INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.
Gene expression strategies of barley yellow dwarf virus

Dinesh-Kumar, S. P., Ph.D.

Iowa State University, 1993
Gene expression strategies of barley yellow dwarf virus

by

S.P. Dinesh-Kumar

A Dissertation Submitted to the
Graduate Faculty in Partial Fulfillment of the
Requirements for the Degree of
DOCTOR OF PHILOSOPHY

Department: Plant Pathology
Interdepartmental Major: Molecular, Cellular, and Developmental Biology

Approved:

Signature was redacted for privacy.

In Charge of Major Work

Signature was redacted for privacy.

For the Interdepartmental Major

Signature was redacted for privacy.

For the Major Department

Signature was redacted for privacy.

For the Graduate College

Iowa State University
Ames, Iowa

1993
I dedicate this dissertation to my parents B.Puttaswamy and Savithramma
### TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>GENERAL INTRODUCTION TO LUTEOVIRUS GENE EXPRESSION</td>
<td>1</td>
</tr>
<tr>
<td>Luteoviruses</td>
<td>1</td>
</tr>
<tr>
<td>Genome Organization</td>
<td>3</td>
</tr>
<tr>
<td>Gene Functions</td>
<td>7</td>
</tr>
<tr>
<td>Gene Expression Strategies</td>
<td>16</td>
</tr>
<tr>
<td>Explanation of Thesis Format</td>
<td>30</td>
</tr>
<tr>
<td><strong>PAPER I</strong> PRECISE MAPPING AND IN VITRO TRANSLATION</td>
<td>32</td>
</tr>
<tr>
<td>OF A TRIFUNCTIONAL SUBGENOMIC RNA OF BARLEY YELLOW DWARF VIRUS</td>
<td></td>
</tr>
<tr>
<td><strong>ABSTRACT</strong></td>
<td>34</td>
</tr>
<tr>
<td><strong>INTRODUCTION</strong></td>
<td>35</td>
</tr>
<tr>
<td><strong>MATERIALS AND METHODS</strong></td>
<td>37</td>
</tr>
<tr>
<td>Virus and Viral RNA Purification</td>
<td>37</td>
</tr>
<tr>
<td>Electroporation of Protoplasts with BYDV RNA</td>
<td>37</td>
</tr>
<tr>
<td>Plasmid Construction</td>
<td>38</td>
</tr>
<tr>
<td>In vitro Transcription</td>
<td>40</td>
</tr>
<tr>
<td>RNA Extraction</td>
<td>41</td>
</tr>
<tr>
<td>Northern Blot Analysis</td>
<td>42</td>
</tr>
<tr>
<td>RNase Protection</td>
<td>42</td>
</tr>
<tr>
<td>Primer Extension</td>
<td>43</td>
</tr>
<tr>
<td>In vitro Translation and Immunoprecipitation</td>
<td>43</td>
</tr>
<tr>
<td><strong>RESULTS</strong></td>
<td>45</td>
</tr>
<tr>
<td>Characterization of BYDV-PAV RNAs</td>
<td>45</td>
</tr>
<tr>
<td>Identification of the 5' end of sgRNA1</td>
<td>48</td>
</tr>
<tr>
<td>In vitro Translation Products of Synthetic Subgenomic RNA</td>
<td>50</td>
</tr>
<tr>
<td>Section</td>
<td>Page</td>
</tr>
<tr>
<td>----------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>DISCUSSION</td>
<td>56</td>
</tr>
<tr>
<td>Subgenomic RNA Mapping</td>
<td>56</td>
</tr>
<tr>
<td>Translation Initiation at Two Codons</td>
<td>58</td>
</tr>
<tr>
<td>Readthrough</td>
<td>64</td>
</tr>
<tr>
<td>ACKNOWLEDGMENTS</td>
<td>66</td>
</tr>
<tr>
<td>LITERATURE CITED</td>
<td>67</td>
</tr>
<tr>
<td>PAPER II CONTROL OF START CODON CHOICE ON A PLANT VIRAL RNA ENCODING OVERLAPPING GENES</td>
<td>74</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>76</td>
</tr>
<tr>
<td>INTRODUCTION</td>
<td>78</td>
</tr>
<tr>
<td>RESULTS</td>
<td>82</td>
</tr>
<tr>
<td>Effects of mutations near the CP start codon on secondary structure</td>
<td>82</td>
</tr>
<tr>
<td>Effects of mutations on translation of CP and 17K ORFs in reticulocyte lysates</td>
<td>89</td>
</tr>
<tr>
<td>Effects of mutations on start codon choice in protoplasts</td>
<td>93</td>
</tr>
<tr>
<td>Efficient initiation at the 17K AUG is needed for maximum initiation at the CP AUG</td>
<td>99</td>
</tr>
<tr>
<td>DISCUSSION</td>
<td>101</td>
</tr>
<tr>
<td>Differences between cell-free and protoplast assay systems</td>
<td>101</td>
</tr>
<tr>
<td>Effect of secondary structure</td>
<td>103</td>
</tr>
<tr>
<td>Roles of bases at positions -3 and +4</td>
<td>104</td>
</tr>
<tr>
<td>Initiation at ACG</td>
<td>105</td>
</tr>
<tr>
<td>Ribosome pausing model</td>
<td>105</td>
</tr>
<tr>
<td>Regulation of CP and 17K levels in infected cells</td>
<td>110</td>
</tr>
<tr>
<td>METHODS</td>
<td>112</td>
</tr>
<tr>
<td>Construction of mutant plasmids for in vitro studies</td>
<td>112</td>
</tr>
<tr>
<td>Plasmids pAGUS1 and pAdhGUS</td>
<td>112</td>
</tr>
<tr>
<td>Construction of CP/17K-GUS fusions</td>
<td>114</td>
</tr>
</tbody>
</table>
ACKNOWLEDGMENTS 117

REFERENCES

GENERAL SUMMARY 125

REFERENCES 130

ACKNOWLEDGMENTS 146

APPENDIX I IN VITRO SYNTHESIS OF FULL-LENGTH INFECTIOUS TRANSCRIPTS FROM A cDNA CLONE OF BARLEY YELLOW DWARF VIRUS 147

APPENDIX II OF 39K ORF STOP CODON ON BYDV-PAV FRAMESHIFTING AND REPLICATION 166

APPENDIX III SIGNALS THAT CONTROL TRANSLATIONAL READTHROUGH OF COAT PROTEIN STOP CODON IN BARLEY YELLOW DWARF VIRUS IN VIVO 174
GENERAL INTRODUCTION TO LUTEOVIRUS GENE EXPRESSION

