This is to certify that the

thesis entitled

Genomic localization of the 5'-linked protein (VPg)
of Pea Enation Mosaic Enamovirus

presented by

Christiane Elisabeth Wobus

has been accepted towards fulfillment
of the requirements for

Master's degree in Botany & Plant Pathology

Major professor

Date 3/5/97

Genomic localization of the 5'-linked protein (VPg) of Pea Enation Mosaic *Enamovirus*

By

Christiane Elisabeth Wobus

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

1997

ABSTRACT

Genomic localization of the 5'-linked protein (VPg)

of Pea Enation Mosaic *Enamovirus*

By

Christiane Elisabeth Wobus

This study reports that the protein sequence of Pea Enation Mosaic Virus (PEMV) genome-linked, viral protein (VPg) is encoded within the 84K ORF 1 of RNA 1. It is composed of 28 amino acids with a molecular mass (M_r) of 3157 Da. Position and size of the VPg suggest that ORF 1 is expressed as a polyprotein. This is the first evidence of proteolytic cleavage of a protein encoded by PEMV RNA 1 or by *Luteoviruses*. Computer-generated VPg sequence comparisons did not reveal sequence similarity to other viruses including *Luteoviruses*. Anti-VPg antibodies were produced and used in the identification and characterization of the VPg. A 15 kDa protein doublet was detected in protein extracts from RNA 1 and RNA 2 infected plants but not in extracts of plants infected with RNA 2 alone. Biotinylated anti-VPg antibodies detected a similar size protein in RNase A-treated preparations of viral RNA.

ACKNOWLEDGEMENTS

I gratefully acknowledge my professor Dr. Gustaaf de Zoeten for his support in making this thesis possible. I also thank Dr. Richard Allison and Dr. Rebecca Grumet for their help on my graduate committee. I thank Dr. Jihad Skaf and Debby Rucker for their assistance in the laboratory. Thank you to Joe Leykam and the staff of the MSU macromolecular structure facility for protein sequencing and scientific assistance. Thanks to the staff of the University Laboratory Animal Resources (ULAR) for the care of the rabbit, to Kurt Stepnitz and Marlene Cameron for their photographic expertise, and to the department office staff for their help. Lastly, I like to thank my husband Erik and my family for their love and support, without whose help this work would not have been possible.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Literature review

Pea Enation Mosaic Virus (PEMV)

1. *Genomic organization*

Pea Enation Mosaic Virus is the sole member of the *Enamovirus* genus. The two linear single-stranded (ss), plus-sense (+) RNAs are encapsidated in two icosahedrons of different sizes. The 5706 nucleotides (nt) of RNA 1 are encapsidated in a polyhedron of 180 subunits (T=3) with the diameter of 28nm. While the 25nm diameter particle consisting of 150 subunits contains the RNA 2 genome of 4253 nt. In viral preparations the particles were found to sediment at different rates, leading to the naming of bottom (B) and top (T) component, for RNA 1 and RNA 2 respectively (for review see Demler *et al.*, 1996a). The RNAs could not be aminoacylated and contain no poly (A) sequence at their 3' end (German *et al.*, 1978). The 5' end of the viral genome is linked covalently to a protein (VPg) of 17.5kDa (Reisman & de Zoeten, 1982). Some strains also contain a third RNA of 717nt, which is a satellite RNA of RNA 2 (for review see Demler *et al.*, 1996a).

RNA 1 contains five open reading frames (ORF) and encodes all structural proteins (the numbering system is based on Miller *et al.*, 1995) (Fig. 1). ORF 0 starts after a 183 nt non-coding region and has the potential to encode a 34kDa protein. The C-terminus of the putative protein is acidic while a

small region towards the N-terminus is characteristic of transmembrane proteins (Demler & de Zoeten, 1991). However, no significant sequence homology was found in database searches leaving the function of the putative protein unknown. ORF 1 (84K) is initiated out of frame from ORF 0 at a second AUG at nt 265. The putative 84kDa protein contains several hydrophobic stretches at its N-terminus characteristic for membrane spanning domains. A helicase IV-like motif and a 3C-like protease motif were identified towards the center of the ORF. The sequence at the C-terminus did not resemble any known protein sequence (Demler & de Zoeten, 1991). ORF 2 encodes the viral polymerase and extensively overlaps at its 5' end with ORF 1. The occurrence of a frameshift-motif led to the hypothesis that the product ORF 2 is expressed as a frameshift fusion protein with ORF 1 (Demler & de Zoeten, 1991). A different mechanism is used to express the 3' half of the genome encoding the structural proteins. The intergenic region between ORFs 1, 2 and ORFs 3, 4, 5 resembles a signal for subgenomic RNA production (Demler & de Zoeten, 1991). ORF 3 encodes the major capsid protein (cp) of 21kDa (Demler & de Zoeten, 1991), previously identified as a 22 kDa protein (Shepherd *et al.*, 1968; Hull & Lane, 1973; Hill & Shepherd, 1972). In addition to the 21kDa cp, aphid transmissible isolates also contain a minor capsid protein of 54 kDa. The latter is the result of read-through (RT) of the 21kDa cp (ORF 3) and a 33kDa protein (ORF 5) and is involved in aphid transmission. At the N-terminus of the 33kDa protein a 4kDa proline rich region connects both proteins; this might form a hinge to allow for better interaction of the aphid transmission protein with the

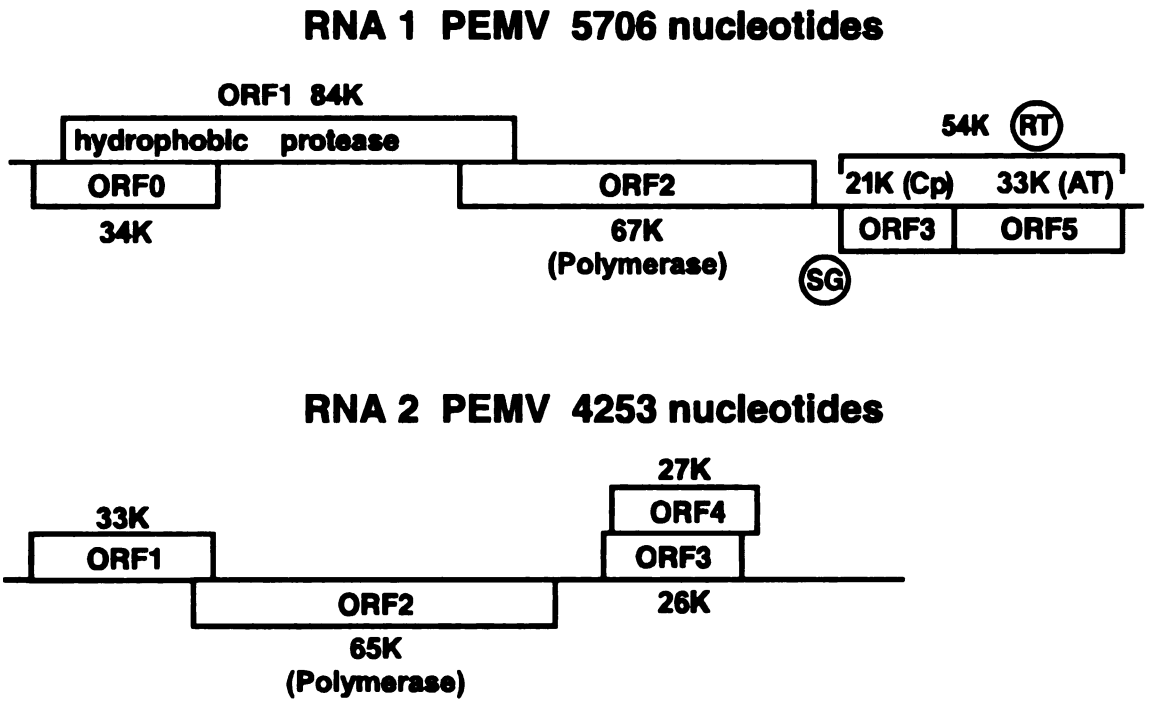

Figure 1: Localization of PEMV RNA1 and RNA2 open reading frames (ORFs). The localization of RNA1 and RNA2 ORFs is based on Demler & de Zoeten (1991) and Demler et al. (1993), respectively. The numbering system of RNA1 ORFs is based on Miller et al. (1995). The hydrophobic region and the protease motif, identified by Demler & de Zoeten (1991) are indicated. Abbreviations: coat protein (Cp), aphid transmission factor (AT), subgenomic RNA signal (SG), read-through (RT), K=kDa

aphid receptors (Demler & de Zoeten, 1991). The most convincing evidence for the expression of the 54 kDa fusion protein via RT was obtained when N-terminal protein sequencing of the 21kDa and 54kDa proteins resulted in identical sequences (Demler *et al.*, 1996a). The involvement of the 54kDa protein in aphid transmission was shown by the use of infectious cDNA clones. These clones of the 3' portion of RNA 1, including ORF 3 and ORF 5, of an aphid-transmissible (AT^+), aphid-nontransmissible (AT^-) and a naturally occurring, nontransmissible deletion strain (AT_Δ^-) were used to generate cDNA clones with various exchanges of portions of the parental clones. When these infectious clones were inoculated together with RNA 2, it was determined that the 110nt near the 3' end of the 33K ORF are critical for aphid transmission (Demler *et al.*, 1996a). The 3' end of RNA 1 ends in a 227 nt non-coding region.

RNA 2 has four confirmed ORFs (Fig. 1) and a potential 15 K 3' terminal ORF that lacks a termination codon is dispensable in infections and expression of this protein is unlikely (Demler *et al.*, 1993). ORF 1 is initiated after a short 5'terminal non-coding region of 21 nt. This ORF can potentially encode a 33 kDa protein but no function is assigned due to lack of statistically significant sequence homology to known sequences (Demler *et al.*, 1993). It is believed that the protein product of this ORF migrates anomalously and corresponds to a 45 kDa protein detected after *in vitro* translation of viral RNA (Gabriel & de Zoeten, 1984; Demler *et al.*, 1993). The second ORF encodes a 65kDa protein, the second viral polymerase. ORF 2 overlaps slightly with ORF 1, however, in a different reading frame. The occurrence of a minor translation product

corresponding in size to a fusion protein derived from ORF 1 and 2 suggests that expression occurs via frameshifting (Demler *et al.*, 1993). An intergenic region thought to encode a subgenomic promoter separates ORF 2 from ORFs 3 and 4 which have the potential to encode a 26kDa and a 27kDa protein, respectively. The reading frames are superimposed upon each other but are in different reading frames (Demler *et al.*, 1993). Preliminary results provide evidence that these proteins are involved in movement (J. Skaf, unpublished results).

Some strains also contain a satellite RNA of 717nt (Demler & deZoeten, 1989). Like other satellite RNAs, this RNA is characterized by a lack of sequence homology with the host plant or viral genome, it is non-essential in virus infection and depends on the helper virus for replication and encapsidation (Roossinck *et al.*, 1992). No extensive sequence homology was found between host RNA, RNA 1 or RNA 2 and the satellite (Demler & de Zoeten, 1989) or to other satellites (Demler *et al.*, 1994b). It is non-infectious by itself and has no influence on symptom expression, aphid transmission or particle morphology in peas when inoculated together with RNA 1 and 2 (Demler & de Zoeten, 1989). RNA 1 provides the structural and aphid transmission functions while RNA 2 is solely responsible for replication of the satellite. Very limited sequence homology exists between the satellite RNA and RNA 2, only 12 out of 14 nt at the 5' end and 7 of 8 nt at the 3' terminus are identical. This is thought to be required by the polymerase of RNA 2 in order to recognize the satellite RNA for replication (Demler *et al.*, 1994b). No multimeric or circular forms were found

suggesting a replication mechanism involving linear, monomeric intermediates. Computer analysis of the satellite sequence identified several ORFs, up to 4.3K on the plus-sense and up to 7.2 K ORFs on the minus-sense, but *in vitro* translation of purified satellite or transcript RNA showed no protein products (Demler *et al.*, 1994b).

2. *Interaction of RNAs and symptoms*

PEMV historically has been characterized as a bipartite virus since both particles are required for infectivity. However, after the elucidation of both genomic sequences, the interdependencies of RNA 1 and 2 became more obvious. The production of infectious transcripts and with it the possibility of infecting plants and protoplasts with individual RNAs or with a combination of RNAs brought clarification to the functions provided by each RNA. RNA 1 is solely responsible for encapsidation and aphid transmission. Evidence shows that genomic and subgenomic RNAs of RNA 1 were formed in the absence of RNA 2 and that RNA 1 is independently capable of cp expression and virion assembly (Demler *et al.*, 1994a). RNA 2, on the other hand, is needed for mechanical inoculation and phloem evasion. Protoplasts inoculated with RNA 1 or RNA 2 alone, showed replication of the respective genomes, indicating the replicative independence of each RNA. However, inoculation of plants with RNA 2 resulted in a systematic, although almost symptomless, infection (Demler *et al.*, 1994a). Only after prolonged infection (14-28 days post-inoculation [dpi]) did the RNA 2 infected plants show chlorotic yellowing and mosaic patterns or mild

mottling of upper leaves (Demler *et al.*, 1994a). The development of symptoms correlates with electron microscope (EM) findings that only RNA 1 infection causes the formation of double-membrane bound replication complexes in the cytoplasm of infected protoplasts (see below) (Demler *et al.*, 1994a).

3. *Economic importance, host range, environment, control*

PEMV occurs worldwide and causes significant losses in pea (*Pisum sativum*), chickpea (*Cicer arietinum*), lentils (*Lens culinaris*) and field bean (*Vicia faba*) in the United States as well as in Europe (for review see Hagedorn, 1996).

Its natural host range is restricted to leguminous plants. Alfalfa (*Medicago sativa*) is the most important and widespread host, while some members from the *Vicia* , *Lathyrus*, *Medicago*, *Pisum*, *Melilotus*, *Lupinus*, *Trifolium* and *Lens* genera have also been identified as natural hosts. However, some non-leguminous plant species are local lesion hosts, mainly members of the genera *Chenopodium* and *Nicotiana* (for review see Hagedorn, 1996).

In addition to the plant species, environmental factors also influence occurrence and severity of symptoms. Due to the aphid-transmissibility of the virus, factors influencing the life cycle of aphids become especially important. For instance, peas sown later in the growing season always display higher incidence of infection by PEMV due to an increase in aphid populations. The weather also has a profound effect. A high rate of transmission is observed on warm, humid and stormless days as many alatae (winged) aphids develop and move with the wind to uninfected peas. However, cold or hot, dry weather and

rainy days reduce aphid multiplication and the development of winged individuals, which results in low PEMV occurrence (for review see Hagedorn, 1996). Most efficient aphid transmission occurs around 20°C. In addition to the aphid, the outcome of a PEMV infection is also determined by the temperatures following inoculation. Generally, lower temperatures (12-14°C) reduce and higher temperatures (28-30°C) increase the manifestation and severity of symptoms in plants (for review see Hagedorn, 1996).

The different parameters influencing incidence and spread of PEMV can be controlled to reduce crop losses. Today, the development of host plants tolerant to PEMV infections is used commercially. Genetic resistance of varying degrees is determined by a single, dominant locus En. Introduction of this gene into susceptible cultivars resulted in excellent resistance. In recent years, the marker locus Adh-1 (coding for alcohol dehydrogenase) has been identified, that is tightly linked to En, permitting the screening for PEMV resistance by Adh-1 signals (for review see Hagedorn, 1996). The following control measures influencing different parameters are also possible. Aphid transmission of this virus is in a persistent manner and the use of insecticides can reduce the vector population and limit virus spread. However, due to the erratic occurrence of severe outbreaks, this practice is not widespread (for review see Hagedorn, 1996). PEMV has to overwinter in leguminous perennials in order for aphids to carry PEMV to new plants in the spring. A solution would be the development of aphid-resistant perennial legumes but this is not an active control effort at this time (for review see Hagedorn, 1996).

4. *Aphid transmission and virus-vector relationship*

As described earlier, the aphid vector plays an important role in the dissemination of naturally occurring PEMV infections. Therefore, a considerable amount of research has focused on the interaction of the virus and its vector.

Aphid transmissible (AT^+) and nontransmissible (AT^-) strains are known and naturally occurring PEMV infections contain a mixture of AT^+ and AT^- strains (Demler *et al.*, 1996a). Both strains in these mixtures can be transmitted by the aphid due to a process called phenotypic mixing, where the AT^- strain is transported within a virion containing cp determinants of the AT^+ strain necessary for aphid transmission during encapsidation (Adam, 1978). In the laboratory, an AT^+ strain is often converted into an AT^- strain after repeated mechanical passage. This is manifested upon purification of the virus in a higher yield of aphid-nontransmissible compared to aphid-transmissible strains (reviewed in Demler *et al.*, 1996a).

PEMV is transmitted in a circulative, nonpropagative manner. Circulative, also called persistent, transmission is characterized by the following: (I) The virus is acquired through the food canal and subsequently moves to the aphid's intestines. (II) After the virus has been taken up by the aphid through feeding, termed the acquisition feed, a latent period follows during which the virus passes through the hindgut wall into the hemocoel and subsequently into the salivary gland. (III) Finally, the virus is ejected with the saliva from the maxillary salivary canal and can infect another plant (Matthews, 1991). Support for this type of

virus movement through the aphid comes from many EM observations (Harris & Bath, 1972; Harris *et al.*, 1975; Harris, 1979; Harris, 1990). Virus was found concentrated after a 24 hr. acquisition period in the stomach, the intestine and the hindgut lamina but not in the digestive tract or foregut (reviewed by Demler *et al.*, 1996a). It is thought that, in the case of PEMV, crossing the barrier between hindgut and hemocoel is accomplished by simple penetration of the gut epithelial cells, since no evidence could be obtained indicating an interaction of the virus with the plasma membrane of these cells as characteristic for phagocytosis. Once in the hemocoel the virus circulates until membrane mediated endocytosis allows the virus to cross a second barrier to reach the salivary system. AT^+ and AT^- virions have been observed crossing the gut-hemocoel barrier but only AT^+ virions could be detected in the salivary system. Therefore, aphid transmissibility seems to be determined by a successful interaction with the accessory salivary gland membranes (reviewed in Demler *et al.*, 1996a).

Circulative transmission is divided into propagative and nonpropagative transmission, depending on virus replication in the aphid vector. Only a few aphid species transmit viruses circulatively; and these viruses commonly produce yellowing and leaf-rolling symptoms (Matthews, 1991). *Myzus persicae* and *Acyrthosiphon pisum* are the two most important vectors of PEMV which produces mosaic patterns (yellow spots) in the leaves as well as rolling of the infected leaves (see below).

PEMV is a circulative, nonpropagative virus. Although, its short acquisition and inoculation feeding periods as well as the ability to be inoculated during brief aphid test probes of non-phloem tissue are characteristic of non-circulative viruses. However, a latent period prior to successful aphid inoculation, typical for circulative transmission has been observed (reviewed in Demler *et al.*, 1996a). Other evidence for circulative transmission is that prolonged acquisition feeding leads to an increased transmission efficiency which results in a higher incidence of infected plants (Fargette *et al.*, 1982). Support for the nonpropagative nature of PEMV, e.g. no virus replication in the aphid, comes from the following observations: One, transmission efficiency decreases with time in aphids denied subsequent acquisition feedings. Two, aphids lose inoculativity (the ability of an aphid or insect to deliver infectious virus into a healthy plant [Matthews, 1991]) after serial passage of hemolymph between aphids. Third, transmission efficiency correlates with the amount of virus acquired by the aphid (reviewed in Demler *et al.*, 1996a). Adam (1978) hypothesized that PEMV is a nonpropagative virus that is stored but not replicated in fat bodies of certain cells of the mesoderm. Electron-microscopy showed virus inclusions in the cytoplasm of stomach and midgut cells, as well as in connective tissue, fat and basophilic mesodermal cells. Phagocytosis causes the inclusions to be incorporated into endocytotic vacuoles and other parts of the lysosomal apparatus. However, no virus breakdown was observed (reviewed in Demler *et al.*, 1996a). Heterologous encapsidation during infections of plants with AT^- and AT^+ mixtures was observed. Transencapsidation might take place

during replication in an aphid. However, injections of aphids with a mixture of strains does not yield heterologous virions (Adam, 1978). Furthermore, no virus multiplication was detected by enzyme-linked immunosorbent assays of aphids (Fargette *et al.*, 1982).

