolate (1588 bp) 11/94	**Z34265**
Chicory yellow mottle virus	**CYMV**	**na**	**na**
Chicory yellow mottle virus satellite RNA	**sCYMV (4 sequences)**	**T isolate small satellite RNA (457 bp) 6/91**	**D00685**
		C isolate large satellite RNA (1165 bp) 6/91	**D00686**
		sRNA S1 (457 bp) 7/94	**D00721**
	sCYMV	**sRNA L1 (1145 bp) 12/91**	**D00722**
Grapevine Bulgarian latent virus	**GBLV**	**na**	**na**
Hibiscus latent ringspot virus	**HLRV**	**na**	**na**
Lucerne Australian latent virus	**LALV**	**na**	**na**
Myrolaban latent ringspot virus	**MLRSV**	**na**	**na**

Appendix A (cont'd).

Peach rosette mosaic virus	**PRMV (1 sequence)**	**RNA1 (see Fig.4 above)**	**na**
Potato virus U	**PVU**	**na**	**na**
Tomato ringspot virus	**TomRSV (5 sequences)**	**RNA1 (8114 bp) 9/94**	**L19655**
		RNA2 (7273 bp) 4/94	**D12477**
		RNA1 3' non-coding region (1546 bp) 2/90	**M27936**
		RNA2 3' non-coding region (1550 bp) 2/90	**M27935**
		RNA1 5' terminus (1140 bp) 12/91	**M73822**

B. Tentative Nepovirus Species

Arracacha virus B	**AVB**	**na**	**na**
Artichoke vein banding virus	**AVBV**	**na**	**na**
Cherry rasp leaf virus	**CRLV**	**na**	**na**
Lucerne Australian symptomless virus	**LASV**	**na**	**na**
Rubus Chinese seed-borne virus	**RCSV**	**na**	**na**
Satsuma dwarf virus	**SDV**	**na**	**na**
Strawberry latent ringspot virus	**SLRSV (2 sequences)**	**RNA2 (3824 bp) 1/95**	**X77466**
		43K/27K capsid proteins (2424 bp) 2/95	**X75165**
Strawberry latent ringspot satellite RNA	**sSLRSV (1 sequence)**	**sRNA encoding 36K protein (1118 bp) 6/93**	**X69826**
Tomato top necrosis virus	**TTNV**	**na**	**na**

APPENDIX B

Appendix B

Nucleotide sequence of the oligonucleotide primers used for cloning and sequencing PRMV RNA1.

Primer	Nucleotide Sequence and binding location on PRMV RNA1	cDNA clone
Oligo $(dT)_{12-18}$	**d(TTT...$_{12-18}$)**	**5-9 (3.2 kb)**
	poly (A) tail	
RA42	**d(AAATCATCATCGATCTCAAC)**	**2.1 (1.2 kb)**
	4838-4857	
RA50	**d(ACCACTAGTACCCACCAATC)**	**50-3.9 (1.4 kb)**
	3636-3655	
RA52	**d(CCTTCTGTGGGATCCTTGGAGTAGAC)**	**52-4.2 (1.1 kb)**
	2222-2247	
RA68	**d(GAAATATAGGAACCAACC)**	**68-2.90 (0.9 kb)**
	1099-1116	
RA70	**d(CCCATTGCAACCGCCAGAGCTAC)**	**70.20 (0.15 kb)**
	212-234	
RA75	**d(GTCCAAATATTCCATCAC)**	**RNA sequencing**
	50-67	

Appendix C

Appendix C

Cloning and Sequencing of PRMV RNA2

Introduction

Peach rosette mosaic virus (PRMV), a nepovirus, was first recognized as the cause of a disease of peaches (*Prunus persica* L.) in Michigan in 1917 (Klos *et al.*, 1976). Nepoviruses are considered a genus within the picornavirus-like supergroup of plant viruses which includes the potyviruses, comoviruses, and picornaviruses (Goldbach, *et al.,* 1987). Common features within this supergroup include genomic structure and organization, as well as nucleotide and amino acid sequence similarity.

Most nepoviruses, including PRMV consist of three distinct particle types: a top (T) component consisting of empty polyhedral capsid proteins; and a middle component (M) and a bottom (B) component. M and B components each contain identical capsid proteins plus single molecules of RNA2 and RNA1, respectively (Martelli and Taylor, 1990). Nepoviruses have been subdivided based on their respective RN2 length. In subgroup I, RNA2 is less than 5.4 kb. In subgroup II, which includes PRMV and TomRSV, RNA2 is greater than 5.4kb (Francki *et al.,* 1985). Nepoviruses have a bipartite genome with a polyprotein expression strategy (Matthew, 1991)

Many nepovirus subgroup I and II genomic RNAs have been sequenced completely (see Appendix A) and genomic sequence analysis has provided further criteria to separate the two subgroups. In subgroup II TomRSV, PRMV, BBLMV, CLRV, for example, the 3'-untranslated region (UTR) is 1.4 kb or greater compared to 0.5 kb or less in subgroup I GFLV, GCMV, TBRV. Polyprotein processing in subgroup I and II also appears to differ: the subgroup II RNA1-encoded proteolytic enzyme functions more similar to that of como-, poty-, and animal picorna-like viruses than to that of nepovirus

subgroup I. Evidence demonstrates that a single amino acid substitution (Leu to His) in putative subgroup II proteinase substrate-binding pockets may be responsible for differences in proteolytic activity (Bazan and Fletterick, 1988).