Luteoviruses

Barley yellow dwarf virus (BYDV) is the type member of luteovirus group of plant viruses. The virions of members of this group are icosahedral and 24-30 nm in diameter. They are made up of 180 subunits of a 22 kDa coat protein surrounding one molecule of about 6 kb single-stranded RNA. Based on genome organization, serological relatedness and cytopathological effects, luteoviruses are classified into two subgroups (D'Arcy, 1986). Subgroup I includes: BYDV-MAV, BYDV-PAV, BYDV-SGV, soybean dwarf virus (SDV) and others. Subgroup II includes: BYDV-RPV, beet western yellow virus (BWYV), potato leaf roll virus (PLRV) and others. Luteoviruses are widespread and cause heavy economic losses of sugar beets, potatoes and small grain cereals worldwide (Rochow and Duffus, 1981; Conti et al., 1990). BYDV and PLRV infects all members of the Graminae and Solanaceae respectively. BWYV infects wide range of dicotyledons and some monocotyledons crops (Rochow and Duffus, 1981)

Luteoviruses are confined to phloem, mostly in sieve elements, companion cells, and occasionally in phloem parenchyma (Rochow, 1969). They have not been observed in epidermal, mesophyll or xylem tissues of infected plants. However, luteoviruses can replicate efficiently in the protoplasts derived from mesophyll tissues or suspension cells (Barker and Harrison, 1982; Young et al., 1989, 1991; Dinesh-Kumar et. al., 1992; Veidt et al., 1992). Different strains of BYDV induce distinct cytopathological
events (Gill and Chong, 1979). In cells infected with strains belong to subgroup I, virus particles move from sieve elements through plasmodesmata into adjacent companion cells, new virus accumulates in the cytoplasm near the plasmodesmata, single membraned vesicles containing fibrils appear in the cytoplasm, the nucleus becomes distorted, aggregation and accumulation of densely staining, heterochromatin-like material occurs. In contrast, in cells infected with members of subgroup II, vesicles containing fibrils are bounded by a second membrane which is continuous with the endoplasmic reticulum, virus accumulates in the nucleolus, eventually the heterochromatin slowly disintegrates, extensive wall thickening occurs in the infected parenchyma cells. The cytopathological changes induced by BWYV (Esau and Hoefert, 1972) and PLRV (Shepardson et al., 1980) were the same as those induced by subgroup II strains of BYDV.

Luteoviruses are transmitted only by aphid vectors in persistently-circulative manner (Rochow, 1969). Aphids after acquiring the virus, can transmit to plants throughout the life of the aphid, even after molting (Gildow, 1985). Luteovirus particles from infected phloem cells are ingested into the aphid's stylet and accumulate in the hindgut. The virus particles are transported from the haemocoel to the basal membrane of the accessory salivary gland, from which it is exuded into the salivary duct. Virions are then excreted along with salivary cell secretions into the phloem cells of healthy plants during feeding (Gildow, 1982; Gildow and Rochow, 1980). There is a strong evidence that luteoviruses do not replicate in the aphid.
vectors (Paliwal and Sinha, 1970). Luteoviruses have a high degree of specificity between the virus and the aphid species transmit them. Each luteovirus is transmitted efficiently by only one or a few species of aphid. Luteoviruses are difficult to control, because of these very specific persistant vector relationships.

**Genome Organization**

The genome of luteoviruses consist of a single positive sense RNA of 5.6-6.0 kb. Some subgroup II viruses, like PLRV and RPV, contain a 5'-terminal genome linked protein (VPg) (Mayo et al., 1982; Murphy et al., 1989). The 3' end of the genome does not contain a poly(A) tail or t-RNA-like structure. The complete nucleotide sequences of a number of luteoviruses, including BYDV-PAV (Miller et al., 1988a; Ueng et al., 1992), BYDV-RPV (Vincent et al., 1991), BYDV-MAV (Ueng et al., 1992), SDV (Rathjen et al., 1994), PLRV (Mayo et al., 1989; van der Wilk, et al., 1989; Keese et al., 1990), and BWYV (Veldt et al., 1988) are known. Based on the knowledge of nucleotide sequences, subgroup I and subgroup II luteoviruses were found to have different genome organization (Figure 1).

Subgroup II viruses possess an open reading frame (ORF 0) at the 5' terminus, which is not present in subgroup I luteos. ORF 0 encodes 28-29 kDa product (P0). The 3' sequence of members of subgroup I is longer (654-687 nt) compare to subgroup II (150 nt). In BYDV-PAV and BYDV-MAV, this results in an additional open reading frame (ORF 6), which encodes a 4.3 to 6.7 kDa product (P6). SDV has no ORFs in this region.
Luteoviruses

Subgroup I

RNA1

RNA2

Subgroup II

PEMV

Figure 1 Genotype organization of luteoviruses (Subgroup I and Subgroup II) and enamovirus (pea enation mosaic virus, PEMV). Boxes represents open reading frames (ORFs). Numbering system followed here is similar to that used by Martin, et al., (1990). Homology between ORFs of different groups are represented by similar shading. PO to P6 indicate the translation products of respective ORFs, expected sizes of the product is shown in parentheses. pol, polymerase; CP, coat protein. Translational strategies are indicated by abbreviations, Is, leaky scanning; fs, frameshifting; rt, stop codon readthrough.
ORF 0 of subgroup II overlaps extensively with the ORF1, with the latter being expressed by leaky scanning. ORFs 1 and 2 code for components of the viral RNA-dependent RNA polymerases. ORF2 is expressed by translational frameshifting (Brault and Miller, 1992; Prufer et al., 1992). ORF 1 of subgroup I codes for a 39 kDa polypeptide (P1), whereas ORF 1 in subgroup II ranges from 66-70 kDa. ORFs 1 and 2 in subgroup I overlap by 13 nt, compared to 298-474 nt in subgroup II. The amino acid sequences of subgroup I polymerases (ORF 2) are more similar to those of diantho-, carmo- and tombusviruses than to the subgroup II polymerases. The polymerases of subgroup II are more similar to those of sobemoviruses than to the subgroup I luteoviruses (Habili and Symons, 1989; Koonin, 1991). This difference in the polymerase divides luteoviruses into two separate subgroups.

In all luteoviruses, three ORFs at the 3' end of the genome are expressed from a large single subgenomic RNA (Tacke et al., 1990; Miller and Mayo, 1991; Dinesh-Kumar et al., 1992). The intergenic region between ORF 2 and ORF 3 ranges from 116 nt to 202 nt. ORF 3 encodes viral coat protein (CP) of 22 kDa. ORF 4, which encodes a 17-19 kDa product is nested within the CP coding sequence but in another reading frame. The overlapping ORF 3 and ORF 4 are expressed by the leaky scanning mechanism (Kozak, 1986; Kozak, 1989; Dinesh-Kumar and Miller, 1993). The CP ORF is followed by a 50 kDa ORF (ORF5) in the same reading frame and separated by a single amber termination codon. ORF 5 is expressed by translational readthrough of the coat protein stop codon (Veidt
et al., 1988; Bahner et al., 1990; Tacke et al., 1990; Dinesh-Kumar et al., 1992).