5. *Cytopathology, replication and translation*

Changes in symptoms of PEMV infected plants are accompanied by distinct changes in ultrastructure. No symptoms and no ultrastructural changes are visible up to 4 days post-inoculation (dpi). Plants infected for four to nine days show signs of necrosis, a downward curling of upper leaves and an increase of vesiculation in the sieve elements (Demler *et al.*, 1994a; de Zoeten & Gaard, 1983). At the same time electron-dense structures appear in the vascular bundles and PEMV virions are detected in the cytoplasm of all cell types as well as in the nuclei of these cells especially in the region of the nucleolus (de Zoeten & Gaard, 1983). However, it is currently not known which of these sites, nucleus or cytoplasm, is the place for virion assembly. Virions in the cytoplasm appear randomly scattered in loose aggregates or are contained in electron dense, dagger-like tubules in the plasmodesmata. During this time, 7-10 dpi, visible symptoms like vein clearing and yellow mosaic patterns appear while stunting and epinasty of aerial parts is seen in combination with an amplification of the foliar symptoms 7 - 14 dpi (Demler *et al.*, 1994a). Elongated feather-like crystalline inclusions are found in the perinuclear space and in epidermal cells starting 14 days after inoculation (reviewed in Demler *et al.*,

1996a). Enations on veins on the lower leaf surface do not become evident until late in the infection (starting at 20 dpi) (Demler *et al.*, 1994a).

The typical symptom of PEMV infected tissue at the ultrastructural level of resolution is the presence of double membrane bound vesicles in phloem related tissues and other parenchymatic cell types. These vesicles originate at the inner envelope of the nucleus and migrate into the perinuclear space. On their way into the cytoplasm the vesicles gain their second membrane when crossing through the outer envelope of the nucleus. Each of these single-membrane bound vesicle contains 2-30 individual single-bound vesicles that contain aggregated granular material and DNase sensitive fibrils (de Zoeten *et al.*, 1972). These vesicular structures often appear near plasmodesmata and are especially numerous in phloem tissue and sieve elements, maybe indicating a role in cell-to-cell or long distance movement (Demler *et al.*, 1996a). The extent of vesiculation and the amount of nuclei containing virus particles increase with time after inoculation (de Zoeten *et al.*, 1972). This and the following observations lead to the conclusion that replication is linked to the nucleus and the vesicles originating from it.

Positive strand viruses produce (-) strands during RNA replication and subsequently more (+) strands via the formation of replicative forms (RF) and replicative intermediates (RI). The RF is a structure of full-length (+) and (-) strand basepaired to each other while RI only contains partially basepaired structures (Matthews, 1991). De Zoeten *et al.* (1976) used *in situ* hybridization and ferritin labeled antibodies (Abs) to determine the site of viral RNA synthesis

by localizing double-stranded (ds) viral RNA to the nuclei of infected cells. However, the applied techniques were not able to establish whether RF and/or RI were present. Both forms were previously shown to be present in PEMV infected plants (German & de Zoeten, 1975). The involvement of vesicles in replication in addition to the nucleus was shown by the association of negative strand RNA with these structures (Powell *et al.*, 1977). The hybridization studies were accompanied by experiments showing incorporation of radiolabeled precursors into nuclei and vesicles in the presence of actinomycin D (AMD) (de Zoeten *et al.*, 1976; Powell & de Zoeten, 1977). AMD binds tightly and specifically to ds DNA but not to ss or ds RNA, therefore, preventing RNA synthesis of the host without affecting viral RNA synthesis (Stryer, 1988). It could also be shown that the viral RNA polymerase was localized to nuclei and vesicles and required the presence of all four nucleotide-triphosphates and Mg^{2+} for maximum efficiency (Powell *et al.*, 1977; Powell & de Zoeten, 1977). Extended experiments used isolated nuclei to study PEMV replication (Powell & de Zoeten, 1977). It was determined that 60% of virus-specific RNA is (+) ss and 40% ds and that (-) ss RNA exists only in ds form or bound by protein but not as free RNA. Intact progeny virions were not observed indicating that only early stages of the virus life cycle occur in isolated nuclei. There is an absolute requirement for uncoating of virions prior to replication in this nuclear system. Although the viral polymerase is AMD insensitive, host protein synthesis is still required for full activity, pointing to the involvement of host gene products in viral replication. The accumulation of double stranded (ds) RNA followed the

infectivity pattern (an increase in infectivity above the initially infected leaves) and coincided with the appearance of vesicles in the phloem (de Zoeten & Gaard, 1983).

Viral RNA was translated using rabbit reticulocyte lysate and wheat germ extract (Gabriel & de Zoeten, 1984). RNA 1 supported the translation of three major proteins: vp1 (147kDa), vp2 (88kDa) and vp4 (36kDa) while translation of RNA 2 resulted in only one protein, vp3 (45kDa). After the genomic sequence became available, ORFs could be assigned to the protein products. Vp1 corresponded to the fusion protein of ORF1 and 2, vp2 to ORF1 and vp4 to ORF0. No protein product encoded by ORFs of RNA 2 with the size of vp3 was found but additional translation experiments with deletion clones demonstrated that the 45kDa protein corresponded to ORF1 (Demler *et al.*, 1993).

Relationship of PEMV to other virus groups

1. *Evolutionary links*

Since no fossil records exist, the study of virus evolution became possible only after the advent of comparative sequence analysis. Positive strand RNA virus evolution is generally believed to be influenced by two trends, 1) conservation of distinct gene combinations involved in viral replication and expression and 2) recombination and reshuffling of conserved gene blocks (Koonin & Dolja, 1993; Gorbalenya & Koonin, 1993). In an exhaustive review, Koonin and Dolja (1993) analyzed the origin of all positive strand virus groups known to that date. Since RNA dependent RNA polymerases (RdRps) contain

the only universal sequence motifs, "RdRp is the only domain of positive-strand RNA viruses allowing an all-inclusive phylogenetic analysis..." (Dolja & Koonin, 1993). Based on RdRps, three supergroups were delineated. PEMV RNA 1 has a polymerase similar to *Luteovirus* subgroup II viruses (Demler & de Zoeten, 1991) and therefore was classified in supergroup 1 as part of the sobemo-lineage (Dolja & Koonin, 1993), while the polymerase of PEMV RNA 2 is related to viruses of the *Carmo-*, *Necro-*, *Diantho-*, *Tombusvirus* group and *Luteovirus* subgroup I (Demler *et al.*, 1993) which are all part of the carmo-lineage in supergroup 2 (Dolja & Koonin, 1993). It therefore appears that PEMV with its genomic origin from two separate supergroups is providing a live example for the hypothesis of gene block shuffling. The different origins of the two PEMV RNAs is also supported by the order of genes involved in replication. PEMV RNA 1 in addition to a viral polymerase, related to the supergroup 1 RdRp, also possesses a protease while only a polymerase, related to supergroup 2 RdRp, could be identified as part of the RNA 2 replicase. The latter is typical for other members of the newly proposed order of *Carmovirales* (Dolja & Koonin, 1993). The putative protease of PEMV RNA 1 (Demler & de Zoeten, 1991) is part of a 3C-like protease family, a chymotrypsin-related protease (Gorbalenya & Koonin, 1993). Other proteases in this family belong to the *Picorna-*, *Como-*, *Nepo-*, *Poty-*, *Sobemo-* and *Luteoviruses* subgroup II (Gorbalenya & Koonin, 1993). There is no direct experimental evidence for protease activity of *Sobemo-*, *Luteo-* and *Enamoviruses*. However, all these proteases contain a conserved catalytic triad which prompted the authors to

suggest a similar cleavage specificity for *Sobemo-*, *Luteo-* and *Enamoviruses* as is found in proteases of *Picorna-*, *Como-*, *Nepo-*, *Potyviruses* (Gorbalenya & Koonin, 1993). In all these virus groups the protease is located upstream of the polymerase. *Picorna-*, *Como-*, *Nepo-*, *Potyviruses* also contain a VPg and a helicase as part of their replicase unit, resulting in a domain order of helicase-VPg-protease-polymerase. *Sobemo-* and *Luteoviruses* lack the helicase and the location of their VPg gene is unknown. Dolja and Koonin (1993), therefore, predicted that the VPg domain in the latter viral groups is upstream of the protease. Demler *et al.* (1996a) came to the same conclusion when proposing the location of PEMV-VPg upstream of the protease.

2. *Similarities and dissimilarities between PEMV and Luteoviruses*

As mentioned above, PEMV is a virus that contains two genome parts derived from different supergroups (Dolja & Koonin, 1993). The same phenomenon is seen in *Luteoviruses*, where *Luteovirus* subgroup II has a polymerase as found in RdRp of supergroup 1 and is part of the order *Sobemovirales* in the *Picornavirata* class while *Luteovirus* subgroup I has a supergroup 2 polymerase and belongs to the order *Carmovirales* in the *Flavivirata* class (Dolja & Koonin, 1993). Similarities between PEMV (RNA 1 in particular) and *Luteoviruses* exist especially in respect to genomic organization, cytopathology, and aphid transmission. The following paragraphs are based on a summary by Demler *et al.* (1996a) and the reader is referred to that review for individual references.

The general characteristics of the virus provide a good example of both similarities and dissimilarities that provided the basis for creating a new virus group for PEMV (*Enamovirus*). *Luteoviruses* are monopartite viruses that have isometric particles containing a ss (+) RNA with a 3' OH group and a 5' VPg (e.g. Randles & Rathjen, 1995). PEMV, on the other hand, has isometric particles each encapsidating a different ss (+) RNA with a 3' OH and a 5' VPg (e.g. de Zoeten & Demler, 1995).

Sequencing of the viral genome of PEMV revealed strong similarities between RNA 1 and *Luteoviruses* subgroup II. Statistically significant sequence homology was found after comparisons of the viral polymerase, the protease domain, the cp and the cp readthrough (RT). The main differences between PEMV RNA 1 and *Luteovirus* subgroup II occurred in the 3' region of the genomic RNAs. The *Luteovirus* ORF 4 (17-19K) is missing from the PEMV genome. For PLRV, the protein product of ORF 4 was shown to bind to single-stranded nucleic acid (Tacke *et al.*, 1991) and the authors later suggested that this protein may act as the phloem-specific movement protein (Tacke *et al.*, 1993). Furthermore, the RT of PEMV is reduced by about 20 kDa and the conserved 12 nucleotide sequence surrounding the stop codon in *Luteovirsues* is not conserved in PEMV.

The differences in cytopathology can be related to different virus titers and infected tissues of *Enamo-* and *Luteovirus* groups. PEMV accumulates to higher concentration in the plant compared to *Luteoviruses* because it infects cells of all cell types and is not limited to the phloem as is the latter. However,

despite a broader range of infected tissue types, PEMV preferentially accumulates in phloem cells and sieve elements as well as in plasmodesmata connecting these cells. PEMV and *Luteoviruses* induce similar double membrane bound vesicles during the infection that contain the viral RNA. *Luteovirus* induced vesicles can fuse with the nuclear envelope and particles are sometimes seen in the nucleus. However, it was not determined whether *Luteovirus* replication is associated with the nucleus. As for PEMV, virus replication in the nucleus was unequivocally proven (see above).

Aphid transmission of both virus groups is circulative, nonpropagative, but the dynamics of these processes differ. *Luteoviruses* have a longer acquisition time and inoculation threshold and inoculation of plants only occurs during feeding of the aphid in the phloem. Inoculation with PEMV already occurs during probing of the aphid of non-phloem tissue. The fate of these viruses in the aphid vector is very similar. They are acquired by their respective vector and move through the hindgut into the hemocoel and later to the accessory salivary glands. The site of specificity in these glands, however, differs. BYDV aphid uptake is determined by the hindgut and by the plasmalemma of the salivary accessory gland, where the latter is responsible for the high level of vector specificity (Gildow & Rochow, 1980). In PEMV, the basal lamina, not the plasmalemma, of the accessory salivary gland is responsible for the exclusion of AT^- but not AT^+ strains whereas the hindgut barrier can be crossed by both strains. The uptake mechanism of both viruses also differs; *Luteoviruses* use receptor mediated endocytosis to move across the membrane barriers of the

hindgut and the accessory salivary gland, while PEMV crosses the former by simple penetration. It is not known which mechanism mediates movement of PEMV into the salivary gland system. *Luteoviruses* can only be transmitted by aphids, whereas PEMV is also mechanically transmissible. However, when plants are mechanically inoculated with RNA 1 alone, the RNA more closely related to *Luteoviruses*, no systemic infection can be established. This is viewed as an evolutionary relict of *Luteovirus* ancestry of PEMV, that was overcome by the acquisition of RNA 2.

3. *What type of complex does PEMV represent?*

Many different terms have been used to describe the complex of RNA 1 and RNA 2 of PEMV. Although often described as a bipartite virus because of its two particles, PEMV is not truly bipartite. Bipartite viruses have a single polymerase while PEMV possesses two. In addition, PEMV shows no sequence homology at the 5' and 3' ends of its genomic RNAs while in bipartite viruses similar sequences at the 3' and 5' ends of the two RNAs are thought necessary for the single viral polymerase to recognize and replicate the genome. It is not a mixed infection either which would require two independent viruses. However, RNA 1 can not systemically move in the plant nor can it be mechanically inoculated without RNA 2 which in turn can not be encapsidated and transmitted by aphids without RNA 1. Therefore, one could describe this relationship as mutualistic symbiosis between two defective viruses (Demler *et al.*, 1996a).

The relationship of the two RNAs of PEMV also shows some resemblance to the helper-dependent complexes observed between some *Luteo-* and *Umbraviruses*. Due to transencapsidation, a mechanically transmitted virus gains aphid transmissibility when both viruses are acquired simultaneously (Falk & Duffus, 1981). The *Umbravirus* replicates independently and is capable of systemic infections but lacks a cp. The *Luteovirus* replicates also independently but is aphid transmissible because of the ability to express a cp. PEMV RNA 1 represents the luteoviral part and PEMV RNA 2 the umbraviral part of the complex. However, since RNA 1 is not an independent virus and it is not phloem limitated in the complex this relationship does not fit the accepted definition of a helper-dependent complex. PEMV, therefore, can not be classified in the existing categories without broadening their definitions.

Evolutionarily PEMV seems most closely related to the groundnut rosette complex, made up of groundnut rosette assistor virus (GRAV), a subgroup II *Luteovirus*, groundnut rosette virus (GRV), an umbravirus, and a satellite virus. The satellite depends on GRV for multiplication and both, GRV and the satellite, depend on GRAV for aphid transmission (Demler *et al.*, 1996a). RNA 1, related to subgroup II *Luteoviruses*, resembles GRAV while RNA 2, related to *Umbraviruses*, is similar to GRV. Like the satellite of the groundnut rosette complex, the satellite of PEMV also depends on the umbraviral part for replication while the satellite and RNA 2 require the luteoviral part for encapsidation and aphid transmission. The evolutionary relationship between PEMV and the groundnut disease complex was further strengthened when

Demler *et al.* (1996b) showed sequence similarities between the satellites of GRV and PEMV and that GRV and PEMV could replicate each others satellite.

VPg (viral protein, genome-linked)

1. *General characteristics*

VPgs are small proteins covalently attached to the 5' end of many single-stranded, positive sense, viral RNAs. To date two animal virus families, *Picorna-* and *Caliciviruses*, and six plant virus groups, *Enamo-*, *Luteo-*, *Como-*, *Nepo-*, *Poty-* and *Sobemoviruses* possess VPgs (for review see Vartepetian & Bogdanov, 1987). This would be expected since all of the above mentioned virus groups belong to the *Picornavirata* with *Luteovirus* and PEMV representing a link to the *Flavivirata* (Dolja & Koonin, 1993). It was reasoned that protein-primed RNA replication mechanisms (see below for possible VPg functions) only arose once, early in evolution (Dolja & Koonin, 1993). As mentioned earlier the domain localization of VPg, upstream of a 3C-like protease, is identical for *Picorna-*, *Como-*, *Nepo-* and *Potyviruses* (Dolja & Koonin, 1993; Gorbalenya & Koonin, 1993). Despite sequence similarities within a group, no common sequence elements were found, hampering efforts to predict the VPg localization of *Sobemo-* and *Luteoviruses* (Gorbalenya & Koonin, 1993).

VPgs often run anomalously on SDS-PAGE gels due to their cationic character (Vartepetian & Bogdanov, 1987). The electrophoretically derived molecular weights, therefore, frequently predict much larger proteins than protein sequencing reveals. The sizes of the VPgs range primarily between 20

and 30 amino acids (aa) (e.g. Wimmer, 1982; Mayo & Fritsch, 1994; Jaegle *et al.*, 1987; Drygin *et al.*, 1987). VPgs are not only found on viral genomes but also on at least two satellite RNA (Meyer *et al.*, 1984; Koenig & Fritsch, 1982; Liu *et al.*, 1991). The aa involved in the covalent linkage of the protein with the RNA can be either tyrosine, serine or threonine due to the free hydroxyl group in their sidechains. Tyrosine is involved in *Picornaviruses* (for reviews see Wimmer, 1982; Vartepetian & Bogdanov, 1987) and *Potyviruses* (Murphy *et al.*, 1991; Jayaram *et al.*, 1992; Vance *et al.*, 1992) while serine has been implicated for viruses belonging to *Como-* (Jaegle *et al.*, 1987; Drygin *et al.*, 1987) and *Nepoviruses* (Pinck *et al.*, 1991; Margis *et al.*, 1993; Zalluoa *et al.*, 1996). Threonine, although theoretically possible, has not yet been found as an aa involved in RNA linkage. The function of VPgs remains obscure despite intensive research especially with poliovirus. Possible functions include: primer of RNA synthesis, signal for virion encapsidation, and in some instances an involvement in infectivity (e.g. Vartepetian & Bogdanov, 1987). The VPg is needed for infectivity of *Sobemo-*, *Nepo-*, *Poty-* and *Caliciviruses* but removal of the protein from the genomes of *Picorna-*, *Como-*, *Enamo-*, *Luteoviruses* has no effect on infectivity (for review see Vartepetian & Bogdanov, 1987). Since most research on the function of VPg has involved poliovirus, I will provide an overview of poliovirus replication and the involvement of its VPg in the viral life cycle before summarizing data obtained with the VPgs of plant viruses.

2. *Poliovirus VPg*

Poliovirus VPg is a basic, acid-soluble, 22 aa protein (for review see Wimmer, 1982). The basic amino acids are located in the center and C-terminal portion of the protein (Vartepetian & Bogdanov, 1987). Tyrosine at position 3 from the N-terminus is linked via its O^4-hydroxyl to the terminal phosphate of uracil, the most 5' terminal nucleotide of the RNA (for review see Wimmer, 1982).

The presence and absence of VPg influences several stages of the poliovirus life cycle. Poliovirus VPg is not needed during very early stages (e.g. entry and uncoating) since removal of the protein has no influence on infectivity of the RNA (Flanegan *et al.*, 1977). The protein is found on (+) and (-) ss RNA, on replicative forms and replicative intermediates as well as on nascent strands suggesting a role in viral replication (Wimmer, 1982). It is not necessary for translation and a 27kDa unlinking enzyme of the host is involved in the removal of VPg from RNA to become mRNA (Vartepetian & Bogdanov, 1987). However, it is again needed during virion formation. Only RNA linked to a VPg is found in virions indicating that it might serve as a signal determining encapsidation of RNA (Nomoto *et al.*, 1977a).

Two different approaches have been used to elucidate the mechanisms underlying poliovirus replication. The first one uses isolated crude membranous RNA replication complexes from virus infected HeLa cells. This is mainly used to study the synthesis of (+) strand RNA from (-) strands (Ehrenfeld & Richards, 1989). A second approach uses *in vitro* studies to elucidate primarily (-) strand from (+) strand RNA synthesis. Components involved in poliovirus replication

have been isolated and purified and individual steps were reconstituted (Ehrenfeld & Richards, 1989). The ultimate goal to construct a complete *in vitro* replication system has not been achieved and caution should be exercised when analyzing *in vitro* results (Hey *et al.*, 1987).

Different poliovirus replication models have been proposed but each takes the strict template and primer dependence of the viral polymerase $3D^{pol}$ into account (Flanegan & Baltimore, 1979; van Dyke & Flanegan, 1980). Also, the formation of membranous vesicles and their involvement in viral replication is well documented (for review see Mirzayan & Wimmer, 1994). One model states that positive strand RNA synthesis occurs in a membranous environment where uridylylated VPg serves as primer for RNA elongation by the viral polymerase $3D^{pol}$ (for review see Mirzayan & Wimmer, 1994). Evidence showing the involvement of VPg-pU as a primer is: I) VPg is attached to the 5' end of genomic, (+), (-) and nascent RNA but not to mRNA (Nomoto *et al.*, 1977a; Pettersson *et al.*, 1978; Flanegan *et al.*, 1977). II) Antibodies (Abs) against VPg (anti-VPg Abs) immunoprecipitate VPg and VPg-pUpU (Crawford & Baltimore, 1983). III) Anti-VPg Abs specifically inhibit initiation of viral RNA synthesis *in vitro* (Baron & Baltimore, 1982). Data indicating the importance of membranes during replication led to the discovery of a VPg (3B) precursor protein 3AB, a 12kDa protein with a hydrophobic stretch of 22 aa in the 3A portion. This hydrophobic stretch is involved in attaching 3AB to membranes of the replication complex and delivering VPg to its site of function. The viral proteinase $3C^{pro}$ cleaves 3AB during or after (+) RNA elongation. This releases the VPg and the

associated viral RNA from the membrane (for review see Mirzayan & Wimmer, 1994).