PRMV RNA1 was sequenced in order to confirm the subgroup II status of PRMV (chapter 2). RNA1 consists of 7977 nucleotides not including its 3'-poly (A) tail. The 5'- and 3'- untranslated regions consist of 52 and 1474 nucleotides, respectively. Analysis of the PRMV RNA1 nucleotide sequence unveiled a single long open reading frame of 6450 nucleotides capable of encoding a 240 kD polyprotein. Motifs characteristic of a replicase, a proteinase, an NTP-binding protein and a proteinase cofactor were detected in RNA1 putative amino acid sequence and the order and identity of these putative proteins are consistent with other nepoviruses. RNA1 genomic characteristics confirm PRMV subgroup II status.

The partial nucleotide sequence of PRMV RNA2 was analyzed for nepovirus features including the presence of extensive sequence homology between the 3' UTRs of RNA1 and RNA2. RNA2 nucleotide sequence analysis was also performed to confirm the subgroup II characteristic 3'-UTR (greater than 1.4 kb). Nucleotide sequence analysis of BBLMV, CLRV, and TomRSV, respectively, indicates that the 3'-terminal 1.4 kb of the 3'-UTR are nearly identical (e.g. TomRSV RNA1 and RNA2 3'-1533 nucleotides differ at only 3 positions) (Bacher *et al.*, 1994; Scott *et al.*, 1992; Rott *et al.*, 1991 and Sanfaçon, 1995). Although subgroup I nepoviruses show extensive nucleotide sequence homology among their 3'-UTRs (80-100%), the extent of homology is limited to a few hundred nucleotides (Sanfaçon, 1995).

Sequence analysis of the 3'-terminal region of RNA2 adds to our understanding of the PRMV genome and further confirms PRMV subgroup II status.

RNA2 cDNA Cloning and Sequence Analysis

The Michigan 'Concord' grapevine PRMV isolate used in RNA1 cDNA synthesis and sequencing was also used for RNA2 cDNA synthesis and sequencing. Materials and

methods including virion and RNA purification, cDNA synthesis, cloning of cDNA into KS- EcoRV site, exo III deletion of cDNA and nucleotide sequencing of cDNA and exo III subclones are exactly as described in Chapter 2. Pooled PRMV RNA1 and 2 were used as template for cDNA synthesis.

Results and Discussion

A cDNA clone, 4-2.2, contained a cDNA insert of approximately 4000 nucleotides, as estimated electrophoretically. Nucleotide sequence analysis of 4-2.2 detected two tandem-ligated cDNA fragments whose respective nucleotide sequences were nearly identical. The 3'-cDNA fragment contained 1501 unique nucleotides excluding a 3'-poly (A) tail of 41 residues; a 42-residue 3'-poly(A) tail of the upstream portion (1220 unique nucleotides) separated the two fragments of the hybrid RNA. The cDNA nucleotide sequences of the two distinct fragments are compared with the 3'-UTR sequence presented in chapter two (Fig.9).

Nucleotide sequence comparison to RNA1 was used to determine the RNA origin of the two cDNA fragments. GCG BESTFIT and GAP analysis (Devereaux *et al.* 1984) were utilized with default gap and length weights of 5.0 and 0.3, respectively. The entire 1220 nucleotides of the upstream fragment of 4-2.2 shared perfect identity with the corresponding 3'-terminal nucleotides of PRMV RNA1. Upstream and downstream cDNA fragments were 89.6% identical. Poly (A) tails were not included in the alignment. Therefore, the origin of the unique downstream segment of cDNA clone 4-2.2 was assigned to PRMV RNA2. The three potential ORFs of the 3'-terminal 1501 nucleotides were analyzed for their coding capacity; the longest reading frame consisted of 273 nucleotides with a coding capacity of 91 amino acids (RNA2 cDNA positions 59-332).

```
     1                                                            50
(a)  CTCAAAGACC AATATTTGAT ATGTCAAGTA TTAAGCAGCA TCTAGCTGCT
(b)  .......... .......... .......... .......... ..........
(c)  .......... .......... .........G TCTGCTGACC TAACGGTCTT

     51                                                          100
(a)  TCCTAAGCGC AGGGGGTCTC TTAGCGCCAG TTTCTAGTCC TGTAGGCTAG
(b)  .......... .......... .......... .......... ..........
(c)  TACCCAGCGC AGGGGGTCTT TTAGCGCCAG TTTCTAGTCC TGTAGGCT.G

     101                                                         150
(a)  AGGTCTTGTG GGCCTAACCC ACATCCAAGA GGTTGTCATC AATTAGCATT
(b)  .......... .......... .......... .......... ..........
(c)  AGGCTTTGTG GGCCTAACCC ACTTCCAAGA GGAAGTCATC AATTAACATT

     151                                                         200
(a)  TTACCTTCGG GTTGAAGATG TGAATGGAAG AGTGATGCCC TTCCAGACCT
(b)  .......... .......... .......... .......... ..........
(c)  TTACCTACGG GTTAAAGATG AGAATGGAAG AGTAATGCCC TTCCAGACCT

     201                                                         250
(a)  CTCCTTTGGA GAACCATGAG TCAACACA.. .TGGTCTTGG AGGTCACAGT
(b)  .......... .......... .......... .......... ..........
(c)  CTCCTTTGGA GAACCATGAG TCCACACATG GTGGTCTTGG AGGTCACAGT

     251                                                         300
(a)  TCCGATTCTA ACTGTGTGCT TTTACCAATT TTAAAG.AAA TGGAAGAGTA
(b)  .......... .......... .......... .......... ..........
(c)  TCCGATTCTA ACTGTGTGCT TTTACCAATT TTAAAGAAAA TGGAGGAGCA

     301                                                         350
(a)  GGAGATGCTC TTGTGTGATG AGTGTGTAGA TACCTTCATG TTGCTCATTA
(b)  .......... ..GTGTGATG AGTGTGTAGA TACCTTCATG TTGCTCATTA
(c)  GGAGATGCTT TTACGTGGTG CGCATGTAAA TACCTTCATG TTGCTCATTA

     351                                                         400
(a)  CAACACATTA ATGAATTCAT TAATAGTTAT GTGTTTGTGG TGGCATGTTG
(b)  CAACACATTA ATGAATTCAT TAATAGTTAT GTGTTTGTGG TGGCATGTTG
(c)  CAACACACTT ATGAATTCAT TAATAGTGGT GTGTTTGTGG TGGCATGTTG

     401                                                         450
(a)  GGTGTGTTTA TCTATACATG ATTTGAAAAT CTCAAATGAC TAGGGAGAAA
(b)  GGTGTGTTTA TCTATACATG ATTTGAAAAT CTCAAATGAC TAGGGAGAAA
(c)  GGTGTGTTTA TTTACATATA ACCCAAAAAT CTCGAATGAC TAGGGAGAAA

     451                                                         500
(a)  GATCCTGTAG GTGTGGAAAT CACCCGCTTT GTTGGAGAGC CAATTCCAAC
(b)  GATCCTGTAG GTGTGGAAAT CACCCGCTTT GTTGGAGAGC CAATTCCAAC
(c)  GATCCCTTGG GTGTGTGATT CACCTGCTTA ATTGGAGAGC CAATTCCAAT

     501                                                         550
(a)  T.CTTTGCTA CCTTCAAGAA AGGAGATTGT ACTGGTGAAA TTCCAGTCCT
(b)  T.CTTTGCTA CCTTCAAGAA AGGAGATTGT ACTGGTGAAA TTCCAGTCCT
(c)  TGCTTTGCTA CCTTCAAGAA AGGAGATTAT .CTGGTGAAA TTCCAGATTT
```