Pea enation mosaic mosaic virus (PEMV), a monotypic enamovirus is included in this review because it has some properties similar to luteoviruses. It is transmitted by aphids in a persistant circulative manner. Unlike luteoviruses, it is also mechanically transmissible. The virus is not limited to phloem cells, but it is mostly concentrated in mature phloem sieve elements (de Zoeten and Gaard, 1983). It has bipartite genome with two RNAs, which are encapsidated in separate, morphologically distinct particles (German and de Zoeten, 1975). The 5' end of these RNAs contains a 17.5 kDa genome-linked viral protein (Reisman and de Zoeten, 1982). Like luteoviruses, they also lack a poly(A) tail or t-RNA like structure at the 3' end of the genome (German et al., 1978). Both of these RNAs are required for systemic infection, but they can replicate individually in the protoplasts (Demler et al., 1993). RNA1 has a similar genome organization as luteovirus subgroup II, except for the absence of ORF 4, which is nested with the CP ORF (ORF 3) (Figure 1). The ORF 5 encodes 33 kDa product compared to the approx. 50 kDa product in luteoviruses. The 5' half of the RNA2 is similar to subgroup I type of luteoviruses. The polymerase (ORF 2) sequence of RNA2 is more similar to luteo-, carmo-, diantho- and tombus virus groups than to RNA1. Thus, PEMV contains two distinct type polymerase-like proteins. The function of the downstream ORFs in RNA2 is unknown.
Gene Functions

1. Polymerase/helicase

Amino acid sequence analyses of ORFs 1 and 2 indicate that these two ORFs code for helicase-like and polymerase-like RNA replication complex (Habili and Symons, 1989). The ORF 2 of both subgroups appears to encode core polymerase, because it contains a highly conserved amino acid sequence; GXXXTXXXN\(\times\)\(_{20-40}\)GDD, a signature sequence of all RNA-dependent RNA polymerases (Habili and Symons, 1989; Kamer and Argos, 1984). ORFs 1 and 2 contain amino acid motifs similar to viral and cellular helicases (Habili and Symons, 1989). Nucleic acid helicases (or helicase subunits) are involved in unwinding of double stranded templates during replication and recombination (Gorbalenya et al., 1988). Helicase motifs I, Ia and III are present in ORF 1 of subgroup I, whereas motif IV and VI are present in ORF 2 (Habili and Symons, 1989). In subgroup II motif IV is present in ORF 1 and motif VI in ORF 2. Motif VI present in all known helicases provides NTP-binding domain (Gorbalenya, et al., 1988).

In BWYV, frameshift mutations in ORF 1 or ORF 2 were lethal for replication in protoplasts (Reutenauer et al., 1993). Further, the deletion mutants with only ORF 1 and ORF 2 were capable of replication in protoplasts. This indicated that polypeptide encoded by these two ORFs function as viral replicase complex. In contrast, Young et al., (1991) reported that the ORF1 is dispensable for infection in case of BYDV-PAV (subgroup I). However, our results show that ORF1 product is essential for
infection (Dinesh-Kumar, and Miller, unpublished; see Appendix I of this thesis).

2. Coat protein

The primary function of the coat protein is as a structural unit of virions. Several findings show that the ORF3 of both subgroups codes for the 22 kDa polypeptide, and that functions as coat protein. In case of BYDV-PAV the direct amino acid sequence obtained from the tryptic peptides derived from purified viral coat proteins were identical to those deduced from the nucleic acid sequencing (Miller et al., 1988b). Similarly the N-terminal amino acid sequence analysis of the purified PEMV virions were identical to the N terminal nucleotide sequence of the ORF3 of PEMV (Demler and de Zoeten, 1991). In BWYV, the antiserum raised against the CP expressed as a fusion protein in bacteria was reactive with coat protein accumulated in BWYV infected protoplasts (Veidt et al., 1992). All these observations indicate that the ORF3 encodes the 22 kDa coat protein.

3. Genome linked viral protein (VPg)

Instead of the commonly found cap structure (m7GpppG) at the 5' end, the RNA of some viruses contains a viral-derived protein called genome-linked viral protein. (VPg). VPg has been suggested to have a role in viral replication (Wimmer, 1982). PLRV (Mayo et al., 1982) and RPV (Murphy et al., 1989) contain VPg's of 7 kDa and 17 kDa respectively. The VPg protein identified for BYDV-RPV corresponds to the size of ORF 4
product. Hence ORF 4 which is nested with in the CP ORF has been proposed to encode VPg in luteoviruses (Miller et al., 1988a). In the case of PLRV, it has been suggested that the product of ORF 4 may be a precursor protein to VPg, which upon proteolytic cleavage may yield 7 kDa protein, as in case of cowpea mosaic virus (van der Wijk et al., 1989). The 17 kDa protein product of ORF 4 in PLRV has nucleic acid binding properties, hence (Tacke et al., 1991) predicted that it may function as VPg and may play a role in viral replication. However, point mutations introduced in the start codon of ORF 4 of BWYV or BYDV-PAV did not affected the replication in protoplasts, indicating that the ORF 4 may not encode VPg (Reutenauer et al., 1993; Dinesh-Kumar and Miller, unpublished). Furthermore, the antibody raised against the ORF4 fusion protein, fail to detect any low molecular weight product of about 7 kDa from BWYV infected protoplast extracts (Reutenauer et al., 1993).

In the earlier observation Reisman and de Zoeten, 1982 showed that both RNAs of PEMV contain VPg of 17.5 kDa at the 5'end of the RNA. The recent sequence data indicate that the ORF4 that encodes this size protein is absent in PEMV genome (Demler et al., 1993). Hence, authors are doubtful about the presence of VPg on both species of PEMV RNA. These observations indicate that the coding capacity elsewhere in the genome may encode VPg in case of PEMV and luteoviruses (see below).
4. Protease

The N-terminal region of ORF1 of PEMV RNA1 and subgroup II luteoviruses possesses hydrophobic residues Demler and de Zoeten, (1991). Hence authors suggested that the ORF 1 and ORF 2 products together can function in replication as membrane-bound replicase system similar to that of picornoviruses (Takeda et al., 1986) and comoviruses (Goldbach and Van Kammen, 1985). The membrane-bound replicase subunit-VPg-protease-polymerase complex carries out VPg-primed replication and leaves behind the VPg after proteolytic cleavage from the protease.