Negative strand synthesis involves the uridylylation at the 3' poly A of the (+) RNA by host factor (HF), resulting in the formation of a polyA - polyU hairpin that serves as a primer. The attachment of VPg to the (-) strand takes place by a self-catalyzed transesterification reaction that concomitantly cleaves the hairpin and separates both strands (Mirzayan & Wimmer, 1994). Two main observations support this model. One, negative strand covalently linked to (+) strand RNA was seen after *in vitro* synthesis in the presence of viral polymerase and HF (Young *et al.*, 1986; Young *et al.*, 1985). Two, HF has a terminal uridylyltransferase activity that stimulates initiation of RNA synthesis (Andrews & Baltimore, 1986; Andrews *et al.*, 1985).

The evidence obtained for poliovirus replication led people to postulate the following two functions of VPg. First, VPg or uridylylated VPg serve as primer for (+) strand RNA synthesis. Second, VPg has a nucleolytic activity that cleaves the RNA hairpin during (-) strand synthesis and a linkage activity to attach itself to the 5' end of the (-) RNA. The strongest evidence against the essential role of VPg during RNA replication comes from mutational experiments in which a VPg mutation was replication positive but nonviable. This suggests that an essential function other than viral RNA replication was affected. As indicated earlier, the VPg could also serve as a signal for encapsidation although no direct experimental evidence for this is available. None of these functions could be positively verified and other functions are possible as well.

3. *VPgs of plant viruses*

The plant viruses closely related to poliovirus and the *Picornaviruses* are *Como-*, *Nepo-* and *Potyviruses* which are combined in the superfamily of picorna-like plant viruses (Matthews, 1991). A comparison of genome organization and nucleotide sequence led to the following common features of a (+) ssRNA genome, a 5' VPg and a 3' poly (A) tail. The viral expression strategy is via a genome containing a long single ORF that codes for a polyprotein which is processed by viral proteases. Several nonstructural proteins of similar function also share significant sequence similarity and these conserved proteins are arranged in the order of membrane binding region - VPg - protease - RdRp (for review see Matthews, 1991). The following paragraphs will provide a discussion on the knowledge of VPgs from each group of the picorna-like viruses.

VPgs have been identified on the *Comoviruses* CPMV, cowpea severe mosaic virus (CPSMV) and red clover mottle virus (RCMV) (for review see Mayo & Fritsch, 1994), radish mosaic virus (RaMV) (Drygin *et al.*, 1987), squash mosaic virus (SqMV) and Echtes Ackerbohnenmosaik Virus (EAMV) (Daubert & Bruening, 1979). Protein sequencing of VPgs from CPMV, CPSMV and RCMV revealed basic, hydrophilic proteins of 28 aa with a relatively similar sequence (Mayo & Fritsch, 1994). The linkage to the viral RNA in the case of the 28 aa CPMV-VPg and the 36 aa RaMV-VPg was a β-OH (serine) linked via a phosphodiester bond to the 5' terminal phosphate of the last nucleotide of the

viral RNA (Jaegle *et al.*, 1987; Drygin *et al.*, 1987, respectively). CPMV, the type member of *Comoviruses*, was the first plant virus for which the VPg was identified (Daubert *et al.*, 1978). However, data obtained from poliovirus VPg nevertheless guide the experimental analysis of all VPg functions.

CPMV-VPg has a positive charge of +5. If VPg functions as a primer during CPMV RNA replication as was proposed for poliovirus, the positive charge could enable the protein to interact with the negatively charged viral RNA (Jaegle *et al.*, 1987). Not all 5' ends of RF are bound to VPg (Lomonossoff *et al.*, 1985). It was hypothesized that if the VPg is initially present and later removed, those (+) strands could serve as mRNA, in analogy to poliovirus where the mRNA lacks a VPg. Also, like *Picornaviruses*, the VPg was not required for infectivity or *in vitro* translation (Stanley *et al.*, 1978). Early research with anti-VPg Abs indicated that CPMV-VPg was part of a membrane bound 60kDa precursor molecule (Zabel *et al.*, 1982; Goldbach *et al.*, 1982; Wellink *et al.*, 1986) that was responsible for bringing the VPg to the replication complex. More recent evidence indicates a 112kDa VPg precursor protein is present during the initiation of replication (Peters *et al.*, 1995). It is unknown whether the precursor is directly involved, or whether involvement is via a 26kDa protein (VPg + 24kDa protease) derived from this precursor. Free VPg is rapidly degraded or undetectable (de Varennes *et al.*, 1986; Peters *et al.*, 1995; Peters *et al.*, 1992) but viral RNA of as few as 17 nucleotides linked to VPg provided protection from proteolytic degradation (de Varennes *et al.*, 1986).

Nepoviruses are closely related to *Comoviruses* and are classified within the picorna-like plant virus superfamily as *Comoviridae*. The *Comoviridae* have isometric particles with a bipartite genome. Like *Comoviruses*, the VPg is encoded by the larger RNA (RNA 1) (Mayo *et al.*, 1982a), and it plays no role in translation (Koenig & Fritsch, 1982; Chu *et al.*, 1981), and covalent linkage of the VPg occurs via the serine β-OH to the 5' terminal phosphate of the viral RNA (Pinck *et al.*, 1991; Margis *et al.*, 1993; Zalluoa *et al.*, 1996). However, unlike *Comoviruses*, infectivity is greatly reduced or eliminated by protease treatment (Mayo *et al.*, 1982a; Harrison & Barker, 1978; Chu *et al.*, 1981). The size of six different nepoviral VPgs is around 4kDa on SDS-PAGE (Mayo *et al.*, 1982a). Microsequencing revealed a 24aa VPg for grapevine fanleaf virus (GFLV) (Pinck *et al.*, 1991) and 27aa VPg in case of tomato black ring virus (TBRV) (Mayo & Fritsch, 1994). A 26 aa VPg sequence for tomato ringspot virus (TomRSV) (Mayo & Fritsch, 1994) and the 29 aa sequence of grapevine chrome mosaic virus (GCMV) were inferred based on strong sequence similarities between the RNA 1 polyproteins. Despite this, nepoviral VPgs show relative sequence diversity and only the insertion of gaps led to a consensus (Mayo & Fritsch 1994). Mutations in the 5' end of satellite RNA from cDNA clones of arabis mosaic virus (ArMV) were corrected after infection of plants providing indirect evidence that VPg - 5' terminal RNA bases might serve as a primer in (+) strand RNA synthesis (Liu *et al.*, 1991). A previously unknown possible VPg function emerged from translation studies of the GFLV polyprotein. Cleavage by the viral encoded 24kDa protease (Pro) is two times more efficient after cleavage of the

VPg-Pro polyprotein and release of the VPg. It was therefore postulated that the VPg could modulate the cleavage activity of the 24kDa Pro and ultimately regulate protein expression (Margis *et al.*, 1994).

Potyviruses like *Como-* and *Nepoviruses* belong to the same superfamily but unlike *Como-* and *Nepoviruses* they are filamentous with a monopartite ss (+) RNA genome (Matthews, 1991). The VPg has been identified and localized to the N-terminus of the small nuclear inclusion protein NIa for tobacco etch virus (TEV) (Dougherty & Parks, 1991), tobacco vein mottling virus (TVMV) (Shahabuddin *et al.*, 1988) and turnip mosaic virus (TuMV) (Laliberte *et al.*, 1992). The C-terminus of this protein is a serine-like 3C protease that can utilize a suboptimal internal cleavage site of NIa to release the protease from the VPg (Laliberte *et al.*, 1992; Carrington *et al.*, 1993). The VPg is linked via a phosphodiester bond of a tyrosine hydroxyl group to the viral RNA (Murphy *et al.*, 1991; Jayaram *et al.*, 1992; Vance *et al.*, 1992) and mutations of the conserved tyrosine results in a lethal phenotype (Murphy *et al.*, 1996). Like *Nepoviruses* the infectivity of the RNA decreases with the removal of the protein (Riechmann *et al.*, 1989). The size of the VPg is between 22kDa and 24kDa on SDS-PAGE (Riechmann *et al.*, 1989; Laliberte *et al.*, 1992; Murphy *et al.*, 1990; Shahabuddin *et al.*, 1988). A 21kDa (189 aa) VPg for tobacco etch virus (TEV) was deduced from computer sequence analysis (Dougherty & Parks, 1991). *Potyvirus* VPgs are much larger than any other identified VPgs. This might be due to the multifunctional activities of the NIa precursor protein. In addition to the VPg, a nuclear localization signal has been identified within the N-terminus

of NIa (Carrington *et al.*, 1993; Schaad *et al.*, 1996). Both the 49kDa NIa proteinase and a 24kDa protein, the N-terminal part of the proteinase, were found attached to the viral RNA (Murphy *et al.*, 1990) indicating that cleavage is not required for VPg activity (Carrington *et al.*, 1993). However, attachment of either protein to the RNA might not be sufficient for VPg function as it is believed that the functional VPg form includes an additional 6kDa protein that serves as a membrane anchor like the 3A protein of poliovirus (Restrepo-Hartwig *et al.*, 1994).

Sobemoviruses are related to *Luteoviruses* (Koonin, 1991) and both groups belong with the picorna-like plant viruses to the *Picornavirata* (Dolja & Koonin, 1993). Members of the two groups have isometric particles with a monopartite, ss (+) RNA genome that contains a 5' VPg but is not polyadenylated at the 3' end (Matthews, 1991). Based on electrophoretic mobility, the type member of the *Sobemoviruses*, southern bean mosaic virus (SBMV), has a 12kDa VPg (Ghosh *et al.*, 1979). Sequencing of the genomes of rice yellow mottle virus (RYMV) (Yassi *et al.*, 1994) and cocksfoot mottle virus (CfMV) (Mäkinen *et al.*, 1995), led to the discovery of a conserved VPg motif that is located upstream of the protease and polymerase in the second ORF. The aa involved in the protein-RNA linkage is unknown but Yassi *et al.* (1994) proposed the involvement of a conserved tyrosine residue in analogy with poliovirus data.

Luteoviruses are divided into two subgroups which are distinguished based on the genomic organization and sequence similarity (e.g. Martin, 1990). They contain viruses related not only to the *Picornavirata* (subgroup II) but also

to the *Flavivirata* (subgroup I) (Dolja & Koonin, 1993). VPgs have been found at the 5' end of RNAs of subgroup II viruses. The RPV isolate of barley yellow dwarf (BYDV) and potato leafroll virus (PLRV) have a 17kDa and 7kDa VPg, respectively, attached to the 5' end of their genome (Murphy *et al.*, 1989; Mayo *et al.*, 1982b, respectively). No VPg has been found attached to the RNA of a subgroup I *Luteovirus*. This is not unexpected since no virus of the *Flavivirata* class contains this protein. Like poliovirus, *Como-* and *Potyviruses*, the protein is not required for infectivity of the RNA (Mayo *et al.*, 1982b). No protein sequence has been determined and the aa involved in the VPg linkage to the RNA as well as the precise genomic localization of this protein are unknown. However, hypotheses regarding the genomic localization of the VPg have been proposed. The first complete aa sequence of a *Luteovirus* was determined for subgroup I BYDV-PAV (Miller *et al.*, 1988). Since ORF 4 has the potential to encode a 17kDa protein and the M_r of BYDV-RPV VPg was of the same size, Miller (1994) proposed that the VPg is encoded in ORF 4. A 17K ORF 4 was also found after sequencing of PLRV. The authors suggested proteolytic cleavage was the most logical explanation to derive the 7kDa PLRV-VPg (van der Wilk *et al.*, 1989). However, more recent evidence points to ORF 1 as the VPg coding region (Miller *et al.*, 1995, Mayo & Ziegler-Graff, 1996). Frameshift mutations of BWYV-ORFs were introduced and RNA transcripts used to infect protoplasts. Only mutations in ORF 1 and 2 are lethal (Reutenauer *et al.*, 1993). The essential nature of the two ORFs points to their involvement in RNA replication, a function also proposed for VPg. Also, subgroup II

Luteoviruses show a protease motif in ORF 1 and a polymerase motif in ORF 2 and are part of the picornavirus-like superfamily which possesses a replicase with the domain order of VPg-protease-polymerase (Dolja & Carrington, 1992; Koonin & Dolja, 1993) placing the VPg in ORF 1.

4. *Rationale for project*

As mentioned earlier, *Enamoviruses* contain a divided genome with RNA 1 related to subgroup II *Luteoviruses* and *Picornavirata* and RNA 2 to the *Carmovirales* in the *Flavivirata* (Dolja & Koonin, 1993). PEMV-VPg was identified on SDS-PAGE gel by iodination as a 17.5kDa protein which was covalently attached to the 5' end of the viral genome (Reisman & de Zoeten, 1982). It is unknown whether the protein is attached to both RNAs. The actual size of the protein is most likely smaller as other VPgs have turned out to be smaller than the size predicted by SDS-PAGE (for review see Vartepetian & Bogdanov, 1987). No ORF of 17K / 18K or smaller was identified within the PEMV genomic sequence (Demler & de Zoeten, 1991). The protein is not required for infectivity since *in vitro* RNA transcripts were successfully used for mechanical inoculations (Demler *et al.*, 1996a). It was speculated that the VPg is localized upstream of the protease motif of RNA 1 as part of the conserved domain arrangement of membrane binding - VPg - protease - polymerase in analogy with *Picorna-* and picorna-like viruses (Demler *et al.*, 1996a). However, the actual size and the genomic localization of the VPg and the aa involved in the protein-RNA linkage is not known.

The goal of my project was to isolate and characterize the PEMV-VPg and to obtain protein sequence to determine the genomic localization of the VPg gene.

Chapter 2

Genomic localization of the 5'-linked protein (VPg) of Pea Enation Mosaic *Enamovirus*

Introduction

Pea enation mosaic virus (PEMV) is the sole member of the *Enamovirus* genus (de Zoeten & Demler, 1995). Its genome consists of two taxonomically distinct, single-stranded, (+) sense RNAs. RNA 1 (5706 nucleotides) shares significant homologies with subgroup II *Luteoviruses* in the arrangement and sequences of RNA polymerase, putative protease, coat protein (cp), and coat protein readthrough domain (Demler & de Zoeten, 1991; Demler *et al.*, 1996a). RNA 2 (4253 nucleotides) encodes a polymerase, distinct from the polymerase of RNA 1, that reveals a closer relationship to members of the *Luteovirus* subgroup I as well as the *Umbra-*, *Carmo-*, *Tombus-*, *Diantho-* and *Necroviruses* (Demler *et al.*, 1993). Despite the independent replication of RNA 1 and RNA 2, both are required for wild-type infections. RNA 1 provides for encapsidation and vector transmission while RNA 2 facilitates systemic invasion and mechanical transmission (Demler *et al.*, 1994).

Many single-stranded viral RNAs, including PEMV (Reisman & de Zoeten, 1982), encode, and are covalently linked to a small protein (VPg) at the 5' terminus (for reviews see Wimmer, 1982; Vartapetian & Bogdanov, 1987 and

references therein). The function of VPgs remains under investigation but experimental data from poliovirus has led to several hypothesis. Poliovirus VPg was proposed to act as a primer for RNA synthesis (*e.g.* Nomoto *et al.*, 1977a; Mirzayan & Wimmer, 1994). Poliovirus RNA polymerase is strictly primer-dependent (Van Dyke & Flanegan, 1980) and antibodies (Abs) raised against the poliovirus VPg inhibit RNA replication (Baron & Baltimore, 1982; Barton *et al.*, 1995). In addition, poliovirus VPg may also play a role in virus encapsidation since only poliovirus-RNA that is covalently linked to the VPg is packaged (Nomoto *et al.*, 1977a; Nomoto *et al.*, 1977b).

The nucleotide sequence of the PEMV genome (Demler & de Zoeten, 1991) did not reveal a discrete open reading frame (ORF) corresponding to the 17.5 kDa PEMV-VPg reported by Reisman & de Zoeten (1982). However, the predicted amino acid sequence of the 84K protein of RNA 1 ORF 1 revealed a hydrophobic region near the N-terminus of the protein that is characteristic of membrane spanning proteins. The central region of this 84K protein is similar to the catalytic site of a serine-based member of the 3C-like proteinases although no proteinase activity had been reported for PEMV or any *Luteovirus*. The N-terminus and the central part are separated by a very basic region, a characteristic common to VPgs (Demler & de Zoeten, 1991). This linear arrangement of transmembrane region - proteinase - polymerase domains in PEMV was similar to *Picorna*-like viruses (membrane binding domain - VPg - proteinase - polymerase) (Demler & de Zoeten, 1991; Domier *et al.*, 1987, respectively) which utilize proteinases in protein processing. Based on this

analogy, Demler *et al.* (1996a) suggested that the PEMV-VPg gene is encoded upstream of the protease motif.

Several members of the *Luteovirus* subgroup II possess a VPg (Mayo *et al.*, 1982; Murphy *et al.*, 1989). However, attempts to determine ORFs encoding these VPgs led to conflicting and inconclusive results. Miller *et al.* (1988) and Miller (1994) proposed that the VPg may be encoded by the 17K ORF 4, since a 17 kDa VPg was reported for the RPV isolate of barley yellow dwarf virus (BYDV-RPV) (Murphy *et al.*, 1989). Similarly, van der Wilk *et al.* (1989) inferred that the 7kDa VPg of potato leafroll virus (PLRV) is produced via proteolytic processing of the PLRV-ORF 4 product. Experimental proof for both conclusions is lacking and subsequent data prompted researchers to propose the involvement of ORF 1 (Mayo & Ziegler-Graff, 1996; Miller *et al.*, 1995) in the production of VPg. The similarities of PEMV RNA 1 to subgroup II *Luteovirus* RNAs and the relatively high yield of purified virus from peas make PEMV an attractive system to address questions surrounding the genomic origin of the VPg.

In this study we report for the first time the amino acid (aa) composition of PEMV VPg. Subsequent computer sequence alignments mapped the VPg to the 84K ORF of RNA 1, downstream of the protease motif. This provides the first indication for proteolytic processing as part of the PEMV replication process.

Materials and Methods

Virus propagation and purification:

The naturally occurring, aphid-nontransmissible "deletion" strain PEMV-AT-Δ (Demler *et al.*, 1996a) was propagated in *Pisum sativum* L. cv. 8221 by mechanical inoculation of the lower two sets of leaves of 9 to 11 day old seedlings. Pea plants (*Pisum sativum*) were grown from seeds in a growth chamber with a 16 hour light regime, daytime temperature of 22°C and night temperature of 18°C. PEMV infected fresh leaves (2g) or dried leaves (0.5g) were ground in 10ml of 0.1M phosphate buffer pH 6.5 with celite (diatomaceous earth) in a chilled mortar and pestle and used for mechanical inoculation.

Infected plants were harvested ten to twelve days post-inoculation and PEMV virions were purified according to a protocol from German & de Zoeten (1975). All extraction equipment was placed at 4°C prior to the virus isolation and solutions were made from a stock solution of 2M sodium acetate (NaOAc) buffer, pH 6.0 (total volume of 2 liters: 230 ml of glacial acetic acid [17.4M] were added to 1600 ml of dH_2O in a beaker on ice, 130g of NaOH were slowly added under constant stirring and the pH was adjusted with 10N NaOH). Plants were harvested by cutting them at the soil level with a razor blade and the tissue weight was determined. A volume of extraction buffer (0.2 M NaOAc, pH 6.0) equal to 1.5 times the weight of the tissue was added to the harvested tissue in a commercially available Waring blender and well homogenized. One volume of chloroform, based on the weight of the tissue, was added and the tissue was homogenized for one minute at high speed. The homogenate was poured into 250ml centrifuge bottles(Beckman Instruments Inc., Palo Alto, CA), placed in a

JA 14 rotor (Beckman) and centrifuged at 4°C for 20 minutes at 7000rpm (7K) in a Beckman J2-21 centrifuge. The upper layer was then decanted into a 1l flask on ice and dialyzed overnight in 10 l of 0.2M NaOAc in the cold room (4°C). On the following day, the solution was centrifuged in 250ml centrifuge bottles in a JA-14 rotor at 4°C and 7K for 20 minutes. The supernatant was then centrifuged in a Beckman L-70 ultracentrifuge in 26.3ml polycarbonate tubes with 3 piece cap assembly (Beckman, No. 355616) in a 50.2Ti rotor (Beckman) at 4°C and 30K for 3 hours. The resulting supernatant was decanted and discarded. Two milliliters of 5% sucrose in 0.1M NaOAc buffer were added to the pellet to resuspend it overnight on a shaker (Gyrotory shaker model G2, New Brunswick Scientific Co., Inc., Edison, NJ) in the cold room. The resuspended pellets were combined the next day in 30ml glass Corex tubes (Corning Incorporated, Corning, NY) and spun in a JA-20 rotor (Beckman) at 4°C and 7K for 20 minutes in a Beckman J2-21 centrifuge. The high speed centrifugation and the resuspension of the pellets were repeated. The virus concentration was calculated after determining the absorbency of a 1:100 dilution of an virus aliquot with a Beckman DU-64 spectrophotometer at 260nm, using an extinction coefficient of 7.5mg/ml. Linear sucrose gradients of 10% to 40% were prepared from stock solutions of 10% and 40% sucrose in 0.1M NaOAc buffer in Ultra-clear centrifuge tubes (25x89mm, Beckman) and 2-20 mg of virus were loaded. The gradients were spun in a SW 28 rotor (Beckman) at 4°C and 25K for 3 hours in a Beckman L-70 ultracentrifuge and fractionated afterwards using an ISCO density gradient fractionator, model 185, (ISCO Inc., Lincoln Nebraska) and the

ISCO UA-5 absorbance/fluorescence detector. PEMV peaks were collected, combined and centrifuged overnight in a 50.2Ti rotor (Beckman) at 4°C and 50K rpm. The resulting pellets were resuspended in diethyl pyrocarbonate (DEPC) - treated 20mM Tris-HCl, pH 7.2 and used for viral RNA isolation.