Figure 9 (cont'd).

```
     551                                                        600
(a)  TATATTTATT GCTTTCTAGG ACTTGAGTCT TTTAGTTTTG CAATCTTGCA
(b)  TATATTTATT GCTTTCTAGG ACTTGAGTCT TTTAGTTTTG CAATCTTGCA
(c)  TATGTTTATT GC.TTCTAGG ATTTGAGTCT TTTAGTTTTG CAATCCTGCA

     601                                                        650
(a)  GAGTTGCTTT AGTAGATCTG CACGTGAAGT GCGTCAACGT TATGGCGTAA
(b)  GAGTTGCTTT AGTAGATCTG CACGTGAAGT GCGTCAACGT TATGGCGTAA
(c)  GAGTTGCGTT AGTAAATCTG TACGTGAAGT ACGTCAACGT AATGACGTAA

     651                                                        700
(a)  TAGTGTGTTG TGTCTCCCAC ACAATAAGTA ATGAGACAAC GCTGGGTTAG
(b)  TAGTGTGTTG TGTCTCCCAC ACAATAAGTA ATGAGACAAC GCTGGGTTAG
(c)  TAGTGTGTTG TGTCTCCCAC ACTATAAGTA ATGAGACAAC GCTGGGTTAG

     701                                                        750
(a)  ATCCC.GGGA GGGTGGTTCC CTCTGACAAC ATTTGTGCTT TAGTAGATAA
(b)  ATCCC.GGGA GGGTGGTTCC CTCTGACAAC ATTTGTGCTT TAGTAGATAA
(c)  ATCCCGGGGA GGGTGGTTCC TCCTGTGAAC ATTTGTGCTT TAGTAAATAA

     751                                                        800
(a)  GCACCCTTTT CTTCCAGTCT TACTGAGGCA GGATATCAAA AGTAGGCTTG
(b)  GCACCCTTTT CTTCCAGTCT TACTGAGGCA GGATATCAAA AGTAGGCTTG
(c)  GCACCCTTTT CTTCCAGTCT TACTGAGACG GAATATCAAA AGTAGGCTTG

     801                                                        850
(a)  CAGATTATAG ATTTGTGGTT AACTGATTAG ACTTTGAGTA ATTGTAAGAA
(b)  CAGATTATAG ATTTGTGGTT AACTGATTAG ACTTTGAGTA ATTGTAAGAA
(c)  CAGATAACAG ATTTATGGTT AACTGATCTA GATTTGAGCA ATTGTAAGAA

     851                                                        900
(a)  CTATCCATAA GATTATCTTG GATTGTTTAA TACTCTCATG CTTATCAGCT
(b)  CTATCCATAA GATTATCTTG GATTGTTTAA TACTCTCATG CTTATCAGCT
(c)  CTGTCCATAA .TTTGCTTTG GACTGTTTAA TGCTCTCATC TTTACCAGCT

     901                                                        950
(a)  CTTTCCATGA ATACTACTGC GATACCGCTG GCGTATTCTA GTTTTAAAGA
(b)  CTTTCCATGA ATACTACTGC GATACCGCTG GCGTATTCTA GTTTTAAAGA
(c)  CTTTCCATGA ATACTGCTGC GATACCG.TG GCGTATTCTA GTTTTAAAGA

     951                                                       1000
(a)  CGGTATGCTG CTTCCAGCAT ATAAAAGCAG ATATAGTAGC CATAAGCATG
(b)  CGGTATGCTG CTTCCAGCAT ATAAAAGCAG ATATAGTAGC CATAAGCATG
(c)  CGGTATGCTG TTTCCGGCAT ATAAAAGTGG ATATAGTAAC CGTAAGTATA

     1001                                                      1050
(a)  ATGGTTAAGC TAAATT.... ....CACCGA TGAGTCGGAG GAGCCATCAT
(b)  ATGGTTAAGC TAAATT.... ....CACCGA TGAGTCGGAG GAGCCATCAT
(c)  GTGGTTTTGT GAAAATTCCA TCGACATGAA TTTGTAGGTG GAGCCACTAT
```