The chymotrypsin-like serine proteases of picarno- and como- viruses contain three conserved catalytic sites called a catalytic triad (Gorbalenya et al., 1989). Demler and de Zoeten, (1991) found one catalytic site motif, TR/KXGXSG in P1 of PEMV RNA1. We analyzed P1 sequences of subgroup II luteoviruses and found that all of them contain the conserved triad catalytic motifs; H(X_{25})E/D(X_{70-80})TK/RXGXSG (Figure 2). These findings indicate that the ORF1 product may contain a protease function. However, no proteolytic cleavage products have been detected so far in any subgroup II viral translation products. Therefore careful analysis of the translation products using antibody raised against different regions of the P1/P2 of subgroup II viruses may provide evidence for existence of such a replication complex. Analysis of the P1/P2 sequence of subgroup I did not reveal any protease-like motifs. Therefore one can hypothesize that the membrane-bound VPg based replication occurs only in subgroup II viruses but not in subgroup I. Several findings indirectly support this hypothesis.
<table>
<thead>
<tr>
<th></th>
<th>BWYV</th>
<th>PLRV</th>
<th>RPV</th>
<th>PEMV</th>
<th>Cons</th>
</tr>
</thead>
</table>
| 1 | TLTQKSVGFAKFLW KDLRTWSSLL LASFSAIWL VSNITTP... | VFAKFLSVFAKLW KDLRTWSSLL LASFSAIWL VSNITTP... | VFAKFLSVFAKLW KDLRTWSSLL LASFSAIWL VSNITTP... | VFAKFLSVFAKLW KDLRTWSSLL LASFSAIWL VSNITTP... | VFAKFLSVFAKLW KDLRTWSSLL LASFSAIWL VSNITTP...
| 10 | **QIISIAFSLL** KKSFSAL... RSPFKLYEK AIDGPKFTIF | **QIISIAFSLL** KKSFSAL... RSPFKLYEK AIDGPKFTIF | **QIISIAFSLL** KKSFSAL... RSPFKLYEK AIDGPKFTIF | **QIISIAFSLL** KKSFSAL... RSPFKLYEK AIDGPKFTIF | **QIISIAFSLL** KKSFSAL...
| 20 | **FLV...**MPL KIIRV...FSDKNNKFLVE AVEGNYETF | **FLV...**MPL KIIRV...FSDKNNKFLVE AVEGNYETF | **FLV...**MPL KIIRV...FSDKNNKFLVE AVEGNYETF | **FLV...**MPL KIIRV...FSDKNNKFLVE AVEGNYETF | **FLV...**MPL KIIRV...FSDKNNKFLVE AVEGNYETF
| 30 | **LL...**CITVLKKNFIRL...TFKFRFNEK TV3GSDYISI | **LL...**CITVLKKNFIRL...TFKFRFNEK TV3GSDYISI | **LL...**CITVLKKNFIRL...TFKFRFNEK TV3GSDYISI | **LL...**CITVLKKNFIRL...TFKFRFNEK TV3GSDYISI | **LL...**CITVLKKNFIRL...TFKFRFNEK TV3GSDYISI
| 40 | **EPEAKIATAK SDLGVTLJL** | **EPEAKIATAK SDLGVTLJL** | **EPEAKIATAK SDLGVTLJL** | **EPEAKIATAK SDLGVTLJL** | **EPEAKIATAK SDLGVTLJL**
| 50 | **LQVMPWSHL LTVSPLA...** | **LQVMPWSHL LTVSPLA...** | **LQVMPWSHL LTVSPLA...** | **LQVMPWSHL LTVSPLA...** | **LQVMPWSHL LTVSPLA...**
| 60 | **-----LW-----L-----** | **-----LW-----L-----** | **-----LW-----L-----** | **-----LW-----L-----** | **-----LW-----L-----**

Figure 2 Alignment of indicated portions of P1 protein of subgroup II luteoviruses (BWYV, PLRV, and RPV) and pea enation mosaic virus RNA1 with putative chymotrypsin-like serine protease of southern bean mosaic virus (SBMV). The derived consensus (Cons) is shown below the aligned sequence. Consensus is defined if similar residue occurs in 4 out of 5 sequences. * indicates residues which form conserved catalytic triads of H, Histidine; D, Aspartate; and S, serine. References for sequences included in the alignment are BWYV (Veidt et al., 1988), PLRV (Keese et al., 1990), RPV (Vincent et al., 1991), and SBMV (Wu et al., 1987).
The ultrastructure observation indicates that subgroup II luteoviruses and PEMV are localized in the continuous membrane systems of the nuclear envelop and endoplasmic reticulum (Esau and Hoefert, 1972; Powell and de Zoeten, 1977; Shepardson et al., 1980). In BWYV (subgroup II), capping of the infectious transcripts results in 20 fold higher infection compare to uncapped transcripts (Veldt et al., 1992). Whereas in BYDV-PAV (subgroup I) capped transcripts are not infectious (Dinesh-Kumar and Miller, unpublished; see Appendix I of this thesis) and interestingly the BYDV-PAV genome translates in a cap-independent manner in wheat germ translation system (Wang and Miller, unpublished).

5. Aphid transmission and encapsidation

The ORF 5 product has been suggested to play a role in specificity of aphid transmission and particle assembly (Massalski and Harrison 1987; Bahner et al., 1990). The role of readthrough protein in aphid transmission comes from the observation that a possible product of ORF 5 was eliminated from the genome of a aphid transmissible strain of PEMV upon many generations of mechanical inoculation (Hull, 1977; Adam et al., 1979). However, the antibody made against the P5 protein of PLRV aphid transmissible strain detected the same size fusion product of P3 and P5 from the potato tissue infected with aphid transmissible and non-transmissible strains (Bahner et al., 1990). Furthermore, Demler and de Zoeten, 1991 found ORF 5 (P5, 33 kDa) in RNA1 of both aphid non-transmissible and aphid-transmissible strains of PEMV, which is also separated by amber stop
codon from coat protein (P3). These observations argue against the complete loss of P5 in non-aphid transmissible strains.

The N-terminal sequence of P5 display great similarity (Mayo et al., 1989; Bahner et al., 1990; Demler and de Zoeten, 1991) but the C-terminal halves are very distinct (Mayo et al., 1989). The C-terminal part of P5 is proteolytically removed in case of PLRV (Bahner et al., 1990). Because of the lack of C-terminal region of P5 in case of PEMV and proteolytic cleavage of this region in case of other luteoviruses may indicate that the N-terminal part of the P5 may contain a domain responsible for aphid transmission (Bahner et al., 1990; Demler and de Zoeten, 1991).