Viral RNA isolation and anti-VPg antibody production:

Viral RNA was purified according to a protocol from Reisman & de Zoeten (1982). Special precautions were taken to prevent contamination with RNases. All steps were carried out in 15ml Corex tubes (Corning) that were autoclaved and baked for 8 hours at 180°C in an oven (Isotemp oven, model 655G, Fisher Scientific, Pittsburgh, PA). The solution containing the resuspended pellet was brought to 2% SDS, 20mM EDTA, mixed and placed at 68°C for 5 min. Two extractions with an equal volume of 10mM Tris-HCl pH 7.2 saturated phenol each followed. Each time the samples were vortexed vigorously and centrifuged for 10 min at 7K rpm in a JA-20 rotor (Beckman). To increase RNA yield, the phenol phase of the first extraction was back-extracted with 10mM Tris-HCl pH 7.2. The aqueous phase was collected from each extraction, pooled, extracted twice with chloroform - isoamyl alcohol (24:1) and later brought to 0.2M NaOAC. To each sample, 2.5 volumes of -20°C, concentrated ethanol were added and the RNA was precipitated at -80°C overnight. The precipitate was pelleted in a JA-20 rotor (Beckman) at 4°C and 7K for 30 min. The RNA was washed by adding 2ml of chilled, 70% ethanol and centrifuging for 10 min. The pellet was resuspended in 500µl of DEPC-treated dH_2O. Three additional phenol-

chloroform extractions followed by ethanol precipitation were performed to remove noncovalently attached proteins. This extensivly purified RNA was then used for antibody production.

To obtain anti-VPg antibodies (Abs), approximately 250μg of precipitated, extensivly purified RNA was resuspended in 400μl of DEPC - treated dH_2O and used for subcutaneous injections of a rabbit. The RNA was emulsified in an equal volume of Freunds complete adjuvant (Sigma Chemical Company, St. Louis, MO) for the first injection and Freunds incomplete adjuvant (Sigma) for booster injections. The care of the rabbit, administration of injections and drawing of blood was performed by the personnel of the University Laboratory Animal Resources (ULAR) at Michigan State University (MSU), East Lansing, MI. Booster injections were administered 28, 46 and 56 days after the first injection. Immune serum was collected before each booster injection and Abs were concentrated following standard protocols (Harlow & Lane, 1988). The obtained blood was placed at 37°C for 1 hour prior to an overnight incubation at 4°C and the serum was separated from the clotted blood the next day by centrifugation in a table top centrifuge (IEC HN-SII centrifuge, International Equipment company, Needham, MA) for 10 min at 1800 x g. The serum was removed and placed into eppendorf tubes in 250 μl aliquots and stored at -20°C. The Abs were tested in Western blots of total protein extracts from infected plants to determine the specificity and strength of the Abs. Proteins were separated on 15% sodium dodecyl sulfate - polyacrylamide gels (SDS-PAGE) and electroblotted onto Imobilon-P membranes (Millipore, Maryland) prior to the

Western blotting protocol (see below). Only anti-VPg Abs collected 56 days after the first injection were used in the described experiments.

Production of RNA transcripts and transcript inoculation of plants:

Pea plants used for total protein isolation and characterization of the putative anti-VPg Abs were infected with transcripts of RNA 1 and RNA 2, RNA 2 alone or mock inoculated. Transcripts of RNA 1 and RNA 2 were prepared from the cDNA clones PEMV-AT$^+$ (Demler *et al.*, 1996a) and pPER2d (Demler *et al.*, 1994a), respectively. Ten μg DNA of a clone was linearized for each transcription reaction, with Pst I (Boehringer Mannheim, Indianapolis, IN) for RNA 1 or Sma I (Boehringer Mannheim) for RNA 2, during a 60 min incubation at 37°C. Care was taken during the reactions to prevent contamination with RNases. For each transcription reaction 165 μl of dH_2O, 100 μl of 5x optimized transcription buffer (Promega Corporation, Madison, WI), 50 μl of 0.1M dithiothreitol (DTT), 50μl of nucleotide triphosphates (4mM ATP, CTP, UTP, each and 0.75mM GTP), 125 μl 2mM 7-methylguanosine (CAP) and 1μl (80ng) of T7 polymerase (Promega) were added to linearized, phenol-chloroform extracted and ethanol precipitated DNA. Following a 45 minute incubation at 37°C, 4 μl of 10mM GTP and 1μl (80ng) of T7 polymerase (Promega) were added and the incubation continued for another 45 min. The DNA template was destroyed by addition of 5 units (U) RQ1 DNase (Promega) during a 15 min incubation at 37°C. The RNA was then extracted twice with phenol-chloroform and precipitated at -20°C for at least one hour with 50 μl of 3M NaOAc and 1ml

chilled ethanol. The precipitated RNA was pelleted at 14000 (14K) rpm in an eppendorf centrifuge 5415C (Brinkmann Instruments, Inc., Westbury, NY) for 20 min, washed with chilled, 70% ethanol and afterwards resuspended in 50 µl DEPC-treated dH_2O. The size and integrity of RNA transcripts was assessed on non-denaturing 1% agarose gels (for detailed protocol see appendixA).

Seven day old pea seedlings were inoculated with various combinations of infectious transcripts or mock inoculated as previously described (Demler *et al.*, 1993). Each inoculum contained 5 to 10 µg of viral RNA transcripts suspended in an equal volume of buffer (10mM Tris-HCl, pH 8.0, 1mM EDTA, 1% w/v celite, 1% w/v bentonite). The seedlings were dusted with 600 mesh carborundum (Fisher Scientific) prior to rub-inoculation. Plants were harvested 10 to 14 days post-inoculation for total protein isolation.

Total protein isolation:

Five leaves from individual pea plants were ground in 750µl of homogenization buffer (700mM sucrose, 500mM Tris-HCl pH8.0, 50mM EDTA, 100mM KCl, 2% β-mercaptoethanol, 2mM phenylmethylsulfonyl fluoride [PMSF] in isopropanol). The homogenate was centrifuged in an eppendorf centrifuge (Brinkman Instruments, Inc.) for 5 minutes at 14K rpm. The supernatant was collected and the protein concentration was determined using the Bio-Rad Protein Assay Kit I (Bio-Rad Laboratories, Hercules, CA). Based on the protocol of the manufacture, dilutions of 0 - 20 µg of bovine gamma globulin (BGG) in 0.8ml of dH_2O were prepared to create a standard curve. Diluted

samples were prepared by adding 1μl and 10μl of each sample to dH_2O to obtain a final volume of 0.8ml. Dye reagent concentrate (0.2ml) (Bio-Rad) was added to all samples and standards and solutions were mixed by gentle inversion of each test tube. The OD_{595} was measured after 5 min in a spectrophotometer (Beckman). A standard curve was created by plotting absorbency versus concentration of standards. The concentration of each sample was then determined from this curve.

Purification and biotinylation of anti VPg-IgGs:

In order to increase the detection limit, anti-VPg IgGs were isolated and biotinylated. Anti-VPg IgGs were purified using protein A low salt columns (Harlow & Lane, 1988). A column was made by filling 200mg protein A beads (Pharmacia Biotech Inc., Piscataway, NJ) dissolved in 100mM Tris-HCl pH 8.0 into a disposable chromatography column (BioRad) and washing the column extensively with the 100mM Tris-buffer. The pH of the serum was adjusted to 8.0 by adding $1/10^{th}$ volume of 1.0M Tris-HCl pH 8.0. After application of 1ml serum, pH 8.0, to the column, washes with 10 column volumes of 100mM Tris-HCl, pH 8.0, and with 10 column volumes of 10mM Tris-HCl, pH 8.0, followed. IgGs were eluted from the column by applying 100mM glycine pH 3.0 in 500μl aliquots. The eluate was collected in eppendorf tubes containing 50 μl of 1M Tris-HCl pH 8.0 to adjust the pH to neutral. The IgG containing fractions were identified by absorbance at 280nm (1OD=approximately 0.8 mg/ml) in a spectrophotometer (Beckman).

Anti-VPg IgGs were biotinylated in solution using the Amersham LifeScience ECL-protein biotinylation module. The provided 20x bicarbonate buffer, pH 8.6, was diluted 1:20 by adding 2.5µl of buffer to 50µl (55µg) of anti-VPg IgGs. After addition of 3µl biotinylation reagent the solution was mixed well and incubated for one hour at room temperature (RT) under constant agitation (eppendorf mixer 5432, Brinkmann Instruments, Inc.).

Unbound biotin was removed first by incubating the reaction mixture with 1/10th volume of 1M Tris, pH 8.0, at room temperature for 15 min, followed by purification of labeled Abs on a Sephadex G-25 spin column. The Sephadex G-25 resin (medium grade, Sigma) was equilibrated overnight at 4°C in an excess volume of phosphate-buffered saline pH 7.5 (PBS; 80mM di-sodium hydrogen orthophosphate anhydrous, 20mM sodium dihydrogen orthophosphate, 100mM sodium chloride) containing 0.01% Thimerosal (mercury-[(o-carboxyphenyl)thio]-ethyl sodium salt). For each biotinylation reaction, a 1ml plastic syringe was stoppered with glass wool and filled with preswelled resin. Each column was placed into a 15ml orange cap tube (Corning Incorporated, Corning, NY) and centrifuged upright in a IEC HN-SII centrifuge. Centrifugations were performed by going up to ¾ speed and down again. The eluate was discarded, the column refilled with preswelled resin and respun until the resin was within 1cm of the top of the syringe. The column was calibrated immediately prior to use by adding a volume of PBS, identical to the reaction volume, to the top of the column and centrifuging at ¾ speed for 2 min. The steps were repeated until the volume of the eluate equaled the volume added. After calibration of the column, labeled

anti-VPg IgGs were added. Again, the centrifugation was performed at ¾ speed for 2 min and the purified, biotinylated protein eluate was transferred to an eppendorf tube.

The biotinylated anti-VPg IgGs (primary antibody preparation) were diluted 1:600 (v/v) in PBS + 0.1% Tween 20, pH 7.5 (PBS-T) containing 2% dried milk and used in western blot analysis as described below (ECL Western blotting detection, Amersham LifeScience).

Gel electrophoresis and western blotting:

In order to test unlabeled anti-VPg Abs, total protein from healthy and infected plants was mixed with Laemmli buffer (Laemmli, 1970) and individual proteins were separated on 15% SDS-PAGE (Sambrook *et al.*, 1989). The protein gel and Immobilon-P membrane (Millipore) were incubated in transfer buffer (a 1:5 dilution of methanol containing 25mM Tris, 192mM glycine, 0.005% SDS) for 5 min prior to assembly of the electroblot sandwich. After completion of the transfer in the cold room at 30V overnight or for 2 hours at 80V but prior to the antibody incubation, the membrane was placed in a plastic tray and washed for 5 min in Tris-buffered Saline + 0.1% Tween 20 (TBS-T; 20mM Tris-HCl pH 7.4, 150mM NaCl). Blots were probed for 45 min with primary Abs diluted TBS-T containing 2% dried milk at a concentration of 1:1000 (v/v) for preimmune serum, 1:5000 (v/v) for anti-PEMV Abs, or 1:1000 (v/v) for anti-VPg Abs. Three washes for 3 min in fresh changes of TBS-T were performed between antibody incubations. Phosphatase-labeled, affinity purified goat anti-rabbit IgGs [diluted

1:2000 (v/v)] (Kirkegaard&Perry Laboratories Inc., Maryland) were used as secondary Abs for another 45 min incubation. Afterwards membranes were washed as previously indicated and proteins were visualized by the use of chromogenic substrates with alkaline phosphatase-coupled antibodies (Sambrook *et al.*, 1989). Bromo chloro indolyl phosphate (BCIP, 40µl of 75mg/ml in 100% dimethyl formamide [DMF]) and nitroblue tetrazolium (NBT, 40µl of 150mg/ml in 70% DMF) were added to 10ml of color buffer, pH 8.5 (0.1M Tris, 0.1M NaCl, 5mM $MgCl_2$). This buffer was used to incubate the blot with agitation until the bands developed to the desired intensity. To remove all reagent, the blot was extensively washed in tap water, once in TBS-T, five times in dH_2O and then placed into 0.02% sodium azide (NaN_3) overnight in the dark before drying. Dry blots were stored in the dark to prevent signal fading.

Western detection with biotinylated anti-VPg IgGs:

Extensively purified viral RNA (55µg), untreated or treated with 10µg RNase A or 125 µg Protease K for 90 min at 37°C, separated electrophoretically and electroblotted as described above, were probed with biotinylated anti-VPg IgGs according to the manufacturer's protocol (ECL-protein biotinylation module, Amersham LifeScience). The membrane was blocked during a one hour incubation in a plastic tray containing 5% blocking reagent after electroblotting. After rinsing the blot twice with PBS-T, it was incubated for another hour, this time with biotinylated anti-VPg IgG, primary antibody solution. Prior to the one hour incubation with streptavidin-horseradish peroxidase (S-HRP; diluted 1:5000

in PBS-T), the membrane was rinsed twice in PBS-T, washed once for 15 min and twice for 5 min in fresh changes of PBS-T. The S-HRP was discarded, the membrane rinsed twice and washed three times for 15 min in PBS-T.

Proteins were detected using ECL reagents and following the manufacturer's protocol (Amersham LifeScience). Equal volumes of ECL detection solution 1 and 2 were mixed (0.125ml/cm^2 membrane) at the end of the last wash. The blot was placed on SaranWrap (DowBrand, Inc.), protein side up, and detection reagents were added for 1 min without agitation. Excess detection reagents were drained and the membrane was wrapped in SaranWrap, smoothing out all air bubbles and placing it , protein side up in a Kodak x-ray exposure holder (20.3x25.4cm, Eastman Kodak Company, Rochester, NY). Autoradiography film (Kodak X-OMAT AR, 20.3x25.4cm) was placed on top and the film was initially exposed for one minute. Film was developed in an automatic developer (Kodak RP X-OMAT Processor, model M7B) and additional pieces of film were exposed for varying length of time based on the strength of the signal.

Sequencing of VPg:

Five hundred μg of extensivly purified viral RNA was incubated with three units of T1 RNase (Sigma) for ten hours (based on the protocol described by Zalloua *et al.*, 1996). The preparation was given to the Biochemistry Macromolecular Structure Facility, MSU for protein sequencing. They applied the sample to a ProSorb membrane (Applied Biosystems) and subjected it to

Edman degradation in a Protein Sequencer, Procise model 494 (Applied Biosystems).

Sequence analysis of VPg:

The obtained protein sequence was compared to other sequences in the database using the BLAST program (Altschul *et al.*, 1990). Protein secondary structure was predicted using the University of Wisconsin (UW) GCG - Peptidestructure program (Devereux *et al.*, 1984) and the PHD (Profile fed neural network systems from Heidelberg) protein mail server at EMBL Heidelberg, Germany (Rost & Sander, 1993; Rost & Sander, 1994a; Rost & Sander, 1994b; Rost, 1996). Additional analysis was performed using the default parameters of individual functions of the UW GCG program (Devereux *et al.*, 1984): Motifs to determine possible protein modification sites, BestFit for protein sequence alignments of PEMV-VPg and other *Luteoviruses* , Isoelectric to calculate the charge of the VPg peptide at neutral pH , and FindPatterns to identify possible cleavage sites in the vicinity of the VPg.

Results

To characterize PEMV-VPg, antibodies (Abs) against the VPg were produced. The viral RNA purified from pea plants infected with the deletion strain of PEMV (AT- Δ) was subjected to extensive phenol-chloroform extractions to minimize the amount of contaminating cp and injected for antibody production.

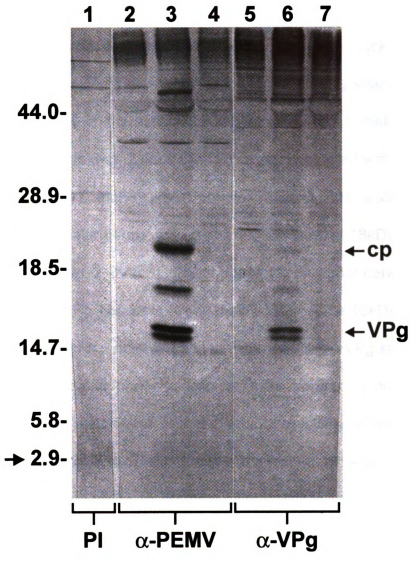

Figure 2: Western blot analysis of total protein extracts isolated from pea leaves. Protein extracts (40μg per lane) were separated by 15% SDS-PAGE (5% stacking gel) and electroblotted onto membranes. These were analyzed using: lane 1, 1:1000 (v/v) rabbit preimmune serum (PI); lanes 2-4, 1:5000 (v/v) rabbit anti-PEMV serum (α-PEMV); lanes 5-7, 1:1000 (v/v) rabbit anti-VPg serum (α-VPg) in a double antibody sandwich with goat anti-rabbit alkaline phosphatase conjugate. Total proteins isolated from leaves of plants inoculated with: RNA 1 and RNA 2 infectious transcripts, lanes 1, 3, 6, respectively; RNA 2 infectious transcripts, lanes 4 and 7; lanes 2 and 5, mock-inoculated plants. The positions of M_r standards are indicated on the left as well as the position of the gel front (arrow). The positions of PEMV coat protein (cp) and VPg are indicated on the right.

Comparisons between total protein extractions from mock inoculated and infected plants probed with anti-VPg Abs indicated a lightly staining band at approximately 18kDa and a heavily staining doublet around 15kDa in RNA 1 + RNA 2 infected plants (Fig. 2, lane 6) that were not detected with preimmune serum (Fig. 2, lane 1). In the same protein extractions, Abs raised against virion PEMV detect four major bands: the cp at 21kDa, a band at 18kDa and the doublet at 15kDa (Fig. 2, lane 3). The reduction in cp signal between the two Abs suggests anti-VPg Abs were raised mainly against the 15kDa doublet. A reduction in cp signal but no change in the strength of the VPg signal was observed with sequential phenol-chloroform extractions (data not shown). We concluded from these experiments that the minor reactivity of the anti-VPg Abs against the cp (Fig. 2, lane 6) was due to a slight contamination of the injected sample with cp. None of the above mentioned bands were observed when total protein extractions from RNA 2 infected plants were probed with anti-PEMV Abs or anti-VPg Abs (Fig. 2, lanes 4 and 7, respectively) or in protein extractions of mock-inoculated plants probed with either antibody (Fig. 2, lanes 2 and 5, respectively). Since the immunoreactive 18kDa band (Fig. 2, lane 3 and 6) appeared erratically in different experiments we concluded that the 15kDa protein doublet most likely represents the VPg.

Biotinylated VPg-IgGs were used to detect the VPg isolated from viral RNA. No protein was detected in the untreated, viral RNA lane (Fig. 3, lane 1). Both, the 18kDa and 15kDa proteins, were detected after RNase A-treated viral

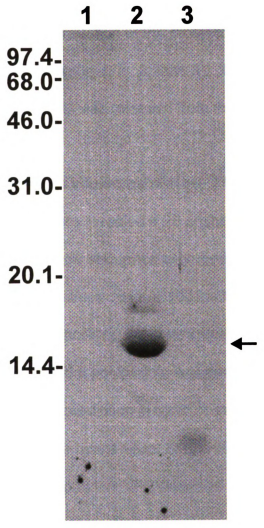

Figure 3: Differentially treated VPg attached to PEMV RNA (55μg per lane) probed with biotin labeled rabbit anti-VPg IgG. Treatments were: lane 1, untreated RNA; lane 2, RNA treated with RNase A; lane 3, RNA treated with Proteinase K. The positions of M_r standards are to the left of this figure while the position of the PEMV-VPg is indicated by an arrow.

RNA was probed with biotinylated anti-VPg IgGs (Fig. 3, lane 2). Protease K treatment resulted in the degradation of the proteins as evident from the increased electrophoretic mobility (Fig. 3, lane 3). This indicates that anti-VPg IgGs recognized a protein that was released from the viral RNA by RNase A treatment.