Figure 9 (cont'd).

```
     1051                                                       1100
(a)  GTGTACAATA GGGGGAAGCC CCTATGGCAA ATTATCTGTA TAGGAGCCCT
(b)  GTGTACAATA GGGGGAAGCC CCTATGGCAA ATTATCTGTA TAGGAGCCCT
(c)  ATGTAAAATA GGGGGAAGCC CCTATGGCGA ATTATCTGTA TAGGAACCCT

     1101                                                       1150
(a)  TTGCTGGGGT TAAAAGCTTA AGGTTTAGTG TAACACAACA TTGGGTGTAC
(b)  TTGCTGGGGT TAAAAGCTTA AGGTTTAGTG TAACACAACA TTGGGTGTAC
(c)  TTGCTGGGGT TAAAAGCTTA AGGTCTAGTG TATCACAACG TTGAGTGTAC

     1151                                                       1200
(a)  TCAAGAGCGT GTGGGGTGGC ACCCACGTGC TTGGATGAGG TCCGGAAATG
(b)  TCAAGAGCGT GTGGGGTGGC ACCCACGTGC TTGGATGAGG TCCGGAAATG
(c)  TCAAGAGTGT GTAGGGTGGC ACCTACATGC TTGGATGAGG TCCGGAGATG

     1201                                                       1250
(a)  AATACCGGGG GATAATTAAT CCCAGCTCAG GCACTAAGCT GACTTTCATG
(b)  AATACCGGGG GATAATTAAT CCCAGCTCAG GCACTAAGCT GACTTTCATG
(c)  AATACCGGGG GATAATCAAT CCCAGCTCAG ACATTAGGCT GACTTTCATG

     1251                                                       1300
(a)  GAAGTGTCCA TGACGCATTT TAAGGTAGGT TTTAGACATA ACCTCCCGGG
(b)  GAAGTGTCCA TGACGCATTT TAAGGTAGGT TTTAGACATA ACCTCCCGGG
(c)  GAAGTGTCCA TGACGCATTT TAAGGTAGGT TTTAGACATA ACCTCCCGGG

     1301                                                       1350
(a)  ATGGAAGTGA TTACCATTTC GTTATTCGTT ATTAGTTTCT TGCAACTATG
(b)  ATGGAAGTGA TTACCATTTC GTTATTCGTT ATTAGTTTCT TGCAACTATG
(c)  ATGGACGTGA TTACCATTTC GTTATTTGTT ATTAGTTTCT TGCAATTATG

     1351                                                       1400
(a)  ATGAGGGGAC CACATCTTAA GCGATGTTGC TGCATTGCGT ACCTATGGTC
(b)  ATGAGGGGAC CACATCTTAA GCGATGTTGC TGCATTGCGT ACCTATGGTC
(c)  ATGAGGGGAC CACATTTTAA GCGATGTTGC TGCATTGCGC ACCTATGGTC

     1401                                                       1450
(a)  ATCTGGTTAG TTGTCGTATT TTCTTTTAGC .TTTTGTGGC GACAGATGAG
(b)  ATCTGGTTAG TTGTCGTATT TTCTTTTAGC .TTTTGTGGC GACAGATGAG
(c)  ATCTGATTAA TTGTCGTATT TTCTTTTAGC TTTTTGTGGC GATAGATGAG

     1451                                                       1500
(a)  GTTTGACTCC TTTTCCTTGA CTCTTGACCT AAGTTGGACA CAAAAATATG
(b)  GTTTGACTCC TTTTCCTTGA CTCTTGACCT AAGTTGGACA CAAAAATATG
(c)  GTTTG..TCC TTTTCCTTGT CT....TGCT AAGTTGGACA CAAAAATTTG
```

Figure 9 (cont'd).

```
     1501                                                  1543
(a)  GTCTTTTGAC TTTCAATAGA GTCGATGAAA ATGTCTGCAT CAC
(b)  GTCTTTTGAC TTTCAATAGA GTCGATGAAA ATGTCTGCAT CAC
(c)  TCTTTTTG... TTTCAATAGA GTCGATGAAA ATGTCTTCAT CAC
```

Fig.9. Nucleotide sequence of the 3'-1527 nucleotides of PRMV RNA1 cDNA (a) (positions 6450-7977) is compared with the nucleotide sequence from a tandem-ligated cDNA clone 4-2.2 This clone contained 3'-terminal regions from both RNA1 (b) and RNA2-cDNA (c), respectively. The entire RNA1 cDNA nucleotide sequence derived from the recombinant 4-2.2 cDNA clone consists of 1220 bases which corresponds to the 3'-UTR 3'-terminus in PRMV RNA1 and 1501 bases from the same region of genomic RNA2. Nucleotide sequence alignment was generated using PILEUP and GAP (GCG) with gap and gap length weights of 5.0 and 0.3, respectively (Devereaux *et al.*, 1984). Nucleotide differences between 3'-termini of RNA1 and 2 are underlined. Numbering above the nucleotide sequence begins relative to position 6450 in PRMV RNA1 (Chapter 2 Fig.4). Gaps created in the nucleotide sequence alignment are indicated (...).