All luteoviruses and PEMV contain amino acid sequence rich in proline residues immediately downstream of the coat protein (see review by Miller et. al., 1994). These proline residues have been suggested to play a role in separating the coat protein domain from the readthrough domain, which may facilitate it to protrude out of the virus particle (Bahner et al., 1990). This protruded readthrough domain may determine aphid specificity by interacting with the receptors in the aphid vector (Bahner et al., 1990). Such protrusions have been observed on the surface of PLRV and BYDV particles (Harrison 1984; Waterhouse et al., 1989).

The readthrough of the CP gene and to generate a downstream ORF has been shown to occur in beet necrotic yellow vein virus (BNYVV) (Bouzoubaa et al., 1986) and soil-borne wheat mosaic virus (Hsu and Brakke, 1985). In BNYVV, readthrough protein is required for efficient transmission by the fungal vector (Tamada and Kusume, 1991) and for particle assembly
(Schmitt, et al., 1992). The deletion mutants in the readthrough protein of BWYV (Reutenauer et al., 1993) and BYDV-PAV (Dinesh-Kumar and Miller, unpublished; see Appendix I of this thesis) were infectious in protoplasts and in the former case they were able to form virus particles. This indicates that coat protein is sufficient for the luteoviruses particle assembly. These results are in contrast to the observation from Young et al., (1991) who claimed that frameshift mutations in the BYDV-PAV readthrough protein were lethal to the virus replication.

6. Movement

Several recent findings indicate that the 17 kDa protein encoded by ORF 4 may function as a phloem specific cell-to-cell movement protein of luteoviruses. The C-terminal domain of the 17 kDa protein of PLRV binds single stranded nucleic acids in a co-operative and sequence non-specific manner (Tacke et al., 1991), similar to that of the well characterized P30 movement protein of tobacco mosaic virus (TMV) (Citovsky et al., 1990). The binding of P30 protein to TMV RNA is proposed to unfold the RNA to facilitate movement through plasmodesmata channels and also to protect it from cellular nucleases (Citovsky et al., 1990). The 17 kDa protein was most abundant in fractions enriched for nuclei, mitochondria, chloroplast and membranous structures but not in cell wall fractions (Tacke et al., 1993). In contrast, TMV movement protein accumulates mainly in the cell wall fraction (Deom et al., 1990). Homodimers of 17 kDa protein were observed in cell extracts of non-transformed potato plants infected with PLRV and transgenic
plants expressing P4. The domain responsible for this dimerization is located in the N-terminal half of the 17 kDa protein (Tacke et al., 1993). The 17 kDa protein expressed in transgenic plants was a phosphoprotein (Tacke, et al., 1993). This is similar to TMV movement protein, which was shown to be phosphorylated by serine/threonine specific protein kinase from cell wall fraction (Citovsky et al., 1993). The phosphorylation of the P30 movement protein of TMV has been suggested to represent a mechanism for the host plant to sequester this protein in the cell wall of mature tissue (Citovsky et al., 1993). Based on all these biochemical properties of 17 kDa protein, it is possible that it is involved in cell-to-cell movement of viral RNA as a ribonucleoprotein complex within the phloem tissue.

7. ORF 6

The ORF 6, which encodes 4.3-6.7 kDa protein is present in BYDV-PAV and BYDV-MAV (subgroup I). The frameshift mutation introduced in this ORF showed that this product is required for infection (Young et al., 1991). The subgenomic message for this protein is highly abundant in early part of infection (Young et al., 1991; Dinesh-Kumar et al., 1992). Protein product has not been identified in infected cells. The function of this protein if it exists is not known.

8. ORF 0

ORF 0 present in subgroup II luteoviruses and PEMV RNA1 encodes protein of 28-29 kDa. The function of this open reading frame is not known.
However, it has been suggested that this protein may determine virus host range (Veidt et al., 1992). There is no sequence homology in this ORF between different subgroup II viruses (Demier and de Zoeten, 1991). Sequence in this ORF has no similarity to helicase, polymerase, protease and transport related sequences. This ORF was dispensable for replication of BWYV in protoplasts (Veidt et al., 1992).

Gene Expression Strategies

Luteoviruses use a variety of unusual gene expression strategies to express their genes (Figure 1) from minimum genetic material. The ORF 2 (polymerase gene) is expressed by translational frameshifting. The coat protein, ORF 4 and ORF 5 are translated from a subgenomic RNA. The overlapping reading frames; CP and ORF 4, and ORF 0 and ORF 1 of subgroup II, and ORF 3 and ORF 4 of PEMV RNA2 are expressed by leaky scanning. ORF 5 is translated by the in-frame readthrough of the amber stop codon of the coat protein.

1. Ribosomal frameshifting

Expression of a single protein from two or more overlapping reading frames due to translational frameshifting mechanism has been observed in many viruses and retrotransposons (Reviewed in Atkins et al., 1990; Hatfield et al., 1992). According to frameshifting, a portion of ribosomes before terminating translation of an upstream open reading frame, either move backward or forward one nucleotide and resume translation in the
overlapping reading frame. Many animal viruses and retrotransposons (Reviewed in (Atkins et al., 1990; Hatfield et al., 1992) and some plant virus including luteoviruses (see review by Miller et al., 1994) induces a -1 translational frameshift events for synthesis of RNA-dependent RNA polymerase or reverse transcriptase. This may be a regulatory mechanism by which the viruses make the required low amount of polymerase during early stages of infection.

Several signals present in the mRNA sequence have been found to play a role in the frameshift event. First, signal constitutes heptanucleotide sequence or slippery sequence at which actual frameshift occurs. These sequences include the run of three adenine, uracil or guanine residues followed by either AAAC, UUUU, UUUU (Jacks et al., 1988) or AAAU (Prufer et al., 1992). This shifty sequence facilitates -1 frameshift event by simultaneous slippage of aminoacyl-tRNA at the ribosomal A site and peptidyl-tRNA at the P site by one nucleotide in the 5' direction (Jacks et al., 1988). Because of this slippage ribosomes resume translation in the new reading frame. The second signal is the presence of the sequence downstream of shifty heptanucleotide which can either form pseudoknot (ten Dam, et al., 1990; Chamorro, et al., 1992; Dinman and Wickner, 1992; Morikawa and Bishop, 1992; Garcia, et al., 1993; Kujawa, et al., 1993) or a stable stem loop structure (Prufer, et al., 1992; Parkin et al., 1992). Presence of such a secondary structure or pseudoknot makes ribosomes to pause at the shifty site, which allows more time for slippage (Tu et al., 1992). However pausing alone was not sufficient for efficient frameshifting.
The presence of these signal sequences in and around the ORF 1 and ORF 2 overlapping region in luteoviruses and PEMV suggested that the polymerase gene is expressed via a -1 frameshift event. In case of BYDV-PAV and PLRV, -1 frameshift occurs in protoplasts at a frequency of about 1% (Brault and Miller, 1992; Prufer et al., 1992). The frequency of frameshifting in other cases varies from 1 to 50%. The virion RNA or in vitro synthesized full length infectious transcripts of BYDV-PAV and BWYV, programmed with wheat germ extract gave expected low amount of transforme protein and large quantity of ORF 1 product (Young et al., 1991; Veldt et al., 1992; Di et al., 1993). The sequence following the shifty sequence in case of BYDV-PAV, PLRV and BWYV can adopt two alternative structures; a simple stem loop or a pseudoknot structure. In case of PLRV, the stem-loop structure was necessary for efficient frameshifting in reticulocyte lysate (Prufer et al., 1992). In contrast the recent results on BWYV (Garcia et al., 1993) and PLRV (Kujawa et al., 1993) indicate the requirement of a pseudoknot. However the role of shifty heptanucleotide sequence and any of the structures has yet to be demonstrated for luteoviruses in vivo.