Purified viral RNA was treated with RNase T1 to release the VPg. Microsequencing of the protein revealed a 28 amino acid (aa) sequence with a M_r of 3157 Da. Using computer sequence alignments, the sequence encoding this protein mapped to nucleotides 1811 to 1894 within ORF 1 of RNA 1, downstream of nucleotides encoding the proteinase (Fig. 4). The small size and location of the VPg suggested a polyprotein expression strategy of ORF1 of RNA1. Secondary structure prediction programs were used to determine a possible folding pattern of the hypothetical 84K protein. In addition, the protease accessibility of the putative VPg cleavage sites was analyzed in order to identify possible cleavage sites. Both cleavage sites, Glu-Ala at the N-terminus and Gln-Ser at the C-terminus of VPg, were solvent accessible and loops were found at or near these sites. In addition, these dipeptide sequences are cleavage sites known to be utilized by other viral 3C-like proteases (Wellink & van Kammen, 1988). These results are consistent with a total length of 28 aa for the VPg.

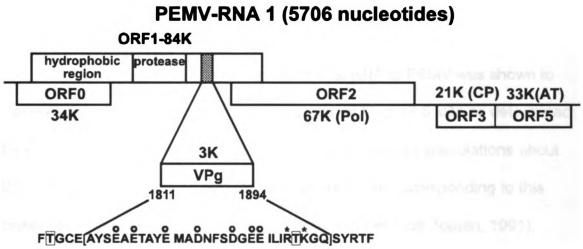

Figure 4: PEMV-VPg protein sequence and its location within the ORF 1 of RNA1. Arrangements of the ORFs of RNA1, the protease, and the N-terminal region containing hydrophobic patches, are based on Demler & de Zoeten (1991). The shaded box indicates the position of the coding region of VPg. The enlarged area shows the nucleotide coordinates and protein sequence of VPg and putative cleavage sites ([ ]). The boxed amino acids indicate the P4 position respective to the putative cleavage sites. Abbreviations: coat protein (CP), aphid transmission factor (AT), RNA polymerase (Pol), basic amino acid (*), acidic amino acid (o), K = kDa

Chapter 3

Discussion

In a previous study, the 5' end of the viral RNA of PEMV was shown to possess a covalently linked VPg with an apparent M_r of 17.5 kDa as determined by electrophoresis (Reisman & de Zoeten, 1982). Despite speculations about the coding region of this protein, no independent ORF corresponding to this protein size was detected in the viral genome (Demler & de Zoeten, 1991). Therefore, the goals of this study were a) to produce polyclonal Abs that could verify the presence of the VPg and b) to obtain and analyze the aa sequence of the VPg, in order to deduce its placement on the viral genome.

The use of extensively purified viral RNA as immunogen was a simple and effective way to obtain polyclonal Abs against the PEMV-VPg. These Abs, either unlabeled or biotinylated, were used to characterize the VPg by western blot analysis. Several lines of evidence support our conclusion that the 15kDa immunoreactive bands correspond to the genome-linked VPg. First, western blot experiments that utilized anti-PEMV Abs and anti-VPg Abs to analyze infected plant protein extracts (Fig. 2, lanes 3 and 6) showed that the VPg has different antigenic sites from those of the cp. Thus, anti-VPg Abs were raised against a protein different from the 21kDa cp. Second, experiments analyzing purified viral RNA with biotinylated anti-VPg IgGs (Fig. 3) and infected plant

protein extracts with anti-VPg Abs (Fig. 2) both labeled a 15kDa doublet as the predominant signal, suggesting that the same protein was recognized when isolated by two different sample extraction procedures. Third, the protein identified in this study as the VPg bound to the RNA. Noncovalently bound proteins were drastically reduced by the extensive viral RNA purification procedure as reflected by the increasingly faint cp-derived signal (Fig. 2, compare lanes 3 and 6). In addition, when viral RNA was treated with RNase A, the 15kDa protein was preserved (Fig. 3, lane 2) but when the same viral RNA was treated with Protease K (Fig. 3, lane 3) the protein was degraded. Thus, the proteinaceous nature of the signal was confirmed. Vartepatian & Bogdanov (1987) suggested that combined nuclease and protease treatment are the best way to identify proteins covalently linked to RNA. However, we were unable to detect a signal when untreated viral RNA was analyzed (Fig. 3, lane 1). This was probably due to the inability of the large RNA-protein complex to enter the 15% polyacrylamide gel.

The detection of a double band in protein extracts (Fig. 2), as opposed to the expected single band, may be due to several reason: 1) Protein degradation could have generated the 15kDa protein doublet. Although a proteinase inhibitor (PMSF) was added during the extraction of total protein, it is possible that not all proteinase activity was inhibited and such activity could have generated proteins containing VPg antigenic sites. 2) The anti-VPg Abs might have recognized a VPg precursor protein in addition to the VPg. One protein of the 15kDa doublet could be a VPg precursor. Poliovirus (Baron & Baltimore,

1982) and cowpea mosaic virus (CPMV) (Zabel *et al.*, 1982) have been shown to posses precursor proteins to the VPg that are recognized by Abs raised against the VPg. 3) A differing number of RNase resistant bases might be left attached to the protein. One residual RNA nucleotide linked to some VPg molecules would be sufficient to change the M_r and charge of the protein.

Protein modifications are also theoretically possible. The UW GCG - Motifs program (Devereux *et al.*, 1984) identified a possible N-glycosylation site at aa 14 (Asn) and two possible casein kinase II phosphorylation sites at aa 7 and 16 (Thr and Ser). However, protein sequencing did not indicate any modifications as aa signal peaks would have been altered.

Interestingly, a similar strength of the VPg and cp signal was found in protein extracts from infected plants probed with anti-PEMV Abs (Fig. 2, lane 3). The reason for this is currently not understood. The amount of each protein needed during the viral life cycle differs widely. Only one molecule of VPg is thought to be attached to each RNA molecule while 180 or 150 subunits of cp are required for formation of the RNA1 and RNA2 capsid, respectively (for review see Demler *et al.*, 1996a). Therefore, it is important to note that western blotting techniques are not quantitative. The relative band intensities of VPg and cp do not necessarily reflect that these proteins are expressed in equal amounts. Hence, these proteins may differ in their antigenicity.

Protein sequencing data and subsequent computer analysis showed that the PEMV-VPg gene was located downstream of the proteinase motif within the 84K ORF of RNA 1 (Fig. 4). This result differs from an earlier proposal by

Demler *et al.* (1996a) that predicted the coding region upstream of the proteinase motif based on similarities of motif arrangements in *Picorna*-like viruses and PEMV (Demler & de Zoeten, 1991). Proteolytic cleavage is a strategy used by the *Picornaviruses* (Dougherty & Semler, 1993). Demler & de Zoeten (1991) identified a potential proteinase sequence motif in the 84K ORF, suggesting the processing of expressed viral proteins in PEMV infected tissues. Size and location of the VPg protein would argue for an involvement of proteolytic activity in the release of the VPg from its precursor protein. This is the first direct indication for polyprotein processing in the expression strategy of *Enamo-* and *Luteoviruses*.

The deletion strain of PEMV (AT-Δ) was cloned from a naturally occurring isolate of PEMV characterized by the lack of 727 nucleotides from the 33K ORF 5 of RNA 1 (Demler, *et al.*, 1996a). This isolate consistently produced three to five times the yield of the aphid-transmissible [AT^+] strains. This high yield made it well suited for the isolation of large quantities of RNA needed for this study (Demler *et al.*, 1996a). Protein sequencing determined that the first aa of the PEMV-AT-Δ-VPg is alanine. The corresponding aa in the published sequence of RNA 1 derived from the WSG strain is threonine (Demler & de Zoeten, 1991). The single base substitution from ACC to GCC causes this change from alanine in the protein sequence of PEMV-AT-Δ -VPg to threonine in the AT-WSG-VPg. Unpublished results (Skaf *et al.*) showed that the same ACC to GCC substitution was found in the aphid transmissible PEMV-AT^+. Hence, the discrepancy is likely due to strain differences.

PEMV VPg consists of 28 aa. This is supported by secondary structure predictions of the hypothetical protein encoded by ORF 1 and by the proteinase accessibility of its putative cleavage. Accessible cleavage sites are most likely exposed as surface loops (e. g. Dougherty & Semler, 1993; Pallai *et al.*, 1989; Arnold *et al.*, 1987). In an analysis with different computer programs (data not shown), the solvent accessibility at the cleavage sites surrounding the VPg was shown to be favorable and loops were found to be at or near these sites.

Additional support comes from the likelihood that the putative VPg cleavage sites are those used preferentially by 3C-like proteases. The majority of proteolytic cleavages of the picornaviral polyprotein are carried out by 3C-like proteases (Wellink & van Kammen, 1988). Most cleavages occur between Gln or Glu and Gly or Ser, less frequently Thr/Ala/Met/Val, amino acid pairs (Wellink & van Kammen, 1988). Named after the 3C protease (3C ^{pro}) of *Picornaviruses*, the family of 3C-like proteases possess a conserved catalytic triad and belong in the chymotrypsin -related protease superfamily (Gorbalenya & Koonin, 1993). In addition to *Picornaviruses*, PEMV (Demler & de Zoeten, 1991) and *Como-*, *Poty-*, and *Nepoviruses* (Dougherty & Semler, 1993) possess 3C-like protease motifs. Hence, the region of the protein surrounding the PEMV-VPg was searched for the presence of cleavage sites similar to those utilized by known 3C-like proteinases (cleavage site sequences were taken from Dougherty & Semler, 1993; Wellink & van Kammen, 1988). This search indicated that the N-terminus of PEMV-AT-Δ-VPg could be generated by a cleavage between glutamic acid

and alanine (Fig. 4), as has been previously observed for *Picorna-* and *Nepoviruses* (for review see Wellink & van Kammen, 1988). In the published sequence of the aphid-nontransmissible AT-WSG strain of PEMV, threonine is encoded instead of alanine at the N-terminus (Demler & de Zoeten, 1991). This Glu-Thr cleavage is also observed in *Picornaviruses* (for review see Wellink & van Kammen, 1988). Therefore, it appears likely that Ala, the first amino acid obtained by protein sequencing, was actually the N-terminus of the VPg. The C-terminus of the PEMV-VPg is believed to result from a Gln-Ser cleavage (Fig. 4). Such a cleavage site is common for *Picornaviruses* as well as *Como-* and *Potyviruses* (for review see Wellink & van Kammen, 1988). In addition, comparative analysis of sequences at tentative protein domain boundaries in polyproteins of *Luteo-* and *Sobemoviruses* previously predicted cleavage sites at Gln or Glu and Gly or Ser or Ala (Gorbalenya & Koonin, 1993). We, therefore, reason that the VPg protein of PEMV is 28 aa long, with a calculated molecular mass of 3157 Da.

The discrepancy between the calculated (3157 Da) and observed (15 kDa or 17.5kDa) M_r of PEMV-VPg may be due to residual RNA nucleotides that remained attached to the VPg thus contributing to the overestimated M_r of the protein when run on SDS-PAGE. Other VPgs run also anomalously in SDS-PAGE due to their cationic character (Daubert & Bruening, 1984). Interestingly, the net charge of the PEMV-VPg is negative (-5.02) at neutral pH, suggesting an anionic character, with one fourth of the aa being acidic (7aa) and only 7% being

basic (2aa) (Fig. 4). At present, it is not clear what factors contribute to the aberrant migration of PEMV-VPg in SDS-PAGE.

The basic amino acids found in the C-terminal half of the poliovirus VPg are thought to be involved in electrostatic interactions of VPg and RNA while the N-terminal portion of the protein is involved in the protein-RNA linkage (for review see Vartepatian & Bogdanov, 1987). Interestingly, the only two basic aa found in PEMV-VPg are close to the C-terminus (see Figure 4). Two residues of tyrosine and serine, amino acids shown previously to be involved in the RNA-protein linkage of other viruses (see Chapter 1 for references), can be found within the PEMV-VPg sequence. The tyrosine and the serine residues are each found towards the N-terminus and in the center of the VPg. Based on the position of the linking aa in the N-terminal half of the VPg in several viruses (Vartepatian & Bogdanov, 1987; Murphy *et al.*, 1991; Jaegle *et al.*, 1987; Pinck *et al.*, 1991; Zalluoa *et al.*, 1996), we predict that the tyrosine at aa 2 or the serine at aa 3 in PEMV-VPg are involved in the covalent linkage of the protein and the RNA. However, at the moment experimental evidence for this hypothesis is lacking.

No significant sequence similarity was found between the protein sequences of PEMV-VPg and *Luteoviruses* or any other virus. This is not surprising, since to date no common VPg sequence elements have been identified and attempts to predict VPg domains on *Luteo-*, *Sobemo-* and *Calicivirus* genomes through comparative sequence analysis have been inconclusive (Gorbalenya & Koonin, 1993). Based on other similarities between

PEMV and *Luteoviruses*, information concerning the PEMV-VPg might prove useful in evaluating the VPgs of *Luteoviruses*.

The cleavage site sequences Glu-Ala (or Glu-Thr) and Gln-Ser utilized by the viral proteinase to release the VPg were found at other sites within the 84K polyprotein. However, not all possible dipeptide sequences within a respective genome are cleaved. For example, only 9 out of 13 Gln-Gly pairs in the poliovirus polyprotein are cleaved (Dewalt & Semler, 1989; Ypma-Wong *et al.*, 1988). It was proposed that in poliovirus $3C^{pro}$ specifically recognizes Gln-Gly pairs while residues flanking this dipeptide together with structural determinants mediate cleavage efficiency (Mirzayan & Wimmer, 1994).

Threonine was found at P4 in both PEMV-VPg cleavage sequences. The cleavage site nomenclature is based on Berger & Schlechter (1970) where P1 corresponds to the newly generated C-terminal residue and P1' to the new N-terminal aa. Studies from *Picorna*-, *Como*- and *Potyviruses* determined aa requirements in the vicinity of cleavage sites and their effect on processing dynamics (for review see Wellink & van Kammen, 1988). The importance of the P4 aa in addition to the dipeptide cleavage sequence is well documented (e. g. Blair & Semler, 1991; Cordingley *et al.*, 1990; Pallai *et al.*, 1989; Harris *et al.*, 1992). In the cases studied, the P4 position is preferentially occupied by a small subset of amino acids with hydrophobic groups, preferentially Ala and less often Val, Leu, Ile, Pro, Phe (for review see Wellink & van Kammen, 1988). In poliovirus the P4 position regulates the cleavage efficiency (Mirzayan & Wimmer, 1994). The P5, P3 and P2 aa appear less critical but are still part of a

minimum length peptide sequence required for cleavage (Long et al., 1989; Cordingley *et al.*, 1990). This data is supported by X-ray crystallography of another *Picornavirus* protease, the human rhinovirus 3C protease (Matthews *et al.*, 1994). A variety of amino acids are tolerated in the P5 and P3 position as aa sidechains point into the solution while a spacious cavity in $3C^{pro}$ is found opposite of P2. However, only small hydrophobic sidechains can be tolerated in the P4 position due to steric constrains. In addition to the favored aa residues, threonine can also be found at P4. Like PEMV-VPg, the poliovirus 3CD protein carries an unfavored threonine in the P4 position of its 3C*D cleavage site (Harris *et al.*, 1992; Pallai *et al.*, 1989). The polyprotein 3CD processing intermediate is the proteolytically active precursor of the 3C protease and the viral polymerase 3D (Mirzayan & Wimmer, 1994). Interestingly, this unfavorable cleavage site still allows for sufficient processing to yield the necessary amounts of cleavage products (Harris *et al.*, 1992). Pallai *et al.* (1989) proposed that the slow processing and the long 3CD half-life were a regulatory mechanism by the virus to govern accessibility and availability of proteins during its life cycle. Also, Semler *et al.* (1988) speculated that the initiation of RNA synthesis *in vivo* may be controlled by proteolytic processing of VPg precursor polypeptides. In PEMV, the viral polymerase of RNA 1 is expressed as a fusion protein with the 84K polyprotein containing the protease and the VPg. It is, therefore, possible that like in poliovirus, different cleavage activities determine the availability of proteins and that PEMV- VPg, whose release from the polyprotein is delayed by

suboptimal cleavage sites, may be activated by proteolysis at a specific point during the viral life cycle, e.g. RNA synthesis.

A similar concept, that proteolysis might temporarily and functionally be connected with the RNA replication process, was proposed by Riechman *et al.* (1993) for *Potyviruses.* A conserved heptapeptide recognition sequence on substrate proteins is absolutely required for proteolysis by potyviral proteases (Dougherty & Parks, 1989). The P6, P3 and P1 aa are strictly conserved in all cleavage sites of a particular potyvirus. They define an active cleavage site while the amino acids at P5, P4, and P2 vary and determine the rate and extent of processing (Dougherty & Parks, 1989; Dougherty *et al.*, 1989). Like other viruses of the *Picornavirata*, the VPg in *Potyviruses* lies upstream of the 3C protease (Riechman *et al.*, 1993). The nuclear inclusion protein a (NIa) of *Potyviruses* consists of a N-terminal VPg domain and a C-terminal protease domain (Riechman *et al.*, 1993). In tobacco etch virus (TEV), the NIa protease catalyzes a suboptimal internal cleavage site, cleaving some but not all NIa molecules at a slow rate. PEMV-VPg is adjacent to the protease in the PEMV genome, but in this case is located downstream rather than upstream of the putative protease motif (see Figure 4). However, based on the appearance of a suboptimal aa at P4 and the adjacent arrangement of VPg and protease, the possibility exists that suboptimal cleavage sites are responsible for the release of PEMV-VPg.

Based on the observations presented above, we suggest that the 84K ORF of PEMV encodes a polyprotein. Therefore, the aa sequence of the 84 kDa

protein was searched for other possible cleavage sites of the utilized dipeptide sequences, Glu-Ala (or Glu-Thr) and Gln-Ser. In addition to the previously identified cleavage sequences, one Glu-Ala, four Glu-Thr and six Gln-Ser dipeptide sequences were found. From those, only two Gln-Ser sites (at aa 600 and 720) downstream of the VPg in the C-terminus of the 84K protein, seem likely to be utilized by the viral protease. These sites have a favorable secondary structure of solvent accessible loops and the appearance of a small hydrophobic aa at P4. Both of these putative cleavage sites are located within the region identified by Demler & de Zoeten (1991) for use as frameshifting site in the expression of the RNA 1 polymerase. Maybe, one of these sites is used to release the polymerase from the 84K-67K fusion protein (see Figure 1). No consensus sequence was found comparing the minimum required peptide cleavage sequences (P6 to P1') of these putative sites with the VPg cleavage sites, however, no consensus flanking sequence could be identified in *Picorna-* and *Comoviruses* either (Wellink & van Kammen, 1988).

The different evolutionary origins of PEMV RNA 1 and RNA 2 were discussed previously (see Chapter 1). The polymerase of RNA 1 is closely related to members of the *Luteovirus* subgroup II (Demler & de Zoeten, 1991) and belongs to the picornavirus-like supergroup (Koonin & Dolja, 1993). Members of this supergroup possess a VPg and are thought to have evolved from a common, VPg containing ancestor (Koonin & Dolja, 1993). Therefore, the polymerase of the ancestor virus and polymerases derived from it, including PEMV RNA 1 polymerase, may have become functionally dependent on a VPg.

On the other hand, PEMV RNA 2 polymerase is related to the carmovirus-related plant viruses (Demler *et al.*, 1994a) that evolved from a different ancestor which did not possess a VPg nor did any of the virus groups evolutionary derived from this ancestor ever acquire a VPg (Koonin & Dolja, 1993). Subsequently, these polymerases may have evolved RNA replication mechanisms independent of a VPg. Based on these considerations, we believe it unlikely that the RNA 2 polymerase depends on a VPg for replication of the RNA 2 genome. No experimental evidence is available but based on these evolutionary considerations, we pose that PEMV-VPg is only attached to RNA 1.

Future research on PEMV-VPg will answer unresolved questions such as: Which aa is involved in the protein-RNA linkage? Is PEMV-VPg covalently linked only to the RNA 1 molecule or does it also attach to RNA 2? What is the function of the VPg? Other topics related to the PEMV-VPg also remain unresolved, such as the unknown cleavage requirements of the viral protease or the relationship between the protease and VPg in the viral life cycle. Mutagenesis experiments could be used to study the function of the VPg and determine the aa involved in the protein-RNA linkage.

In conclusion, we have provided evidence that the VPg of PEMV is encoded in the ORF1 of PEMV RNA1 in a polyprotein arrangement of membrane anchor-protease - VPg - polymerase. Post-translational proteolysis is the most logical explanation for the expression of mature VPg.

Appendix

Appendix A

APPENDIX A

mutagenesis of the 33K ORF of RNA 2

Introduction

Pea Enation Mosaic Virus (PEMV), the sole member in the Enamovirus genus (for review see Demler *et al.*, 1996a), is aphid-transmissible and causes significant losses in the pea-growing regions of the United States and in Europe (for review see Hagedorn, 1996). Two, isometric particles, each encoding a different RNA, make up the infectious virus. RNA 1 is responsible for encapsidation and aphid transmission, while RNA 2 determines mechanical transmissibility and systemic movement (for review see Demler *et al.*, 1996a). Genomic sequencing of RNA 1 and RNA 2 revealed five and four open reading frames (ORFs), respectively (Demler & de Zoeten, 1991; Demler *et al.*, 1994a). The functions of the gene products of individual ORFs was elucidated by comparative sequence and/or experimental analysis. However, the functions of the 34K and 84K ORFs of RNA 1 and of the 33K ORF of RNA 2 (see Figure 1) remain obscure.