Analysis of the RNA2 3'-Terminus

PRMV RNA1 and RNA2 share extensive nucleotide sequence identity in their 3'-UTRs. All nepoviruses whose 3'-termini have been determined have demonstrated this characteristic: subgroup II BBLMV (Bacher *et al.*, 1994) CLRV (Scott *et al.*, 1992) and TomRSV (Rott *et al.*, 1991); and subgroup I nepoviruses GFLV, TBRV, and GCMV (Sanfaçon, 1995). The length of the 3'-UTR distinguishes these two nepovirus subgroups. PRMV may now be added to the list of subgroup II nepoviruses including BBLMV, CLRV, and TomRSV who share 3'-UTRs greater than 1.4 kb. This lends further confirmation to the subgroup II status of PRMV.

Researchers have speculated on the significance of the extraordinary length of subgroup II nepovirus 3'-UTRs (Bacher *et al.*, 1994; Buckley *et al.*, 1993; and Sanfaçon 1995). It is possible that replicase recognition sites and packaging signals are contained in

this region (Buckley *et al.*, 1993) and are conserved in both genomic RNAs. Although little significant nucleotide homology exists between nepovirus 3'-UTRs, certain nucleotide sequences of 8 to 30 nucleotides are conserved (see above). Research is needed to identify their respective functions. Interestingly, nepovirus nucleotide consensus sequences detected in RNA1 3'-UTR are found in the RNA2 3'-terminus at the same position relative to the poly (A) tail. If, in fact, these nucleotide consensus sequences are important for RNA replication and/or packaging, their identical position in the RNA1 and RNA2 genome would indicate that these functions are carried out in similar manner for the entire PRMV genome. A comparison of the 3'-terminal 500 nucleotides presented of RNAs 1 and 2 indicated 32 base differences. These substitutions would destabilize the proposed RNA secondary structure. Our current research is aimed at determining the effect of the nucleotide differences between RNA1 and 2 3'-UTRs on RNA secondary structures.

LIST OF REFERENCES

LIST OF REFERENCES

Allen, W.R. (1986). Effectiveness of Ontario population of *Longidorus diadecturus* and *L. breviannulatus* as vectors of peach rosette mosaic and tomato black ring viruses. *Canadian Journal of Plant Pathology* 8, 49-53.

Allen, W.R., and Ebsary, B.A. (1988). Transmission of raspberry ringspot, tomato black ring, and peach rosette mosaic viruses by an Ontario population of *Longidorus elongatus*. *Canadian Journal of Plant Pathology* 10(1), 1-18.

Allison, R.F., Janda, M., and Ahlquist, P. (1988). Infectious *in vitro* transcripts from cowpea chlorotic mottle virus cDNA clones and exchange of individual RNA components with brome mosaic virus. *Journal of Virology* 62 (10), 3581-3588.

Argos, P. (1988). A sequence motif in many polymerases. *Nucleic Acids Research* 16(21), 9909-9916.

Bacher, J., Warkentin, D., Ramsdell, D., and Hancock, J.F. (1994). Selection versus recombination: what is maintaining identity in the 3' termini of blueberry leaf mottle nepovirus RNA1 and RNA2? *Journal of General Virology* 75, 2133-2138.

Bacher, J., Warkentin, D., Ramsdell, D., and Hancock, J.F. (1994). Sequence analysis of the 3' termini of blueberry leaf mottle virus. *Virus Research* 33, 145-156.

Bardonnet, N., Hans, F., Serghini, M.A. and Pinck, L. (1994). Protection against virus infection in tobacco plants expressing the coat protein of grapevine fanleaf nepovirus. *Plant Cell Reports* 13 (6), 357-360.

Bazan, J.F., and Fletterick, R.J. (1988). Viral cysteine proteinases are homologous to the trypsin-like family of serine proteinases: structural and functional implications. *Proceedings of the National Academy of Science, USA* 85, 7872-7876.

Beachy, R.N., Loech-Fries, S. and Turner, N.E. (1990). Coat protein-mediated resistance against virus infection. *Annual Review of Phytopathology* 28, 451-474.

Bird, G.W., and Ramsdell, D.C. (1985). Population trends and vertical distribution of plant-parasitic nematodes associated with *Vitis labrusca* L. in Michigan. *Journal of Nematology* 17(2), 100-107.

Blok, V.C., Wardell, J., Jolly, C.A., Manoukian, A., Robinson, D.J., Edwards, M.L., and Mayo, M.A. (1992). The nucleotide sequence of RNA2 of raspberry ringspot nepovirus. *Journal of General Virology* 73, 2189-2194.

Brault, V., Hilbrand, L., Candresse, T., LeGall, O., and Dunez, J. (1989). Nucleotide sequence and genetic organization of Hungarian grapevine chrome mosaic nepovirus RNA2. *Nucleic Acids Research* 17(19), 7809-7819.

Brooks, M. and Bruening, G. (1995). A subgenomic RNA associated with cherry leafroll virus infections. *Virology* 211, 33-41.

Brown, T. (1993). Analysis of RNA by northern and slot blot hybridization. *Current Protocols in Molecular Biology* Editors Ausubel, F.M., Brent, R., Kingston, R.E., Moore,

Buckley, B., Silva, S., and Singh, S. (1993). Nucleotide sequence and *in vitro* expression of the capsid protein gene of tobacco ringspot virus. *Virus Research* 30, 335-349.

Cadman, C.H. (1963). Biology of soil-borne viruses. *Annual Review of Phytopathology* 1, 143-172.

Demangeat, G., Hemmer, O., Reinbolt, J., Mayo, and M.A., Fritsch, C. (1992). Virus-specific proteins in cells infected with tomato black ring nepovirus: evidence for proteolytic processing *in vivo*. *Journal of General Virology* 73, 1609-1614.

Dessens, J.T. and Lomonosoff, G.P. (1992). Sequence upstream of the 24K protease enhances cleavage of the cowpea mosaic virus 24K and 87K proteins. *Virology* 189, 225-232.