2. Subgenomic RNAs

Genomic RNAs of most viruses are polycistronic and expression of internal cistrons is facilitated by its own subgenomic RNAs. These are 5'-truncated copies of the genomic RNA, and are encapsidated in some viruses, but not in others. This strategy allows the differential expression of two or
more genes independently from a single genomic RNA. Several in vitro (Miller et al., 1985; Marsh et al., 1988) and in vivo (Gargouri et al., 1989) studies clearly suggest that the subgenomic RNAs arise by internal initiation on the subgenomic promoter on the (-)-strand of genomic RNA.

Two subgenomic RNAs of about 3 kb and 0.8 kb have been identified in BYDV-PAV infected tissue or protoplasts (Young, et al., 1991; Dinesh-Kumar et al., 1992). PLRV and BWYV possess only one large subgenomic RNA of about 2.5 kb (Smith and Harris, 1990; Tacke et al., 1990; Miller and Mayo, 1991; Veldt et al., 1992). The non-coding region between ORF 2 and ORF 3 (CP) in all luteoviruses is A-U-rich, which is a characteristic feature of many subgenomic leaders. Further, (Vincent et al., 1991) found two core regions in the leader sequences of luteoviruses similar to that of well-characterized BMV subgenomic promoter (Marsh et al., 1988).

The 5' ends of the large subgenomic RNA have been precisely mapped for some luteoviruses. In a German isolate of PLRV the subgenomic start is at 40 nucleotide upstream of CP start codon (Tacke et al., 1990). In contrast, the start site of the PLRV-Scottish isolate was mapped to 212 nucleotide upstream of the CP AUG (Miller and Mayo, 1991). These isolates differ at only five bases in the non-coding region. The sequence ACAAAA at the 5' end of the subgenomic RNA in case of PLRV-Scottish isolate is identical to the sequence at the 5' end of the genomic RNA, suggesting that it is a replicase recognition signal on the (-) strand for the synthesis of subgenomic RNA (Miller and Mayo, 1991). In case of BYDV-PAV (Illinois isolate), the start of large subgenomic RNA is 89 nucleotides upstream of CP
start codon (Dinesh-Kumar et al., 1992). The sequence ACAAAA is present about 34 nucleotides upstream of the actual start site and the same sequence is present at position 19 in the genomic RNA. Similarly, in maize chlorotic mottle virus an 11 nucleotide tract similar to the 5' end of genomic RNA was found 111 nucleotides upstream of the subgenomic start site (Lommel et al., 1991). Similarity in the sequences at the 5' end of genomic and subgenomic RNAs applies to some viruses, but it is not a common phenomenon. Therefore it is possible that the start site of the subgenomic RNA is different from the polymerase recognition site.

3. Leaky scanning

Most eukaryotic mRNAs are monocistronic and protein synthesis starts at the first AUG codon (Kozak, 1989). Usually the 40S ribosomal subunit complex including several initiation factors binds at the 5' end of the mRNA and scans until it reaches the first AUG, at which the initiator tRNA\textsuperscript{met} and 60S subunit binds to form 80S ribosome and protein synthesis begins (Kozak, 1989; Merrick, 1992). Several mRNA features such as sequences flanking the start codon, secondary structure upstream and downstream of the start codon and mRNA leader length influence the process of initiation (Kozak, 1991b). In many eukaryotes, including plants, the most favorable sequence context for initiation is usually A at -3 and G at +4 position (Kozak, 1986; Lutcke et al., 1987; Cavener and Ray, 1991). If the first AUG on an mRNA is in unfavorable sequence context, then a portion of the 40s ribosomes may bypass that AUG and initiate at the next downstream AUG if
latter is in a favorable context. Such a process is called leaky scanning (Kozak, 1989). Some viral and cellular mRNAs which encode two different proteins from overlapping reading frames have been speculated to use this process during translation (Kozak, 1991a). However, true bicistronic messages are more commonly found in viruses, because this provides a means of minimizing their genome size and allowing more efficient utilization of the coding region. In addition, this also facilitates coordinated expression of two cistrons to produce required molar amounts of two functional products at the level of translation.