In order to learn molecular cloning techniques, I decided to introduce, in a side-project, a nonsense mutation into the 33K ORF of RNA 2. We hoped that this would possibly provide preliminary information pointing to a function of the 33K protein.

The strategy involved selecting a restriction site near the beginning of the 33K ORF and modifying this site for early translation termination. These modified complete cDNA clones were used for production of infectious transcripts and plant inoculation. Infected plants were tested for viral RNA and proteins and symptoms were observed. However, preliminary data are recorded here and additional experiments are suggested to determine a functional role for the RNA2 ORF1 product.

Materials and Methods

Creation of a frameshift mutation in the 33K ORF:

Five μg of pPER2d cDNA (Demler, *et al.*, 1993) were linearized with 4 units (U) of BseR1 (New England Biolabs Inc., Beverly, MA) in 1x NEB 2 buffer (New England Biolabs) during a 60 min incubation at 37°C. One U of Klenow enzyme (Boehringer Mannheim, Indianapolis, IN) and 1μl of 0.5mM of each dNTP (dATP, dCTP, dGTP, dTTP) were added to create blunt ends and the incubation continued for another 15 min. The reaction was stopped with 1μl of 0.5M EDTA and the extent of DNA linearization evaluated on 1% non-denaturing agarose gel (see below).

The linearized DNA was isolated using the GENECLEAN II kit from BIO 101(Vista, CA), following the manufacturer's recommendations. Initially, three volumes of sodium iodine (NaI) stock solution were added to the reaction and

mixed before 5µl of "Glassmilk", a silica matrix that binds DNA, were added. The tube was vortexed horizontally for approximately 1 minute and placed for 5 min at 55°C while mixing every one to two minutes. DNA bound to Glassmilk was pelleted by centrifugation for 5 seconds in an Eppendorf centrifuge at high speed of 14000 (14K) rpm (Brinkmann Instruments, Inc., Westbury, NY), the supernatant was removed and the pellet was resuspended and washed three times with "New Wash". Each wash involved resuspending the pellet in 10 to 50 volumes of "New Wash" and centrifuging for five seconds at high speed with disposal of the supernatant. After the last wash, the "Glassmilk" was removed from the DNA by resuspending the pellet in a volume of dH_2O at least equal to the volume of "Glassmilk" (5µl) previously added, and incubating at 50°C for three minutes. The mixture was then spun for 30 sec. at 14K rpm and the supernatant, containing the DNA, removed. The resuspension and incubation of the "Glassmilk" pellet was repeated once to increase the efficiency of DNA recovery. The DNA was ready for ligation and transformation reactions.

The ligation was performed by adding 5x ligation buffer (Boehringer Mannheim) and 1 U of T4 DNA Ligase (Boehringer Mannheim) to the DNA followed by incubation of the mixture at room temperature (RT) overnight. The ligated DNA was transformed into competent JM101 *Escherichia coli* cells (Sambrook *et al.*, 1989). Competent cells were thawed in ice-water before 100µl of cells were mixed gently with 50 - 200ng of ligated plasmid. The mixture was placed on ice for 30 min, heated for 1 min at 42°C, and moved back on ice for another 2 min before plating out the cells. Pre-poured LB plates (1% w/v

tryptone, 0.5% w/v yeast extract, 0.55% w/v NaCl, 0.12% v/v 4N NaOH and 1.5% w/v Bacto Agar) containing 40μg/ml ampicillin were warmed to 37°C prior to plating the transformation reactions. One microliter of each transformation reaction was mixed in 50μl of dH_2O and plated out. The remainder of each transformation reaction (roughly 99μl) was plated onto a separate plate. This ensured varying concentrations of colonies growing on the different plates. All plates were incubated overnight at 37°C in a incubator (Precision Gravity Convection Incubator) and individual colonies were picked using sterile toothpicks the next day.

Each selected bacterial colony was grown in 2ml of liquid 2xYT media (1.6% w/v tryptone, 1% w/v yeast extract, 0.5% w/v NaCl) containing 40μg/ml ampicillin overnight at 37°C in an Orbit Environ-Shaker in preparation for plasmid DNA isolation (Lee & Rasheed, 1990). Cultures were centrifuged in 1.5ml microcentrifuge tubes at 14K rpm for 1 min to pellet the cells and the supernatant was discarded. Pelleted cells were resuspended in 100μl lysis buffer (50mM glucose, 25mM Tris-HCl pH 8.0, 10mM EDTA) vortexed and incubated for 5 min at RT. Two hundred μl of freshly prepared alkaline solution (0.2N NaOH pH 5.5, 1% v/v SDS) and later 150μl ice-cold 7.5M ammonium acetate, pH 5.5 were added. Each time the tubes were gently mixed by inversion, followed by a five minute incubation in an ice-water bath. The mixture was centrifuged at 14K rpm for 5 min. The clear supernatant was transferred to a new microcentrifuge tube, containing 0.6 volumes of isopropanol, and the reaction was incubated at RT for a minimum of 10min. A centrifugation at 14K

rpm for 10min followed and 100 μl of 2M ammonium acetate (pH 7.4) were used to resuspend the DNA pellet. After a 5 min incubation in an ice-water bath, the mixture was centrifuged for 5 min at 14K rpm, the supernatant transferred to a new microcentrifuge tube containing 100μl of isopropanol and incubated and centrifuged as described for the previous isopropanol step. The resulting pellet was washed with 70% ethanol, centrifuged, vacuum-dried and resuspended in 50μl of water.

The success of plasmid isolations was evaluated on 1% non-denaturing agarose gels containing 0.1μg/ml of ethidium bromide (EtBr). A 5μl aliquot of plasmid DNA was mixed with 1μl of 6x loading dye (0.25% w/v bromophenol blue, 0.25% w/v xylene cyanol FF, 40% w/v sucrose in water), electrophoresed in 1xTBE (1.08% w/v Trizma base, 0.55% w/v boric acid, 0.037% w/v EDTA) running buffer at 100V (BioRad minisub DNA cell gel system, BioRad power supply model 500/200) and the gel was photographed under UV light (312nm Transilluminator FBTI 614, Fisher Scientific; Polaroid MP-4 Land Camera).

DNA sequencing and sequence analysis:

To confirm the frameshift mutation, the manipulated region was sequenced in both directions using the dideoxynucleotide chain-termination method of Sanger *et al.* (1977). DNA sequencing (Sequenase Version 2.0, United States Biochemical, Cleveland, OH) was based on the manufacturer's protocol. The DNA was denatured by adding 0.2μl of 10N NaOH for each 10μl of DNA (1-2μg). After a 5 min incubation at RT, 4μl of 5N NH_4OAC and three

volumes of ethanol were added. The DNA was mixed well and placed at -80°C for at least 15 min. Precipitated DNA was centrifuged at 14K rpm for 10 min in an Eppendorf centrifuge and the resulting pellet was resuspended in 7µl of sterile dH_2O for each 10µl of starting DNA.

A short region of RNA 2, 50-100 bp upstream of the manipulated region in either direction, was chosen as sites for primer annealing. Each primer was complementary to the respective part of RNA 2. The17 nucleotide (nt) forward primer, dz63, spans nts 269 - 285 of RNA 2 ($^{5'}$ TGT GGC TCT AGT GGA GG $^{3'}$). The 23nt reverse primer, dz64, hybridizes to nts 363 - 385 ($^{5'}$ TTC AGC CAA CAG GCG ACG ATG GA $^{3'}$). Primers were synthesized by the Biochemistry Macromolecular Structure Facility, Michigan State University, East Lansing, MI.

One µl of either primer (25ng/µl) and 2µl of 5X Sequencing buffer (0.2M Tris-HCl pH 7.5, 0.1M $MgCl_2$, 0.25M NaCl) were combined with the denatured DNA, incubated for 2 min at 65°C and cooled to RT by placing it on the bench for at least 15 min. DNA with the annealed primer was added to the labeling reaction containing 1µl of 0.1M DTT, 2µl of a dGTP,dTTP, dCTP mix (diluted 5-fold in dH_2O), 1µl of ^{35}S ATP (diluted 1:1 in dH_2O) and 2µl of Sequenase (diluted 1:8 in ice cold TE buffer [1% v/v 1M Tris-HCl pH 8.0, 0.2% v/v 0.5M EDTA pH 8.0]). The reaction was mixed and placed at RT for 10min.

In preparation of the termination reaction, V-bottom microtiter plates (96 well, 86x128mm, Fisher Scientific, Pittsburgh, PA) containing 2.5µl of the respective dideoxynucleotide triphosphates(ddNTP's: ddCTP, ddGTP, ddTTP, ddATP) were pre-warmed at 42°C for at least 1 min. After completion of the

labeling reaction, 3.5µl of each reaction were added to each ddNTP and incubated at 42°C for 5 min. Reactions were terminated by addition of 4µl stop solution (95% formamide, 20mM EDTA, 0.05% bromophenol blue, 0.05% xylene cyanol) to each well, and samples were heated in boiling water for 2 min prior to loading of the electrophoresis gel.

Electrophoresis was performed on vertical 8% acrylamide gels. Gels were poured at least 45 min prior to electrophoresis to ensure adequate polymerization. The 8% acrylamide gel mixture consisted of 60ml of 20% acrylamide stock solution (96.5g acrylamide, 3.35g N,N-methylene-bis-acrylamide, 233.5g ultra-pure urea in 500ml 1xTBE), 90ml urea mix (233.5g urea in 500ml 1xTBE), 150µl 25% ammonium persulfate and 150µl N,N,N',N'-tetramethyl-ethylene-diamine (TEMED). Half the volume of each sequencing reaction was loaded between teeth of a shark tooth comb and electrophoresed at a constant gel temperature of 50°C in 1xTBE until xylene cyanol reached the bottom of the gel (Sequi Gen sequencing cell, BioRad; electrophoresis constant power supply ECPS 3000/150, Pharmacia, Piscataway, NJ).

The gel was fixed in a solution of 12% methanol and 10% glacial acetic acid for 20min and transferred to 3mm thick Whatman chromatography paper (Whatman International Ltd., Maidstone, England). The fixed gel was placed on a gel dryer (Gel Dryer model 583, BioRad), covered with Saran wrap and dried under vacuum for 1 hour at 80°C (Speed Vac Water Jet SW J120, Savant Instrument Inc., Holbrook, NY). The dried sequencing gel was exposed overnight on autoradiography film (Kodak Scientific Imaging Film X-OMAT AR

35cm x 43cm, Eastman Kodak Company, Rochester, NY) in a Kodak X-ray exposure holder (35.6cm x 43.2 cm) and the film was developed in an automatic film developer (Kodak RP X-OMAT Processor, model M7B) and analyzed.

The obtained sequences were compared to the original PEMV RNA 2 sequence using the default parameters of the University of Wisconsin (UW) GCG - GAP program (Devereux *et al.*, 1984).

Large scale $CsCl_2$ plasmid isolations:

After confirmation of the mutations, large quantities of selected plasmids were purified on cesium chloride ($CsCl_2$) gradients for RNA transcription reactions (Sambrook *et al.*, 1989).

A 1l flask containing 250ml liquid LB with 40μg/ml ampicillin was inoculated with bacterial cultures harboring the mutation and grown overnight in a shaker incubator at 37°C (Orbit Environ-shaker). Each culture was divided in half and poured into 250ml centrifuge tubes. The bacterial cells were pelleted in a JA-14 rotor (Beckman) at 6000 rpm (5520 x g) for 15 min in a Beckman J2-21 centrifuge. The supernatant was aspirated and pellets were transferred to 40ml centrifuge tubes (Beckman) using a spatula. Four ml of ice-cold solution I (2.5% v/v 1M Tris-HCl pH 8.0, 10% v/v 100mM EDTA, 5% v/v1M sterile sucrose) were added to each 250ml centrifuge tube to resuspend the remaining cells, which were subsequently transferred to 40ml centrifuge tubes and vortexed until the pellets were completely resuspended. Eight ml of freshly made solution II (2% v/v 10N NaOH, 10% v/v 10% SDS) were added, mixed by inverting the

tubes and then incubated on ice for 10 min. Another 10 min incubation on ice with occasional shaking followed after addition of six ml of solution III. This solution was made by dissolving 29.4% (w/v) KOAc in a small volume of H_2O and adjusting the pH with glacial acetic acid to 4.8 before bringing the volume to 500ml. At this point, the bacterial mixture was clear but with a white coagulant. After centrifugation in a JA-20 rotor (Beckman) at 15K rpm (27,200 x g) for 20 min, the clear supernatant, was transferred (avoiding debris) to new tubes containing 10ml isopropanol. The now cloudy solution was mixed well, incubated for 5 min at RT and centrifuged at 10K rpm (12,100 x g) for 10 min. The supernatant was discarded and pellets were washed with 70% ethanol, vortexed briefly and spun at 10K rpm for 5 min. The resulting pellets were semi-dried under vacuum for 5 min and completely dissolved in 4ml of 25mM Tris-HCl pH 8.0 before being transferred to 14ml Falcon tubes (Becton Dickinson and Company, Lincoln Park, NJ) containing 4.55g of $CsCl_2$. After the $CsCl_2$ was dissolved, 667µl 10mg/ml ethidium bromide (EtBr) were added; samples were vortexed and incubated in the dark for 20 min. Solutions were spun in a JA-20 rotor (Beckman) at 10K rpm for 10 min and the clear supernatant was transferred to Quick-Seal tubes (Beckman) for the VTi 65.2 rotors (Beckman). The volume usually filled the tubes but 1mg/ml $CsCl_2$ in TE was sometimes used to top off the tubes. All tubes were sealed with a Beckman tube sealer and then centrifuged at 45K rpm for 16 hours at 20°C.

The resulting gradients were handled with great care and appropriate safety equipment was worn for the remainder of the isolation procedure. To

extract the DNA band from each gradient, a tube was securely clamped in front of a hand-held long-wave UV lamp (Mineralight lamp model UVGL-25, UVP, Inc., San Gabrial, CA), above a, EtBr waste beaker. The top of each tube was pierced with a 25G needle to allow air to enter as the sample was withdrawn. A fluorescing horizontal band in the middle of the tube represented plasmid DNA and a vertical dark band the RNA pellet. A 3cc syringe with an 18G needle was carefully inserted, beveled side up and avoiding the pellet, just under the plasmid band which was then slowly drawn into the syringe by sweeping back and forth while filling the syringe. In case the needle became plugged, another needle was inserted. Once the plasmid was collected, the needle was removed and the sample dispensed into 14ml Falcon tubes. After plasmid DNA was collected from all tubes, EtBr was removed by three extractions with 10ml of NaCl-saturated isopropanol each. The plasmid DNA remained in the bottom fraction while the top fraction retaining EtBr was removed. Afterwards, each sample was adjusted to a total volume of 5ml with dH_2O, mixed, transferred to 15ml Corex tubes (Corning Incorporated, Corning, NY) and precipitated with 10ml of EtOH at -20°C for 1 hour. The precipitated DNA was pelleted by centrifugation in a JA-20 rotor for 30min at 8K rpm, washed with 70% EtOH and dried under vacuum. Dried pellets were resuspended in 100µl TE, the concentrations determined by UV-absorption ($1OD_{260}$=50µg/ml DNA) and the remaining DNA stored at -20°C.

The DNA was then used in RNA transcription reactions and infectious transcripts were used to inoculate pea plants as described in chapter 2.

Northern and western blot analysis of infected plants:

Pea plants inoculated with infectious transcripts containing the mutation were analyzed by northern and western blot analysis for the presence of viral RNA and proteins, respectively.

Total RNA was extracted from infected pea plants based on a protocol by Demler & de Zoeten (1989). Care was taken to prevent any contamination with RNases during the procedure and preparation of solutions. Freshly cut plant tissue (0.6 - 1.0g for each sample) was placed at 4°C for approximately 10 min prior to the isolation. The tissue was first powdered by freezing in liquid nitrogen and grinding with a mortar and pestle. The pulverized tissue was transferred to 30ml Corex tubes (Corning) using a spatula. One sample at a time, 5ml of chilled, freshly made grinding buffer (4% w/v paraamino salicylic acid Na salt [PAS, Sigma], 1% triisopropylnaphtalene sulfonic acid [TNS, Kodak], 10mM DTT, 10mM Na meta bisulfite in DEPC-treated 50mM Tris-HCl pH 8.0) were added and later 4ml of Tris-saturated phenol while vortexing the sample between additions. All samples were placed on ice during processing of remaining samples and then centrifuged in a JA-20 rotor (Beckman) for 15 min at 9K rpm and 4°C in a Beckman J2-21 centrifuge. Each supernatant was transferred to another tube and proteins were removed by extraction with 2ml chloroform and centrifugation as indicated above. Again, the supernatant was removed and transferred to a 15ml Corex tube for Lithium Cloride (LiCl) precipitation overnight at -80°C. For each 5ml of supernatant 2.2ml of 10M LiCl

were added. The precipitated RNA was centrifuged in a JA20 rotor for 20min at 9K rpm and 4°C. Using a glass pipette, the supernatant was aspirated and discarded. The pellets were washed with 2ml of 2M LiCl, respun and dried under vacuum. Each pellet was resuspended in 300µl dH_2O, before transfer to an Eppendorf tube. Unsuspended pieces were removed by centrifugation for 5min at 14K rpm in an Eppendorf centrifuge and the supernatant was transferred to a new tube. An ethanol precipitation at -20°C for at least one hour followed after adding 30µl 3M NaOAc and 825µl ethanol. Precipitated RNA was centrifuged at 14K rpm for 20min, the supernatant aspirated and the remaining pellet washed with 70% EtOH, respun, dried under vacuum and resuspended in 200µl dH_2O. The RNA concentration was determined by UV absorbance at 260nm. Each RNA sample was separated on 1% non-denaturing agarose gels as described above.

Electrophoretically separated, total RNA was transferred to a nylon membrane (Micron Seprarations, Inc. (MSI), Westboro, MA) as part of the northern blotting procedure. For this purpose a "sandwich", described below, was assembled. Interfering bubbles between bread slices were removed by rolling a pipette firmly on the surface as each new layer was added. Four pieces of Whatman 3mm paper were cut to double-gel size, saturated in 10xSSC (1.5M NaCl, 0.15M Na citrate) and stacked on an oversized piece of Saran wrap. The gel was placed in the center of the Whatman paper after trimming off excess pieces. The oversized Saran wrap was then used to cover the exposed parts of the Whatman paper before the nylon membrane (MSI) was placed on top of the

gel. The membrane was cut to the same size of the gel, wetted in dH_2O and saturated in 10xSSC prior to the application. Three gel-size pieces of Whatman paper, saturated in 10xSSC, constituted the next layer. A stack of paper towels, cut to the approximate gel-size, and a weight centered on top of a glass plate completed the "sandwich". The next day, the membrane was removed, air dried and the RNA fixed by baking for 2 hours in a 68°C oven (Isotemp oven model 655G, Fisher Scientific). At this point, the membrane was either stored or used immediatelly.

The Genius 3 system "Nonradioactive Nucleic Acid Detection Kit" from Boehringer Mannheim was used to prepare RNA hybridization probes and perform the hybridization, washing and chemiluminescent detection of labeled RNA-RNA hybrids. To make digoxigenin (DIG) - labeled RNA probes, the DNA template was linearized at a restriction site. For the RNA 1 specific probe, 10 μg of DNA were linearized with 2μl HindIII (Boehringer Mannheim), while the same amount was linearized with 2μl PstI (Boehringer Mannheim) for the RNA 2 specific probe. Each restriction enzyme digest was carried out at 37°C for one hour in the optimal buffer as specified by the manufacturer. The DNA template was purified by phenol/chloroform extraction and ethanol precipitation.

During probe preparation great care was taken to prevent contamination of the samples with RNases. RNA was labeled during a transcription reaction for which the following reagents, in the indicated order, were added to 2μg DNA template in a microfuge tube on ice: 4μl 0.1M DTT, 4μl NTP labeling mixture, 8μl 5x transcription buffer (200mM Tris-HCl pH 8.0, 30mM $MgCl_2$, 50mM DTT, 10mM

spermidine, 50mM NaCl), 1μl T7 RNA polymerase and DEPC-dH_2O to adjust the total volume to 40μl. The solution was mixed, incubated at 37°C for 2 hours, and the reaction was stopped with 2μl EDTA. One μl glycogen was added to each reaction tube prior to RNA precipitation with 0.1 volume of 3M NaOAc and 3.0 volumes of chilled EtOH. After completion of a 30 min incubation at -80°C, the reactions were centrifuged for 15 min at 14K rpm in a microcentrifuge, the EtOH decanted, the pellet washed with 70% EtOH and centrifuged again at the same speed for 5min. The EtOH was removed prior to vacuum drying the pellet and resuspension in 100μl dH_2O. The resuspension was carried out at 37°C for 10 min under frequent vortexing. Labeled probes were stored at -80°C.