Devereaux, J., Haeberli, P. and Smithies, O. (1984). A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acid Research* 12, 387-395.

Dias, H.F., and Allen, W.R. (1980). Characterization of the single protein and two nucleic acids of peach rosette mosaic virus. *Canadian Journal of Botany* 58, 1747-1754.

Doolittle, R.R. (1981). Similar amino acid sequences: chance or common ancestry? *Science* 214 (9), 149-159.

Eveleigh, E.S., and Allen, W.R. (1982). Description of *Longidorus diadecturus* n.sp. (Nematoda: Longidoridae), a vector of the peach rosette mosaic virus in peach orchards in southwestern Ontario, Canada. *Canadian Journal of Zoology* 60: 112-115.

Fang, G.W. Allison, R.F., Zamboulim, E.M., Maxwell, D.P., and Gilbertson, R.L. (1995). The complete nucleotide sequence and genome organization of bean common mosaic virus (NL3 strain). *Virus Research* 39, 13-23.

Fitchen, J.H. and Beachy, R.N. (1993). Genetically engineered protection against viruses in transgenic plants. *Annual Review of Microbiology* 47, 739-763.

Francki, R.I.B., Milne, R.G., and Hatta, T. (1985). Nepovirus Group, in *Atlas of Plant Viruses, Vol. II.*, 23-38.

Fuchs, M., Pinck, M., Serghini, M.A., Ravelonadro, M., Walter, B., and Pinck, L. (1989). The nucleotide sequence of satellite RNA in grapevine fanleaf virus strain F13. *Journal of General Virology* 70, 955-962.

Goldbach, R. (1987). Genome similarities between plant and animal viruses. *Microbiological Sciences* 4(7), 197-202.

Goldbach, R., Martelli, G.P. and Milne, R.G. (1995). Genus Nepovirus. *Sixth Report of the International Commission on Taxonomy of Viruses.* Pages 345-347 in: Murphy, F.A., Fauquet, C.M., Bishop, D.H.L., Ghabrial, S.A., Jarvis, A.W., Martelli, G.P., Mayo, M.A., and Summers, M.D., eds., Springer-Verlag Wien, New York,.

Goldbach, R., Rezelman, G., Zabel, P. and van Kammen, A. (1982). Expression of the bottom-component RNA of cowpea mosaic virus. Evidence that the 60-kilodalton VPg precursor is cleaved into single and a 58-kilodalton polypeptide. *Journal of General Virology* 42, p.630.

Golemboski, D.B., Lomonossoff, G.P. and Zaitlin, M. (1990). Plants transformed with a tobacco mosaic virus nonstructural gene sequence are resistant to the virus. *Proceedings of the National Academy of Science USA.* 87, 6311-6315.

Gorbalenya, A.E., and Koonin, E.V. (1989). Viral proteins containing the purine NTP-binding sequence pattern. *Nucleic Acids Research* 17(21), 8413-8440.

Gorbalenya, A.E., Donchenko, A.P., Blinov, V.M., and Koonin, E.V. (1989). Cysteine proteinases of positive strand RNA viruses and chymotrypsin-like serine proteinases: a distinct protein superfamily with a common structural fold. *Federation of European Biochemical Societies* 243(2), 103-114.

Greif, C., Hemmer, O., and Fritsch, C. (1988). Nucleotide sequence of tomato black ring virus RNA1. *Journal of General Virology* 69, 1517-1529.

Grieco, F., Martelli, G.P. and Savino, V. (1995). The nucleotide sequence of RNA3 and RNA4 of olive latent virus 2. *Journal of General Virology* 76, 929-939.

Grumet, R. (1990). Genetically engineered plant virus resistance. *Hortscience* 2, 508-513.

Grumet, R. (1995). Genetic engineering for crop virus-resistance. *Hortscience* 30 (3), 449-455.

Hanahan, D. (1983). Studies on transformation of *Escherichia coli* with plasmids. *Journal of Molecular Biology* 166, 557-580.

Hans, F., and Sanfaçon, H. (1995). Tomato ringspot nepovirus proteinase: characterization and cleavage site specificity. *Journal of General Virology* 76, 917-927.

Harrison B.D. and Murant, A.F. (1977). Nepovirus group. In: Description of plant viruses. *Commonwealth Mycological Institute/Association of Applied Biologists* No. 185.

Harrison, B.D., and Barker, H. (1978). Proteinase-sensitive structure needed for infectivity of nepovirus RNA. *Journal of General Virology* 40, 711-715.

Harrison, B.D., Finch, J.T., Gibbs, A.J., Hollings, M., Shepherd, R.J., Valenta, V., and Wetter, C. (1971). Sixteen groups of plant viruses. *Virology* 45, 356-363.

Hellen, C.U.T., and Wimmer, E. (1992). Maturation of poliovirus capsid proteins. *Virology* 187, 391-397.

Hemmer, O., Greif, C., Dufourcq, P., Reinbolt, J., and Fritsch, C. (1995). Functional characterization of the proteolytic activity of the tomato black ring nepovirus RNA1 encoded polyprotein. *Virology* 206, 362-371.

Henikoff, S. (1984). Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. *Gene* 28, 351-359.

Jaeger, J.A., Turner, D.H. and Zuker, M. (1989). Improved predictions of secondary structures for RNA. *Proceedings of the National Academy of Science, U.S.A.* 86, 7706-7710.

Kikkert, J.R., Hebertsoule, D., Wallace, P.G., Striem, M.J. and Reisch, B.I. (1996). Transgenic plantlets of Chancellor grapevine (*Vitis* spp.) from biolistic transformation of embryogenic cell suspensions. *Plant Cell Reports* 15 (5), 311-316.