**ORF 0 and ORF 1**

In subgroup II luteoviruses and PEMV RNA 1, the ORF 0 at the 5' end of the genomic RNA overlaps with the 5' terminal half of the ORF1. The start codons of these two ORFs are about 130 to 148 nucleotides apart in subgroup II luteoviruses and about 79 nucleotides in RNA1 of PEMV. Similarly tymoviruses (TYMV) and MCMV also contain overlapping reading frames at the 5' end of the genome and the start codons are separated by 4 and 16 nucleotides respectively. In all these cases, the start codon of the first ORF is flanked by suboptimal bases compared to that of second ORF start site (Figure. 3A). Hence it is possible that some of the ribosomes bypass the start codon of the first ORF and initiate at the downstream ORF start codon. In fact expected products from both ORFs were observed upon translation of the genomic RNA of PLRV (Mayo et al., 1989), BWYV (Veidt et al., 1992) and TYMV (Weiland and Dreher, 1989). In *in vitro* studies with
Figure. 3 Nucleotide sequence flanking the dual initiation sites from overlapping reading frames of members of plant viruses. A, overlapping reading frames at the 5' end of the genome; B, overlapping reading frames at the internal region of the genome. Nucleotide sequence [A/G]CNAUGGC in monocots and AN[A/C]AUGGC, in dicots includes the most frequently used bases at positions -3 through +2 (Cavener and Ray, 1991). A base is defined as sole consensus, if it occurs at specific position in >50% of the known start codon sequences. If two nucleotides appear at same position (in brackets) and sum of their frequencies is >75%, then both are considered as consensus. Bases that match the consensus are outlined, and flanking bases which appear at specific position <10% of known start codons are shown in lowercase letters. BYDV-PAV(Vic) (Victorian isolate; Miller et al., 1988a), BYDV-PAV(P) (Purdue isolate; Ueng et al., 1992), BYDV-MAV (Ueng et al., 1992), and SDV (soybean dwarf virus; Rathjen et al., 1994) are subgroup I luteoviruses. BYDV-RMV (Domier et al., 1992), BYDV-RPV (Vincent et al., 1991), PLRV-S (potato leaf roll virus, Scottish isolate; Mayo et al., 1989), PLRV-A & C (potato leaf roll virus, Australian and Canadian isolates; Keese et al., 1990), and BWYV (beet western yellow virus; Veidt et al., 1988) are subgroup II luteoviruses. PEMV-RNA1 & -RNA2 (pea enation mosaic virus; Demler and de Zoeten, 1991; Demler et al., 1993) is a monotypic enamovirus. TYMV (turnip yellow mosaic virus; Weiland and Dreher, 1989), EMV (eggplant mosaic virus; Osirio-Keese et al., 1989), KYMV (kennedy yellow mosaic virus; Ding et al., 1990), and OYMV (ononis yellow mosaic virus; Ding et al., 1989), are tymoviruses. CNV (cucumber necrosis virus; Rochon and Johnston, 1991), TBSV (tomato bushy stunt virus; Hillman et al., 1989), and CyRSV (cymbidium ring spot virus; Grieco et al., 1989) are tombus viruses. MCMV (maize chlorotic mosaic virus; Nutter et al., 1989).
TYMV, mutations in the start codon of the 69K ORF had no effect on the translation of 206K ORF or vice versa (Welland and Dreher, 1989). This shows that the both start codons were used during translation in separate initiation events. Similarly elimination of ORF 0 start codon of BWYV did not affect the expression of ORF 1 \textit{in vitro} (Veidt et al., 1992). More detailed analysis of factors including: flanking bases, mRNA features between the two start codons (especially in case of subgroup II luteoviruses and PEMV RNA1) and the leader sequence between the 5' end and the first start codon, are needed to understand translational efficiencies of these overlapping reading frames.

**ORF 3 and ORF 4:**

In luteoviruses, the 17-19K ORF (ORF4) is embedded within the CP ORF (ORF3) in the internal region of the genome. Similarly in tombusviruses, the 21K and 20K overlapping reading frames exist in an internal region of the genome. These overlapping reading frames are expressed via a single subgenomic RNA message (Tacke et al., 1990; Vincent et al., 1990; Rochon and Johnston, 1991; Dinesh-Kumar et al., 1992). The distance between the two start codons of luteoviruses ranges from 13 to 43 nucleotides and in tombus- viruses it is 29 nucleotides. In all these cases the start codon of the second ORF is flanked by more optimal bases compared to start site of the first ORF, with the exception of SDV (Figure. 3B). In addition the start codon of the BYDV-PAV CP ORF is sequestered in a stem-loop structure (Dinesh-Kumar and Miller, 1993). This led us to propose that these RNAs
follow leaky scanning mechanism of translation initiation (Miller et al., 1988a; Dinesh-Kumar et al., 1992). In BYDV-PAV using systematic site-directed mutagenesis we demonstrated leaky scanning (see paper II of this thesis).

4. Stop codon readthrough

Many viruses, including luteoviruses, use readthrough suppression to regulate the expression of two ORFs which are separated by single termination codon. This allows synthesis of required low amounts of some essential viral protein products. The UGA stop codon supression in Qβ phage is the first known example of readthrough (Weiner and Wever, 1971). Whereas, UAG termination in a plant virus, tobacco mosaic virus (TMV) is the first readthrough observed in eukaryotes (Pelham, 1978). Later this phenomenon was observed in many eukaryotic viruses including retroviruses (reviewed in Hatfield et al., 1992), alphaviruses (Strauss et al., 1983), and a number of plant viruses: tobamo-, furo-, tymo-, tobra-, carmo-, luteoviruses and MCMV (Skuzeski et al., 1990; and therein). In addition to viruses, some cellular genes like β-globin (Geller and Rich, 1980), glutathione peroxidase (Chambers et al., 1986), and maize zein (Wandelt and Feix, 1989) also use in-frame stop codon readthrough.

ORF5 in luteoviruses and PEMV is in-frame with the CP ORF (ORF3) from which it is separated by single amber termination codon. The comparison of sequence surrounding the stop codon of BYDV-PAV luteovirus with the other viruses which are known to undergo readthrough led to the
speculation that ORF5 is most likely to be expressed by translational suppression of the CP stop codon (Miller et al., 1988a). That is most of the time the CP polypeptide releases from the ribosomes at the termination codon but a portion of the ribosomes suppress this stop codon and continue translation to produce CP-ORF5 fusion products. In plant viruses, the readthrough regions can be classified into three different groups based on the sequence surrounding the termination codon (Skuzeski et al., 1990; Figure 4). The sequence context around the stop codon among luteoviruses is highly conserved (AAAUAAGGUAGAC) (Dinesh-Kumar et al., 1992) and belongs to a group other than TMV, a well studied readthrough in plants. In luteoviruses and PEMV, readthrough results in extension of CP, but in other viruses with the same class of readthrough signal it results in the synthesis of replicase. In PEMV the stop codon is UGA and also the sequence surrounding this is poorly conserved compared to other viruses of this class.

Failure to detect a primary translation product of ORF 5 or the mRNA message which encodes this size polypeptide in any luteovirus-infected cells provided indirect evidence that the ORF 5 is expressed only from suppression of CP stop codon. The first direct evidence for readthrough in luteoviruses comes from an in vitro experiment in which the synthetic BWYV subgenomic RNA containing the entire CP ORF and part of the RT region were capable of synthesizing the CP and CP-RT fusion products in rabbit reticulocyte translation system (Veidt et al., 1988). In later studies synthesis of CP and CP-ORF 5 fusion protein (72 kDa) from single message was
Figure. 4 Three groups of readthrough sites in members of plant viruses (updated from Skuzeski et al., 1990). Boxed region includes conserved nucleotide sequence around and downstream of readthrough site. CP, coat protein gene stop codon is involved in readthrough; pol, stop codon in the polymerase gene is involved.