The hybridization procedure followed the labeling of probes and preparation of the northern blots. Duplicate blots were always prepared in order to probe each blot with a separate probe. Each blot was placed into a hybridization bag (Gibco BRL Life Technologies, Inc., Gaithersburg, MD) containing 20ml of northern prehybridization (prehyb) solution (50% formamide, 0.02% SDS, 0.1% N-lauroylsarcosine, 2% [w/v] blocking reagent for nucleic acid hybridization, 20mM sodium maleate pH 7.5 in 5xSSC pH 7.0 [0.75M NaCl, 75mM sodium citrate]) per 100cm^2 of membrane surface. The bags were sealed (Impulse Sealer, TEW Electric Heating Equipment Co., Ltd.), incubated for at least 2 hours in an oven at 68°C and the solution was discarded. Previously made, DIG-labeled probe was added to 25ml of northern prehyb solution to make the northern hybridization solution which was added to the membrane for an overnight incubation at 68°C. The next day, the hybridization solution

containing unannealed DIG-labeled probe was transferred into orange capped tubes (Corning) and stored at -80°C for future experiments. (For reuse, probes were thawed and heated for 10min to 68°C prior to use.) After the annealing step, blots were removed from the bag and placed into plastic trays for washes. Membranes were washed twice for 5 min in 1xSSC containing 0.1% SDS at RT and twice for 15min in 0.1xSSC + 0.1%SDS at 68°C.

As part of the chemiluminescent detection procedure, membranes were first washed in 50ml of maleate buffer (100mM maleic acid pH 7.5, 150mM NaCl) for 1 min prior to a 30 to 60min incubation in 50ml of northern blocking solution (10% blocking reagent stock solution [10% w/v blocking reagent for nucleic acid hybridization, 0.1M sodium maleate pH 7.5] diluted 1:5 in maleate buffer). All remaining steps were performed at RT. Next, anti-DIG-alkaline phosphatase (Boehringer Mannheim) diluted 1:5000 in blocking solution was used to replace the northern blocking solution and the membranes were incubated for 30min. Unbound antibody was discarded; the membrane was moved to a new plastic tray and washed twice with 200ml maleate buffer for 15min. Blots were equilibrated in 50ml of Genius 3 buffer (100mM Tris-HCl and 100mM NaCl adjusted to pH 9.5, 50mM $MgCl_2$) and placed between plastic page protectors. Afterwards, 0.5ml per 100cm^2 of Lumi-Phos 530 (Boehringerr Mannheim) was scattered and agitated gently for 1 min to distribute the reagent over the entire surface. The excess solution was then drained and the top sheet of plastic was lowered. All air bubbles were removed prior to covering with Kodak x-ray film (Kodak X-OMAT AR, 20.3 x 25.4 cm) in a film cassette holder (Kodak x-ray

exposure holder, 20.3 x 25.4 cm). Exposure times varied depending on signal strength. After adequate exposure, the film was developed.

Total protein isolation from leaves of infected pea plants, SDS-PAGE, electroblotting and western blot analysis were carried out as described above (see chapter 2). Anti PEMV Abs (diluted 1:5000 v/v in TBST containing 2% dried milk) were used as primary Abs preparations.

Results and Discussion

In order to introduce a nonsense mutation into the 33K ORF of RNA 2, the restriction enzyme BseR1 was chosen. The enzyme has only one restriction site in the cDNA clone at nucleotide (nt) 342 of RNA 2, one-third into the 33K ORF (nt 21-1001), and none in the plasmid. The following sequence is recognized by BseR1:

$$^{5'}\ GAGGAG(N)_{10}\ ^{3'}$$

$$^{3'}\ CTCCTC(N)_{8}\ ^{5'}$$

This results in a two base pair 3' overhang that was filled by the Klenow unit of DNA polymerase I to create blunt ends. Successful cloning was verified by genomic sequencing in the target region due to the unusual restriction characteristics of the enzyme, which has a recognition sequence different from the cleavage site. Two clones, UORF1-12 and UORF1-13, containing an identical two bp deletion (nt 341 and 342), were identified and used for subsequent experiments. Figure 4 shows the sequence in the target region and

the aa sequences before and after mutagenesis. The early termination of the protein results in a putative protein of 110aa instead of the 326 aa full-length protein.

Both mutant clones were then purified in large scale purifications and infectious transcripts were produced for inoculation of young pea plants. Visual comparison of symptoms between plants infected with wild-type or mutant clones, northern and western blotting techniques to analyze RNA replication and expression of viral proteins were used, respectively, to study the effect of the mutation in plants.

Four different sets of pea seedlings were inoculated with different combinations of wild-type (wt) and mutant clones: mock inoculated, wt RNA 1 and RNA 2, RNA 2 alone, RNA 1 and UORF1-12, RNA 1 and UORF1-13, UORF1-12 alone, UORF1-13 alone. Each time, one pot with four to five plants was used for mock or wt inoculations and two pots (8-10 plants) for each mutant inoculation. The majority of plants in each treatment group showed typical PEMV symptoms after ten days post-inoculation, excluding the mock-inoculated plants which had no visible, virally induced symptoms. No difference in severity of the symptoms could be distinguished between wt and mutant inoculated plants. Symptom onset appeared in some mutant infected plants one to two days delayed. However, due to the subjective nature of these observations no definitive statement can be made.

Total RNA and total protein extractions were obtained from a symptomatic plant of each treatment. Northern and western blot analysis showed differing

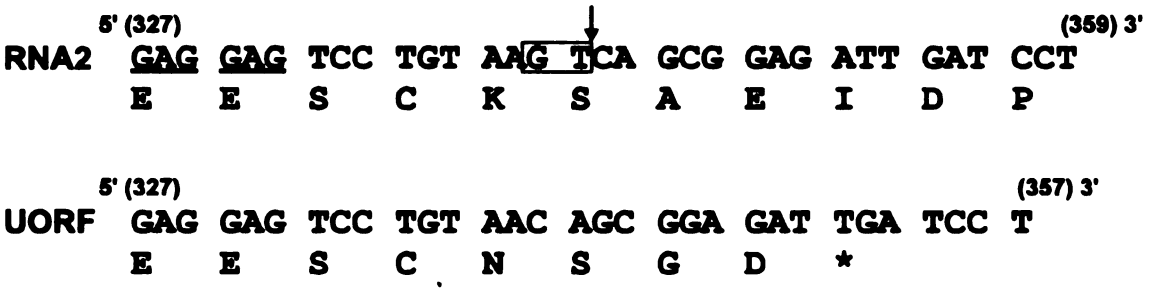

Figure 5: Mutated region of ORF1 (33kDa) of PEMV RNA2. A stop codon was introduced in the ORF1 at nucleotides 351 to 353 that resulted in an early termination of the 33kDa protein after 110 amino acids (see Materials and Methods in Appendix A for details). The nucleotide and amino acids sequence of RNA2 and mutant UORF are shown. The location of the nucleotides in RNA2 is indicated above the nucleotide sequence in brackets as well as the orientation of the RNA. The BseR1 recognition site is underlined and the cleavage site is indicated by the arrow in the RNA2 sequence. The boxed nucleotides were deleted during the mutagenesis, causing the frameshift. The stop codon is indicated by the star (*).

results for each isolation. However, taken all the data together, there appears to be no obvious difference in the RNA or protein pattern between wt and mutant infected plants. Due to the limited number of samples and technical problems, it is difficult to draw general conclusions and the results presented below should be viewed with caution.

Northern blotting procedures used RNA 1 and RNA 2 specific probes which hybridized to the viral polymerase of each genomic RNA. Therefore, a positive signal indicated replication of the respective RNA. Table 1 shows a graphical representation of the obtained results indicating the detection of a positive (+) or negative (-) signal. Total RNA extractions from each of the four sets of inoculations (A-D) were tested. Two of the four RNA extractions were tested twice (A and B) and the others (C and D) only once. As expected, total RNA of mock inoculated plants showed no signal when tested with either RNA 1 or RNA 2 specific probes in all cases, indicating that these plants were not infected with PEMV. Also, no signal was seen when total RNA of RNA 2, UORF1-12 and UORF1-13 inoculated plants were tested with the RNA 1 specific probe. The absence of a RNA 1 specific signal was expected in these samples since the plants were only infected with infectious clones of wt and mutant RNA 2 but not with RNA 1. Therefore no contamination with RNA 1 occurred in these samples. Total RNA of RNA1 + UORF1-12 or UORF1-13 inoculated plants showed the same infectivity pattern (see table 1) as total RNA of RNA1 + RNA2 infected plants. Some but not all samples of the three treatments gave positive

results. The result, that the wt infections (RNA1 + RNA2) were not always successful, suggest that the negative results obtained for some mutant treatments are likely due to technical problems, like bad probes or unsuccessful inoculations, rather than a property that can be attributed to the mutated 33K protein. When UORF1-12 and UORF1-13 where inoculated alone, the isolated total RNA showed fewer positive signals than total RNA from plants infected with RNA2 alone, after probing with the RNA2 specific probe. The reason for this is not known. It could be due to either experimental inconsistencies, or to the mutation in the 33K ORF of RNA2. If the latter was the case, it might suggest that the presence of RNA 1 has a positive effect that overcame the mutation.

Total protein was isolated from three sets of inoculations (B-D). Proteins were isolated, when possible, from leaves of plants also used for total RNA extractions. Two of the three sets of inoculations (C and D) showed virus-specific signals in the positive control (total protein extracts from RNA1 + RNA2 infected plants) when probed with anti-PEMV Abs. One of the western blots did not show the expected signals in the positive control and was, therefore, excluded from the discussion. A typical total protein extraction from RNA1 and RNA2 infected plants probed with anti-PEMV Abs show three to four specific bands: the 21kDa coat protein, a 15kDa doublet representing the VPg, and sometimes an 18kDa protein that appears erratically in different experiments (see Figure 2, lane 3). However, total protein extracts of RNA 2 infected plants do not show these PEMV-specific bands (see Figure 2, lane 4). Therefore as expected, these bands were not seen when total protein extract of mock-

	A		A		B		B		C		D	
probe	RNA1	RNA2	RNA1	RNA2	RNA1	RNA2	RNA1	RNA2	RNA1	RNA2	RNA1	RNA2
treatment												
mock	-	-	-	-	-	-	-	-	-	-	-	-
1+2	-	-	-	+ bs	-	+ s	-	+ bs	+ s	+ bs	+ s	+ bs
RNA 2	-	+ b	-	+ bs	-	+ s	-	+ bs	-	+ bs	-	+ bs
1+12	-	-	-	-	-	+ s	-	+ bs	-	+ bs	-	+ bs
UORF 12	-	-	- g	- g	+ s	+ s	+ bs	+ bs	+ s	+ bs	+ s	+ bs
1+13	-	+ b	-	+ bs	-	+ s	- g	- g	-	-	-	+ bs
UORF 13	-	-	-	+ bs	+ s	+ s	+ bs	+ bs	- g	- g	+ s	+ bs

Table 1: Northern blot analysis of plants infected with wild-type and/or UORF infectious clones. Treatments are indicated in the left column while the different sets of inoculations (A-D) are indicated in the top row. The RNA1 or RNA2 specific probe used during Northern blot analysis is indicated for each treatment. Abbreviations: positive signal (+), negative signal (-), b = clear bands, bs = clear bands but associated with a smear, s = smear but no clear band, g = RNA was degraded in the agarose gel before transfer and no signal is expected.

inoculated, RNA 2, UORF1-12 or UORF1-13 infected plants were tested in this study. However, when mutant or wt RNA 2 were infected together with RNA 1 and total proteins were isolated and tested, the results differed between the two western blots. In one western blot, total protein extracts of RNA 1 and RNA 2 infected plants reacted specifically with the 21kDa cp, the 18kDa protein and the 15kDa VPg doublet. Only total protein extract of RNA1 + UORF1-12 but not RNA1 + UORF1-13 infected plants reacted with the cp and the 15kDa protein doublet. No conclusion can be drawn about the 18kDa protein signal because of its erratic nature. It is not clear why the total protein extraction of the RNA1 + UORF1-13 inoculated plant from set C did not show any virus-specific signals when probed with anti-PEMV Abs. In the other western blot from set D, the 15kDa protein doublet was detected in total protein extracts of RNA1 + 2, RNA1 + UORF1-12 and RNA1 + UORF1-13 infected plants, but neither the cp nor the 18kDa protein. It is not clear, why no cp signal was detected. More repetitions of this experiment are needed to allow a statistical conclusion. However, it appears that no distinctive difference in the virus-specific protein signals was detected in total protein extracts of RNA1 + UORF versus RNA1 + 2 infected plants. The cp and the 15kDa protein were detected in at least one case in the mutant infected plant. This could indicate that these proteins were not affected by the mutation in the 33K ORF of RNA 2. Therefore, the 33K gene product is unlikely to regulate the expression of either protein.

As already indicated, this data is preliminary and more repetitions of northern and western blots are needed to derive any statistical data. The 33K

gene product does not appear to be absolutely essential. Infectious clones carrying the mutation remained infectious either in conjunction with RNA 1 or alone. Analyzing viral RNA from mutant infected plants by PCR and genomic sequencing is one of the several things necessary to come to a definitive conclusion.

In conclusion, an infectious clone containing a mutation in the 33K ORF was constructed. However, we were unable to obtain data indicating a possible function of this ORF gene product. As already indicated, more repetitions of the described experiments are needed to make this preliminary data statistically relevant. In addition, supplemental experiments are needed. This would include PCR and sequence analysis of viral RNA from mutant infected plants but also additional constructs. The described mutant construct results in a protein one third of its original length. It will also be necessary to completely eliminate the 33K ORF to determine which function resides in the N-terminal portion of the protein.

List of References

List of References

Adam, G. (1978). Die gemeinsame Übertragung eines vektoriellen und avektoriellen Stammes des Pea Enation Mosaik Virus durch die Blattlaus *Acyrthisophon pisum*. *Zeitschrift für Pflanzenkrankheiten und Pflanzenschutz* **85**, 586-593.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology* **215**, 403-410.

Andrews, N. C. & Baltimore, D. (1986). Purification of a terminal uridylyl transferase that acts as host factor in the *in vitro* poliovirus replicase reaction. *Proceedings of the National Academy of Sciences of the United States of America* 83, 221-225.

Andrews, N. C., Levin, D. & Baltimore, D. (1985). Poliovirus replicase stimulation by terminal uridylyl transferase. *The Journal of Biological Chemistry* **260**, 7628-7635.

Arnold, E., Luo, M., Vriend, G., Rossmann, M. G., Palmenberg, A.C., Parks, G. D., Nicklin, M. J. H. & Wimmer, E. (1987). Implications of the picornavirus capsid structure for polyprotein processing. *Proceedings of the National Academy of Sciences of the United States of America* **84**, 21-25.

Baron, M. H. & Baltimore, D. (1982). Anti-VPg antibody inhibition of the poliovirus replicase reaction and production of covalent complexes of VPg-related proteins and RNA. *Cell* **30**, 745-752.

Barton, D. J., Black, E. P. & Flanegan, J. B. (1995). Complete replication of poliovirus in vitro: preinitiation of RNA replication complexes require soluble cellular factors for the synthesis of VPg-linked RNA. *Journal of Virology* **69**, 5516-5527.

Berger, A. & Schlechter, I. (1977). Mapping the active site of papain with the aid of peptide substrates and inhibitors. *Philosophical Transactions of the Royal Society of London, Series B* **257**, 249-264.

Blair, W. S. & Semler, B. L. (1991). Role for the P4 amino acid residue in substrate utilization by the poliovirus 3CD proteinase. *Journal of Virology* **65**, 6111-2123.

Carrington, J. C., Haldeman, R. Dolja, V. V. & Retsrepo-Hartwig, M. A. (1993). Internal cleavage and *trans*-proteolytic activities of the VPg-proteinase (Nla) of tobacco etch potyvirus in vivo. *Journal of Virology* **67**, 6995-7000.

Chu, P. W. G., Boccardo, G. & Francki, R. I. B. (1981). Requirement of a genome-associated protein of tobacco ringspot virus for infectivity but not for *in vitro* translation. *Virology* **109**, 428-430.

Cordingley, M. G., Callahan, P. L., Sardana, V. V., Garskky, V. M. & Colonno, R. J. (1990). Substrate requirements of human rhinovirus 3C protease for peptide cleavage in vitro. *The Journal of Biological Chemistry* **265**, 9062-9065.

Crawford, N. M. & Baltimore, D. (1983). Genome-linked protein VPg of poliovirus is present as free VPg and VPg-pUpU in poliovirus-infected cells. *Proceedings of the National Academy of Sciences of the United States of America* **80**, 7452-7455.

Daubert, S. D. & Bruening, G. (1979). Genome-associated proteins of Comoviruses. *Virology* **98**, 246-250.

Daubert, S. D. & Bruening, G. (1984). Detection of genome-linked proteins of plant and animal viruses. *Methods in Virology* **8**, 347-379.

Daubert, S. D., Bruening, G. & Najarian, R. C. (1978). Protein bound to the genomic RNAs of cowpea mosaic virus. *European Journal of Biochemistry* **92**, 45-51.

Demler, S. A. & de Zoeten, G. A. (1989). Characterization of a satellite RNA associated with pea enation mosaic virus. *Journal of General Virology* **10**, 1075-1084.

Demler, S. A. & de Zoeten, G. A. (1991). The nucleotide sequence and luteovirus-like nature of RNA 1 of an aphid non-transmissible strain of pea enation mosaic virus. *Journal of Virology* **72**, 1819-1834.

Demler, S. A., Rucker, D. G. & de Zoeten, G. A. (1993). The chimeric nature of the genome of pea enation mosaic virus: the independent replication of RNA 2. *Journal of Virology* **74**, 1-14.

Demler, S. A., Borkhsenious, O. N., Rucker, D. G. & de Zoeten, G. A. (1994a). Assessment of the autonomy of replicative and structural functions encoded by the luteo-phase of pea enation mosaic virus. *Journal of General Virology* **75**, 997-1007.

Demler, S. A., Rucker, D. G. & de Zoeten, G. A. (1994b). Replication of the satellite RNA of pea enation mosaic virus is controlled by RNA 2-encoded functions. *Journal of General Virology* **75**, 1399-1406.

Demler, S. A., de Zoeten, G. A., Adam, G. & Harris, K. F. (1996a). Pea Enation Mosaic Enamovirus: properties and aphid transmission. In *The plant viruses. Volume 5: Polyhedral virions and bipartite RNA genomes,* 1st edn, pp. 303-344. Edited by B. D. Harrison & A. F. Murrant. New York: Plenum Press.

Demler, S. A., Rucker, D. G., de Zoeten, G. A., Ziegler, A., Robinson, D. J. & Murant, A. F. (1996b). The satellite RNAs associated with the groundnut rosette disease complex and pea enation mosaic virus: sequence similarities and ability of each other's helper virus to support their replication. *Journal of General Virology* **77**, 2847-2855.

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

de Varennes, A., Lomonossoff, G. P., Shanks, M. & Maule, A. J. (1986). The stability of cowpea mosaic virus VPg in reticulocyte lysates. *Journal of General Virology* **67**, 2347-2354.

Dewalt, P. G. & Semler, B. L. (1989). Molecular biology and genetics of poliovirus protein processing. In *Molecular Aspects of Picornavirus Infection and Detection.* Edited by B. L. Semler & E. Ehrenfeld, pp. 73-93, Washington, D. C.: American Society for Microbiology.

de Zoeten, G. A., Gaard, G. & Diez, F. B. (1972). Nuclear vesiculation associated with pea enation mosaic virus-infected plant tissue. *Virology* **48**, 638-647.

de Zoeten, G. A., Powell, C. A., Gaard, G. & German, T. L. (1976). *In situ* localization of pea enation mosaic virus double-stranded ribonucleic acid. *Virology* **70**, 459-469.

de Zoeten, G. A. & Gaard, G. (1983). Mechanisms underlying systemic invasion of pea plants by pea enation mosaic virus. *Intervirology* **19**, 85-94.

de Zoeten, G. A. & Demler, S. A. (1995). Genus Enamovirus. In *Virus Taxonomy - Classification and Nomenclature of viruses. Sixth Report of the International Committee of Taxonomy of Viruses,* 1st edn, pp. 384-387. Edited by F. A. Murphy, C. M. Fauquet, D. H. L. Bishop, S. A. G. Jarvis, G. P. Martelli, M. A. Mayo & M. D. Summer. Vienna: Springer Verlag.

Dolja, V. V. & Carrington, J. C. (1992). Evolution of positive-strand RNA viruses. *Seminars in Virology* **3**, 315-326.

Domier, L. L., Shaw, J. G. & Rhoads, R. E. (1987). Potyviral proteins share amino acid sequence homology with picorna-, como- and caulimoviral proteins. *Virology* **158**, 20-27.

Dougherty, W. G., Cary, S. M. & Parks, T. D. (1989). Molecular and genetic analysis of a plant virus polyprotein cleavage site: a model. *Virology* **171**, 356-364.