Klos, E.J. (1976). Rosette mosaic. *U.S. Department of Agriculture Agricultural Handbook.* No. 437. U.S. Government Printing Office, Wash., D.C. 433 pg.

Klos, E.J., Fronec, F., Knierim, J.M. and Cation, D. (1967). Peach rosette mosaic transmission and control studies. *Michigan Agricultural Experiment Station Quarterly Bulletin.* 49, 287-293.

Krastanova, S., Perrin, M., Barbier, P., Demangeat, G., Cornnet, P., Bardonnet, N., Otten, L., Pinck, L. and Walter, B. (1995). Transformation of grapevine rootstock with the coat protein of grapevine fanleaf nepovirus. *Plant Cell Reports* 14 (9), 550-554.

Kreiah, S., Strunk, G., and Cooper, J.I. (1994). Sequence analysis and location of capsid proteins within RNA2 of strawberry latent ringspot virus. *Journal of General Virology* 75, 2527-2532.

Le Gall, O., Candresse, T., Brault, V., and Dunez, J. (1989). Nucleotide sequence of Hungarian grapevine chrome mosaic nepovirus RNA1. *Nucleic Acids Research* 17(19), 7795-7807.

Le Gall, O., Torregrosa, L., Danglot, Y., Candresse, T., and Bouquet, A. (1994). *Agrobacterium*-mediated genetic transformation of grapevine somatic embryos and regeneration of transgenic plants expressing the coat protein of grapevine chrome mosaic nepovirus (GCMV). *Plant Science* 102, 161-170.

Lee., S. and Rasheed, S. (1990). A simple procedure for maximum yield of high-quality plasmid DNA. *Biotechniques* 9(6), 676-677.

Lomonosoff, G.P., and Shanks, M. (1983). The nucleotide sequence of cowpea mosaic virus B RNA. *EMBO Journal* 2, 2253-2258.

Loudes, A.M., Ritzenthaler, C., Pinck, M., Serghini, M.A., and Pinck, L. (1995). The 19kDa and 124kDa polyproteins of arabis mosaic nepovirus (isolate S) are encoded by two distinct RNA2 species. *Journal of General Virology* 76, 899-906.

Lupo, R., Martelli, G.P., Castellano, M.A., Boscia, D. and Savino, V. (1994). *Agrobacterium rhizogenes*- transformed plant roots as a source of grapevine viruses for purification. *Plant Cell Tissue and Organ Culture* 36 (3), 291-301.

Maniatis, T., Fritsch, E.F., and Sambrook, J. (1982). Molecular Cloning. Cold Spring Harbor Press.

Margis, R., and Pinck, L. (1992). Effects of site-directed mutagenesis on the presumed catalytic triad and substrate-binding pocket of grapevine fanleaf nepovirus 24-kDa proteinase. *Virology* 190, 884-888.

Margis, R., Viry, M., Pinck, M., and Pinck, L. (1991). Cloning and *in vitro* characterization of the grapevine fanleaf virus proteinase cistron. *Virology* 185, 779-787.

Margis, R., Viry, M., Pinck, M., Bardonnet, N., and Pinck, L. (1994). Differential proteolytic activities of precursor and mature forms of the 24K proteinase of grapevine fanleaf nepovirus. *Virology* 200, 79-86.

Martelli, G.P. (1975). Some features of nematode-borne viruses and their relationship with the host plants, *Nematode Vectors of Plant Viruses*. Page 23 in Lamberti, F., Taylor, C.E., and Seinhorst, J.W., eds., *Life Sciences* 2, (NATO Advanced Study Institute, Series A), Plenum Press, New York.

Martelli, G.P. and Taylor, C.E. (1990). Distribution of viruses and their nematode vectors Vol. 6. Pages 151-189 in: *Advances in Disease Vector Research*, Ed. Harris, K.F. Springer-Verlag, New York.

Martinelli, L. and Mandolino, G. (1994). Genetic transformation and regeneration of transgenic plants in grapevine (*Vitis rupestris* S.). *Theoretical and Applied Genetics* 88 (6-7), 621-628.

Matthews, R.E.F. (1982). Classification and nomenclature of viruses. *Intervirology* 71, 1-199.

Mauro, M.C., Toutain, S., Walter, B., Pinck, L., Otten, L., Contosthevenot, P., Deloire, A. and Barbier, P. (1995). High efficiency regeneration of grapevine plants transformed with the GFLV coat protein gene. *Plant Science* 112 (1), 97-106.

Mayo, M.A., Barker, H., and Harrison, B.D. (1982). Specificity and properties of the genome-linked proteins of nepoviruses. *Journal of General Virology* 59, 149-162.

Meyer, M., Hemmer, O., Mayo, M.A., and Fritsch, C. (1986). The nucleotide sequence of tomato black ring virus RNA2. *Journal of General Virology* 67, 1257-1271.

Nakano, M., Hoshino, Y. and Mil, M. (1994). Regeneration of transgenic plants of grapevine (*Vitis vinifera* L.) via *Agrobacterium rhizogenes*- mediated transformation of embryogenic calli. *Journal of Experimental Botany* 45 (274), 649-656.

Palmenberg, A.C. (1990). Proteolytic processing of picornaviral polyprotein. *Annual Review in Microbiology* 44, 603-623.

Peters, S.A., Voorhorst, W.G.B., Wellink, J. and Van Kammen, A. (1992). Processing of VPg-containing polyproteins encoded by the B-RNA from cowpea mosaic virus. *Virology* 191, 90-97.

Pinck, M., Reinbolt, J., Loudes, A.M., Le Ret, M. and Pinck, L. (1991). Primary structure and location of the genome-linked protein (VPg) of grapevine fanleaf nepovirus. *FEBS Letters* 284, 117-119.

Racaniello, V.R., and Baltimore, D. Molecular cloning of poliovirus cDNA and determination of the complete nucleotide sequence of the viral genome. *Proceedings of the National Academy of Science, USA* 78, 4887-4891.

Ramsdell, D.C. and Gillett, J.M. (1983). Superimposed shallow and deep soil fumigation to control *Xiphinema americanum* and peach rosette mosaic virus reinfection in a 'Concord' vineyard. *Plant Disease* 67(6), 625-627.

Ramsdell, D.C., and Gillett, J.M. (1981). Peach rosette mosaic virus in highbush blueberry. *Plant Disease* 65(9), 757-758.

Ramsdell, D.C., and Gillett, J.M. (1985). Relative susceptibility of American, French hybrid and European grape cultivars to infection by peach rosette mosaic virus. *Phytopathologica Mediterranea* 24, 41-43.

Ramsdell, D.C., and Myers, R.L. (1974). Peach rosette mosaic virus, symptomatology and nematodes associated with grapevine 'degeneration' in Michigan. *Phytopathology* 64(9), 1174-1178.

Ramsdell, D.C., and Myers, R.L. (1978). Epidemiology of peach rosette mosaic virus in a 'Concord' grape vineyard. *Phytopathology* 68, 447-450.

Ramsdell, D.C., Gillett, J.M., and Bird, G.W. (1995). Susceptibility of American grapevine scion cultivars and French hybrid rootstock and scion cultivars to infection by peach rosette mosaic nepovirus. *Plant Disease* 79(2), 154-157.

Raski, D.J., Hewitt, W.B., Goheen, A.C, Taylor, C.E., and Taylor, R.H. (1965). Survival of *Xiphinema index* and reservoirs of fanleaf virus in fallowed vineyard soil. *Nematologica* 11, 349-352.

Ritzenthaler, C., Viry, M., Pinck, M., Margis, R., Fuchs, M. and Pinck, L. (1991). Complete nucleotide sequence and genetic organization of grapevine fanleaf nepovirus RNA1. *Journal of General Virology* 72, 2357-2365.

Ritzenthaler, C., Viry, M., Pinck, M., Margis, R., Fuchs, M., and Pinck, L. (1991). Complete nucleotide sequence and genetic organization of grapevine fanleaf nepovirus RNA1. *Journal of General Virology* 72, 2357-2365.

Rott, M.E., Gilchrist, A., Lee, L., and Rochon, D.M. (1995). Nucleotide sequence of tomato ringspot virus RNA1. *Journal of General Virology* 76, 465-473.

Rott, M.E., Tremaine, J.H., and Rochon, D.M. (1991). Comparison of the 5' and 3' termini of tomato ringspot virus RNA1 and RNA2: evidence for RNA recombination. *Virology* 185, 468-472.

Rott, M.E., Tremaine, J.H., and Rochon, D.M. (1991). Nucleotide sequence of tomato ringspot virus RNA2. *Journal of General Virology* 72, 1505-1514.

Sanfaçon, H. (1995). Nepoviruses in *Pathogenesis and Host Specificity in Plant Diseases* Vol. III. Pages 129-141 in: Eds. Singh, R.P., Singh, U.S., and Komuto, K. 129-141.

Sanfaçon, H., Wieczorek, A., and Hans, F. (1995). Expression of the tomato ringspot nepovirus movement and coat proteins in protoplasts. *Journal of General Virology* 76, 2299-2303.

Sanger, F., Nicklen, S., and Coulson, A.R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Science, U.S.A.* 74, 5463-5467.

Scholthof, K.G., Scholthof, H.B. and Jackson, A.O. (1993). Control of plant virus diseases by pathogen-derived resistance in transgenic plants. *Plant Physiology* 102, 7-12;

Serghini, M.A., Fuchs, M., Pinck, M., Reinbolt, J., Walter, B., and Pinck, L. (1990). RNA2 of grapevine fanleaf virus: Sequence analysis and coat protein cistron location. *Journal of General Virology* 71, 1433-1441.

Slatko, B., Heinrich, P., Nixon, B.T., Voytas, D. (1994). *Current Protocols in Molecular Biology* Vol. I. Page 7.2 in: Eds. Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A., Struhl, K. John Wiley and Sons, Inc., New York, NY.

Stace-Smith, R., Ramsdell, D.C. (1987). Nepoviruses of the Americas. *Current Topics in Vector Research* Vol. III. Pages 131-166 in Ed. Harris, K.F.

Stobbs, L.W. and Van Schagen, J.G. (1996). Occurrence of peach rosette mosaic virus on grapevine in southern Ontario. *Plant Disease* 80, 105.

Tabor, S. (1987). Phosphatases and Kinases in *Current Protocols in Molecular Biology* Vol I. Pages 3.10.1-3.10.2 in: Eds. Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A., Struhl, K.

Vos, P., Verver, J., Jaegle, M., Wellink, J., Van Kammen, A. and Goldbach, R. (1988). Two viral proteins involved in the proteolytic processing of the cowpea mosaic virus polyproteins. *Nucleic Acids Research* 16, 1967-1985.

Wellink, J., and van Kammen, A. (1988). Proteinases involved in the processing of viral polyproteins. *Archives of Virology* 98, 1-26.

Wellink, J., Rezelman, G., Goldbach, R., and Beyreuther, K. (1986). Determination of the proteolytic processing sites in the polyprotein encoded by the bottom-component of cowpea mosaic virus. *Journal of General Virology* 59, 50-58.

Wildy, P. (1971). Classification and nomenclature of viruses. First report of the International Committee on Nomenclature of Viruses. *Monographs in Virology* 5, 1.