Dougherty, W. G. & Parks, T. D. (1989). Molecular genetic and biochemical evidence for the involvement of the heptapeptide cleavage sequence in determining the reaction profile at two tobacco etch virus cleavage sites innn cell-free assays. *Virology* **172**, 145-155.

Dougherty, W. G. & Parks, T. D. (1991). Post-translational processing of the tobacco etch virus 49-kDa small nuclear inclusion polyprotein: identification of an internal cleavage site and delimitation of VPg and proteinase domains. *Virology* **183**, 449-456.

Dougherty, W. G. & Semler, B. L. (1993). Expression of virus-encoded proteinases: functional and structural similarities with cellular enzymes. *Microbiological Reviews* **57**, 781-822.

Drygin, Y. F., Sapotsky, M. V. & Bogdanov, A. A. (1987). Radish mosaic virus VPg. *FEBS Letters* **215**, 247-251.

Ehrenfeld, E. & Richards, O. C. (1989). Studies of poliovirus RNA polymerase expressed in Escherichia coli: attempts to understand viral RNA replication. In *Molecular Aspects of Picornavirus Infection and Detection*. Edited by B. L. Semler & E. Ehrenfeld, pp. 95-105, Washington, D. C.: American Society for Microbiology.

Falk, B. W. & Duffus, J. E. (1981). Epidemiology of helper-dependent persistent aphid transmitted virus complexes. In *Plant Diseases and Vectors: Ecology and Epidemiology*. Edited by K. Maramorosh & K. H. Harris, pp. 161-179, New York: Academic Press.

Fargette, D., Jenniskens, M.-J. & Peters, D. (1982). Acquisition and transmission of pea enation mosaic virus by the individual pea aphid. *Phytopathology* **72**, 1386-1390.

Flanegan, J. B. & Baltimore, D. (1979). Poliovirus polyuridylic acid polymerase and RNA replicase have the same viral polypeptide. *Journal of Virology* **29**, 352-360.

Flanegan, J. B., Pettersson, R. F., Ambros, V., Hewlett, M. J. & Baltimore, D. (1977). Covalent linkage of a protein to a defined nucleotide sequence at the 5'-terminus of virion and replicative intermediate RNAs of poliovirus. *Proceedings of the National Academy of Sciences of the United States of America* **74**, 961-965.

Gabriel, C. J. (1983). Investigation into the *in vitro* translation and aphid transmission of pea enation mosaic virus. *Ph.D. thesis*, University of Wisconsin, Madison, WI.

Gabriel, C. J. & de Zoeten, G. A. (1984). The *in vitro* translation of pea enation mosaic virus RNA. *Virology* **139**, 223-230.

German, T. L. & de Zoeten, G. A. (1975). Purification and properties of the replicative forms and replicative intermediates of Pea Enation Mosaic Virus. *Virology* **66**, 172-184.

German, T. L., de Zoeten, G. A. & Hall, T. C. (1978). Pea Enation mosaic virus RNA contains no polyadenylate sequences and can not be aminoacylated. *Intervirology* **9**, 226-230.

Gildow, F. E. & Rochow, W. F. (1980). Role of accessory salivary glands in aphid transmission of barley yellow dwarf virus. *Virology* **104**, 97-108.

Goldbach, R., Rezelman, G., Zabel, P. & 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 VPg and a 58-kilodalton polypeptide. *Journal of Virology* **42**, 630-635.

Gorbalenya, A. E. & Koonin, E. V. (1993). Comparative analysis of amino-acid sequences of key enzymes of replication and expression of positive-strand RNA viruses: validity of approach and functional and evolutionary implications. *Soviet Scientific Review, Section D: Physicochemical Biology* **11**, 1-84.

Gosh, A., Dasgupta, R., Salerno-Rife, T., Rutgers, T. & Kaesberg, P. (1979). Southern bean mosaic virus RNA has a 5'-linked protein but lacks 3-terminal poly(A). *Nucleic Acids Research* **7**, 2137-2146.

Hagedorn, D. J. (1996). Pea enation mosaic virus: ecology and control. In *The plant viruses. Volume 5: Polyhedral virions and bipartite RNA genomes,* 1st edn,

pp. 345-356. Edited by B. D. Harrison & A. F. Murrant. New York: Plenum Press.

Harlow, E. & D. Lane. (1988). *Antibodies: a laboratory manual.* 1st. edn. Cold Spring Harbor, N. Y.: Cold Spring Harbor Press.

Harris, K. F. & Bath, J. E. (1972). The fate of pea enation mosaic virus in its pea aphid vector, *Acyrthosiphon pisum* (Harris). *Virology* **50**, 778-790.

Harris, K. F., Bath, J. E., Thottappilly, G. & Hooper, G. R. (1975). Fate of pea enation mosaic virus in PEMV-injected pea aphids. *Virology* **65**, 148-162.

Harris, K. F. (1979). Leafhoppers and aphids as biological vectors: vector-virus relationships. In *Leafhopper Vectors and Plant Disease Agents*, pp 217-308. Edited by K. Maramorosch & K. F. Harris. New York: Academic Press.

Harris, K. F. (1990). Aphid transmission of plant viruses. In *Plant Viruses, Vol. 2, Pathology*, pp. 177-204. Edited by C. L. Mandahar. Boca Raton, FL: CRC Press.

Harris, K. S., Reddigari, S. R., Nicklin, M. J. H., Hämmerle, T. & Wimmer, E. (1992). Purification and characterization of poliovirus polypeptide 3CD, a proteinase and precursor for RNA polymerase. *Journal of Virology* **66**, 7481-7489.

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

Hey, T. D., Richards, O. C. & Ehrenfeld, E. (1987). Host factor-induced template modification during synthesis of poliovirus RNA in vitro. *Journal of Virology* **61**, 802-811.

Hill, J. H. & Shepherd, R. J. (1972). Molecular weights of plant virus coat proteins by polyacrylamide gel electrophoresis. *Virology* **47**, 817-822.

Hull, R. & Lane, L. C. (1973). The unusual nature of the components of a strain of pea enation mosaic virus. *Virology* **55**, 1-13.

Jaegle, M., Wellink. J. & Goldbach, R. (1987). The genome-linked protein of cowpea mosaic virus is bound to the 5' terminus of virus RNA by a phosphodiester linkage to serine. *Journal of General Virology* **68**, 627-632.

Jayaram, C., Hill, J. H. & Miller, W. A. (1992). Complete nucleotide sequences of two soybean mosaic virus strains differentiated by response of soybean containing the *Rsv* resistance gene. *Journal of General Virology* **73**, 2067-2077.

Koenig, I. & Fritsch. C. (1982). A protein linked at the 5' end of satellite and genomic tomato black ring virus RNAs: Study of in vitro translation after protease treatment. *Journal of General Virology* **60**, 343-353.

Koonin, E. V. (1991). The phylogeny of RNA-dependent RNA polymerases of positive-strand RNA viruses. *Journal of General Virology* **72**, 2197-2206.

Koonin, E. V. & Dolja, V. V. (1993). Evolution and taxonomy of positive-strand RNA viruses: implications of comparative analysis of amino acid sequences. *Critical Reviews in Biochemistry and Molecular Biology* **28**, 375-430.

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

Laliberte, J.-F., Nicolas, O., Chatel, H., Lazure, C. & Morosoli, R. (1992). Release of a 22-kDa protein derived from the amino-terminal domain of the 49-kDa NIa of Turnip Mosaic Potyvirus in *Escherichia coli*. *Virology* **190**, 510-514.

Lee, S & Rasheed, S. (1990). A simple procedure for maximum yield of high-quality plasmid DNA. *Biotechniques* **9**, 676-679.

Lomonossoff, G. P., Shanks, M. & Evans, D. (1985). The structure of cowpea mosaic virus replicative form RNA. *Virology* **144**, 351-362.

Long, A. C., Orr, D. C., Cameron, J. M., Dunn, B. M. and Kay, J. (1989). A consensus sequence for substrate hydrolysis by rhinovirus 3C proteinase. *FEBS Letters* **258**, 75-78.

Liu, Y. Y., Cooper, J. I., Coates, D. & Bauer, G. (1991). Biologically active transcripts of a large satellite RNA from arabis mosaic nepovirus and the importance of 5' end sequences for its replication. *Journal of General Virology* **72**, 2867-2874.

Mäkinen, K., Tamm, T., Naess, V., Truve, E., Puurand, Ü., Munthe, T. Saarma, M. (1995). Characterization of cocksfoot mottle sobemovirus genomic RNA and sequence comparison with related viruses. *Journal of General Virology* **76**, 2817-2825.

Margis, R., Hans, F. & Pinck, L. (1993). VPg northern-immunoblots as a means for detection of viral RNAs in protoplasts or plants infected with grapevine fanleaf virus. *Archives of Virology* **131**, 225-232.

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

Martin, R. R., Keese, P. K., Young, M. J., Waterhouse, P. M. & Gerlach, W. L. (1990). Evolution and molecular biology of luteoviruses. *Annual Reviews in Phytopathology* **28**, 341-363.

Matthews, R. E. F. (1991). Plant Virology. 3rd ed. Academic Press, Inc. San Diego, CA.

Matthews, D. A., Smith, W. W., Ferre, R. A., Condon, B., Budahazi, G., Sisson, W., Villafranca, J. E., Janson, C. A., McElroy, H. E., Gribskov, C. L. & Worland, S. (1994). Structure of human rhinovirus 3C protease reveals a trypsin-like polypeptide fold, RNA binding site, and means for cleaving precursor polyprotein. *Cell* **77**, 761-771.

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

Mayo, M. A., Barker, H., Robinson, D. J., Tamada, T. & Harrison, B. D. (1982). Evidence that potato leafroll virus RNA is positive-stranded, is linked to a small protein and does not contain polyadenylate. *Journal of General Virology* **59**, 163-167.

Mayo, M. A. & Fritsch, C. (1994). A possible consensus for VPg of viruses in the family *Comoviridae*. *FEBS Letters* **354**, 129-130.

Mayo, M. A. & Ziegler-Graff, V. (1996). Molecular biology of luteoviruses. *Advances in Virus Research* **46**, 413-460.

Miller, W. A., Waterhouse, P. M. & Gerlach, W. L. (1988). Sequence and organization of barley yellow dwarf virus genomic RNA. *Nucleic Acids Research* **16**, 6097-6111.

Miller, W. A. (1994). Luteoviruses. In *Ecyclopedia of Virology*, 1st edn, pp. 792-798. Edited by R. G. Webster & A. Granoff. London: Academic Press.

Miller, W. A., Dinesh-Kumar, S. P. & Paul, S. P. (1995). Luteovirus gene expression. *Critical Reviews in Plant Sciences* **41**, 179-211.

Mirzayan, C. & Wimmer, E. (1994). Polioviruses: Molecular Biology. In *Encyclopedia of Virology*, 1st edn, pp. 1119-1132. Edited by G. A. Webster & A. Ganoff. London: Academic Press.

Meyer, M., Hemmer, O. & Fritsch, C. (1984). Complete nucleotide sequence of a satellite RNA of tomato black ring virus. *Journal of General Virology* **65**, 1575-1583.

Murphy, J. F., D'Arcy, C. J. & Clark, J. M. JR. (1989). Barley yellow dwarf virus RNA has a 5' -terminal genome-linked protein. *Journal of General Virology* **70**, 2253-2256.

Murphy, J. F., Rhoads, R. E., Hunt, A. G. & Shaw, J. G. (1990). The VPg of tobacco etch virus RNA is the 49-kDa proteinase or the N-terminal 24-kDa part of the proteinase. *Virology* **178**, 285-288.

Murphy, J. F., Rychlik, W., Rhoads, R. E., Hunt, A. G. & Shaw, J. G. (1991). A tyrosine residue in the small nuclear inclusion protein of tobacco vein mottling virus links the VPg to the viral RNA. *Journal of Virology* **65**, 5511-513.

Murphy, J. F., Klein. P. G., Hunt, A. G. & Shaw, J. G. (1996). Replacement of the tyrosine residue that links a potyviral VPg to the viral RNA is lethal. *Virology* **220**, 535-538.

Nomoto, A., Detjen, B., Pozzatti, R. & Wimmer, E. (1977a). The location of the polio genome protein in viral RNAs and its implication for RNA synthesis. *Nature* **268**, 208-213.

Nomoto, A., Kitamura, N., Golini, F. & Wimmer, E. (1977b). The 5' -terminal structures of poliovirion RNA and poliovirus mRNA differ only in the genome-linked protein VPg. *Proceedings of the National Academy of Sciences of the United States of America* **74**, 5345-5349.

Pallai, P. V., Burkhardt, F., Skoog, M., Schreiner, K., Bax, P., Cohen, K. A., Hansen, G., Palladino, D. E. H., Harris, K. S., Nicklin, M. J. & Wimmer, E. (1989). Cleavage of synthetic peptides by purified poliovirus 3C proteinase. *Journal of Biological Chemistry* **264**, 9738-9741.

Peters, S. A., Voorhorst, W. G. B., Wellink, J. & van Kammen, A. (1992). Processing of VP-containing polyproteins encoded by the B-RNA from cow pea mosaic virus. *Virology* **191**, 90-97.

Peters, S. A., Mesnard, J.-M., Kooter, I. M., Verver, J., Wellink, J. & van Kammen, A. (1995). The cowpea mosaic virus RNA 1-encoded 112kDa protein

may function as a VPg precursor *in vivo*. *Journal of General Virology* **76**, 1807-1813.

Pettersson, R. F., Ambros, V. & Baltimore, D. (1978). Identification of a protein linked to nascent poliovirus RNA and to the polyuridylic acid of negative-strand RNA. *Journal of Virology* **27**, 357-365.

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

Powell, C. A., de Zoeten, G. A. & Gaard, G. (1977). The localization of pea enation mosaic virus-induced RNA-dependent RNA polymerase in infected peas. *Virology* **78**, 135-143.

Powell, C. A. & de Zoeten, G. A. (1977). Replication of pea enation mosaic virus RNA in isolated pea nuclei. *Proceedings of the National Academy of Sciences of the United States of America* **74**, 2919-2922.

Randles, J. W. & Rathjen, J. P. (1995). Genus Luteovirus. In *Virus Taxonomy - Classification and Nomenclature of viruses. Sixth Report of the International Committee of Taxonomy of Viruses,* 1st edn, pp. 379-383. Edited by F. A. Murphy, C. M. Fauquet, D. H. L. Bishop, S. A. G. Jarvis, G. P. Martelli, M. A. Mayo & M. D. Summer. Vienna: Springer Verlag.

Reisman, D. & de Zoeten, G. A. (1982). A covalently linked protein at the 5' end of the genomic RNAs of Pea Enation Mosaic Virus. *Journal of General Virology* **62**, 187-190.

Restrepo-Hartwig, M. A. & Carrington, J. C. (1994). The tobacco etch potyvirus 6-kilodalton protein is membrane associated and involved in viral replication. *Journal of Virology* **68**, 2388-2397.

Reutenauer, A., Ziegler-Graff, V., Lot, H., Scheidecker, D., Guilley, H., Richards, K. & Jonard, J. (1993). Identification of beet western yellows luteovirus genes implicated in viral replication and particle morphogenesis. *Virology* **195**, 692-699.

Riechmann, J. L., Lain, S. & Garcia, J. A. (1989). The genome-linked protein and 55' end RNA sequence of plum pox potyvirus. *Journal of General Virology* **70**, 2785-2789.

Riechmann, J. L., Lain, S. & Garcia, J. A. (1993). Highlights and prospects of potyvirus molecular biology. *Journal of General Virology* **73**, 1-16.

Roosinck, M. J., Sleat, D. & Palukaitis, P. (1992). Satellite RNAs of plant viruses: structures and biological effects. *Microbiological Reviews* **56**, 265-279.

Rost, B. & Sander, C. (1993). Prediction of protein structure at better than 70% accuracy. *Journal of Molecular Biology* **232,** 584-599.

Rost, B. & Sander, C. (1994a). Combining evolutionary information and neural networks to predict protein secondary structure. *Proteins* **19,** 55-72.

Rost, B. & Sander, C. (1994b). Conservation and prediction of solvent accessibility in protein families. *Proteins* **20,** 216-226.

Rost, B. (1996). PHD: predicting one-dimensional protein structures by profile based neural networks. *Methods in Enzymology* **266,** 525-539.

Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989). Molecular Cloning: a laboratory manual, 2nd edn, pp. 18.1 - 18.88. Cold Spring Harbor, N. Y.: Cold Spring Harbor Laboratory Press.

Sanger, F., Nicklen, S & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America* **74**, 5463-5467.

Schaad, M. C., Haldeman-Cahill, R., Cronin, S. & Carrington, J. C. (1996). Analysis of the VPg-proteinase (Nla) encoded by tobacco etch potyvirus: Effects of mutations on subcellular transport, proteolytic processing, and genome amplification. *Journal of Virology* **70**, 7039-7048.

Semler, B. L., Kuhn, R. J. & Wimmer, E. (1988). Replication of the poliovirus genome. In *RNA Genetics*, pp. 23-48. Edited by E. Domingo, J. J. Holland & P. Alquist. Boca Raton, Fl: CRC Press.

Shahabuddin, M., Shaw, R. G. & Rhoads, R. E. (1988). Mapping of the tobacco vein mottling virus VPg cistron. *Virology* **163**, 635-637.

Shepherd, R. J., Wakeman, R. J. & Ghabrial, S. A. (1968). Preparation and properties of the protein and nucleic acid components of pea enation mosaic virus. *Virology* **35**, 255-267.

Stanley, J., Rottier, P., Davies, J. W., Zabel, P. & van Kammen, A. (1978). A protein linked to the 5' termini of both RNA components of the cowpea mosaic virus genome. *Nucleic Acids Research* **5**, 45-5-4522.

Stryer, L. (1988). Biochemistry. 3rd ed. W. H. Freeman and Company, New York.

Tacke, E., Prüfer, D., Schmitz, J. & Rohde, W. (1991). The potato leafroll luteovirus 17K protein is a single-stranded nucleic acid-binding protein. *Journal of General Virology* **72**, 2035-2038.

Tacke, E., Schmitz, J., Prüfer, D. & Rohde W. (1993). Mutational analysis of the nucleic-acid binding 17kDa phosphoprotein of potato leafroll luteovirus identifies an amphipathic α-helix as the domain for protein/protein interactions. *Virology* **197**, 274-282.

Vance, V. B., Moore, D., Turpen, T. H., Bracker, A. & Hollowell, V. C. (1992). The complete nucleotide sequence of pepper mottle virus genomic RNA: Comparison of the encoded polyprotein with those of other sequenced potyviruses. *Virology* **191**, 19-30.

van der Wilk, F., Huisman, M. J., Cornelissen, B. J. C., Huttinga, H. & Goldbach, R. (1989). Nucleotide sequence and organization of potato leafroll virus genomic RNA. *FEBS Letters* **245**, 51-56.

van Dyke, T. A. & Flanegan, J. B. (1980). Identification of poliovirus polypeptide P63 as a soluble RNA - dependent RNA polymerase. *Journal of Virology* **35**, 732-740.

Vartapetian, A. B. & Bogdanov, A. A. (1987). Proteins covalently linked to viral genomes. *Progress in Nucleic Acid Research and Molecular Biology* **34**, 209-251.

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

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

Wimmer, E. (1982). Genome-linked proteins of viruses. *Cell* **28**, 199-201.

Yassi, M. N. A., Ritzenthaler, C., Brugidou, C., Fauquet, C. & Beachy, R. N. (1994). Nucleotide sequence and genome characterization of rice yellow mottle virus RNA. *Journal of General Virology* **75**, 249-257.

Young, D. C., Tuschall, D. M. & Flanegan, J. B. (1985). Poliovirus RNA-dependent RNA polymerase and host cell protein synthesize product RNA twice the size of poliovirus RNA in vitro. *Journal of Virology* **54**, 256-264.

Young, D. C., Dunn, B. M., Tobin, G. J. & Flanegan, J. B. (1986). Anti-VPg antibody precipitation of product RNA synthesized in vitro by the poliovirus polymerase and host factor is mediated by VPg on the poliovirus RNA template. *Journal of Virology* **58**, 715-723.

Ympa-Wong, M. F., Filman, D. J., Hogle, J. M. & Semler, B. L. (1988). Structural domains of the poliovirus polyprotein are major determinants for proteolytic cleavage at Gln-Gly pairs. *The Journal of Biological Chemistry* **263**, 17846-17856.

Zabel, P., Moerman, M., van Straaten, F., Goldbach, R. & van Kammen, A. (1982). Antibodies against the genome-linked protein VPg of Cowpea mosaic virus recognize a 60,000-Dalton Precursor polypeptide. *Journal of Virology* **41**, 1083-1088.

Zalloua, P. A., Buzayan, J. M. & Bruening, G. (1996). Chemical cleavage of 5'-linked protein from tobacco ringspot virus genomic RNAs and characterization of the protein-RNA linkage. *Virology* **219**, 1-8.