References included here are the ones which are not in Skuzeski et al. (1990): PEBV (pea early browning virus, MacFarlane et al., 1989) a tobravirus. BYDV-PAV(P) (Purdue isolate; Ueng et al., 1992), BYDV-MAV (Ueng et al., 1992), and SDV (soybean dwarf virus; Rathjen et al., 1994) are subgroup I luteoviruses. BYDV-RMV (Domier et al., 1992), BYDV-RPV (Vincent et al., 1991), PLRV-S (potato leaf roll virus, Scottish isolate; Mayo et al., 1989), PLRV-A & C (potato leaf roll virus, Australian and Canadian isolates; Keese et al., 1990) are subgroup II luteoviruses. ST9 (Chin et al., 1993) beet western yellow virus associated satellite RNA. PEMV (pea enation mosaic virus; Demier and de Zoeten, 1991) is a monotypic enamovirus. CARN1 and CARN2 (first and second stop codons in the polymerase gene of carnation mottle virus; Guilley et al., 1985), CCFV (cardamine chlorotic fleck virus; Skotnicki et al., 1993), and MNSV (melon necrotic spot virus; Riviere and Rochon, 1990) are carmo-viruses. TCV turnip crinkle virus; Carrington et al., 1989).
demonstrated using BYDV-PAV full-length synthetic subgenomic RNA containing perfect 3' and 5' termini in rabbit reticulocyte lysates (Dinesh-Kumar et al., 1992). This *in vitro* synthesized readthrough product reacted with the BYDV-specific antibody indicating that this protein is present in the purified virus particles. The CP-ORF 5 fusion product was identified in PLRV (Bahner et al., 1990) and BWYV (Reutenauer et al., 1993) infected protoplasts and plant tissues using the antibody raised against the CP-ORF5 fusion protein. Despite identification of readthrough products *in vivo* the actual molar ratio of CP to readthrough product has not been estimated for any luteoviruses. However, in BYDV-PAV efficiency of readthrough *in vitro* varied from 7-15% depending on the salt concentration (magnesium chloride and/or potassium acetate) in the translation mix (Dinesh-Kumar et al., 1992). Similarly in TMV it was possible to enhance the suppression of stop codon from 10% upto 40% *in vitro* just by altering the magnesium ion concentration (Pelham, 1978).

In PLRV to determine *in vivo* readthrough efficiency, the region including 18 nucleotides upstream and 21 nucleotides downstream of the CP stop codon was fused to the β-glucuronidase (GUS) reporter gene under the control of 35S promoter (Tacke et al., 1990). When this construct was electroporated into tobacco and potato protoplasts, the CP stop codon was suppressed at about 1% efficiency. The observed *in vivo* readthrough efficiency is 7-10 fold lower compared to *in vitro* translation experiments with BYDV-PAV full length synthetic subgenomic RNA (Dinesh-Kumar et al., 1992). With TMV using the similar approach, the efficiency of readthrough
observed in vivo (5%, Skuzeski et al., 1990) was 2-4 fold lower than in vitro (Pelham, 1978). However, in the same experiment (Skuzeski et al., 1990), failed to observe readthrough of MCMV and CarMV stop codons which belong to the luteovirus group of readthrough signals. We were unable to observe readthrough of BYDV-PAV stop using the same approach (Dinesh-Kumar and Miller, unpublished; see Appendix III of this thesis). Failure to observe readthrough in our experiments is not due to the inefficient expression level in oat protoplast system because our positive control values were comparable with PLRV and TMV experiments. Further, use of different types of constructs even one including entire CP ORF and the part of readthrough ORF did not give any readthrough. However, we were able to detect readthrough using a construct in which GUS gene was inserted downstream of the CP stop codon in a full-length infectious clone of BYDV-PAV (see Appendix III of this thesis).

Explanation of Thesis Format

This thesis includes general introduction to luteovirus gene expression, two published papers and three appendixes. The paper I was published in 1992, April issue of Virology and paper II in 1993, June issue of The Plant cell. The Figures 1 and 3 of appendix I is my contribution to the paper published in 1993 August issue of Molecular Plant Microbe Interactions by Di, et. al. Appendix II and III include results that will be published after some additional work. Any references cited in the general introduction to
luteovirus gene expression, appendixes and general summary are included in the reference section that follows the general summary.
PAPER I  PRECISE MAPPING AND IN VITRO TRANSLATION OF A TRIFUNCTIONAL SUBGENOMIC RNA OF BARLEY YELLOW DWARF VIRUS
Precise mapping and *in vitro* translation of a trifunctional subgenomic RNA of barley yellow dwarf virus

Dinesh-Kumar, S. P., M.S.
Brault, V., Ph. D.
Miller, W. A., Ph. D.

Plant Pathology Department and
Molecular, Cellular and Developmental Biology Program
Iowa State University
Ames, Iowa 50011
ABSTRACT

The barley yellow dwarf virus genome consists of a single 5.7kb RNA encoding six open reading frames (ORFs). The four ORFs in the 3' half of the genome were proposed to be expressed via subgenomic mRNAs. Here, we show that the PAV serotype of barley yellow dwarf virus (BYDV-PAV) generates two subgenomic RNAs of 2.9 and 0.8kb in infected plants and protoplasts. The 5' end of the larger subgenomic RNA (sgRNA1) was precisely mapped to base 2769 by ribonuclease protection and primer extension methods. Synthetic sgRNA1, containing the exact 5' and 3' ends, was generated by in vitro transcription, translated in vitro, and shown to serve as a message for three ORFs. Translation initiated at the first two AUG codons in sgRNA1, yielding the 22 kilodalton (kDa) viral coat protein (CP) and a 17 kDa protein encoded by an overlapping, out-of-frame ORF. In addition, readthrough of the amber stop codon of the CP ORF gave rise to a 72 kDa fusion product of the CP ORF and an in-frame 50K ORF immediately 3' of the CP termination codon. The 0.8kb sgRNA2 was present in greater abundance than sgRNA1 and likely serves as a message for a 6.7K ORF at the 3' end of the genome. Sequences that may control subgenomic RNA synthesis and the translational events are compared with those of other plant viruses.
INTRODUCTION

Many plant viruses express genes via subgenomic RNAs. These RNAs are 5'-truncated forms of the genomic RNA. By bringing the 5' end in proximity with the start codons of genes located in the 3' end of the genome, subgenomic RNAs act as efficient messages for these genes. The 5' ends of subgenomic RNAs have been precisely mapped for members of the bromo- (French and Ahlquist, 1988), tymo- (Gargouri et al., 1989), tobamo- (Goelet et al., 1982), tobra- (Goulden et al., 1990), cucumo- (Gould and Symons, 1982), tombus- (Rochon and Johnston, 1991), carmo- (Carrington and Morris, 1986), hordei- (Gustafson et al., 1987) and luteo- (Tacke et al., 1990; Miller and Mayo, 1991) virus groups. The subgenomic RNAs of a tombusvirus (Rochon and Johnston, 1991) and a luteovirus (Tacke et al., 1990) have been shown to act as messages for translation of overlapping genes in which initiation occurs at two nearby out-of-frame AUGs.

Luteoviruses have one genomic RNA of about 5.7kb that encodes six open reading frames (ORFs) and is not polyadenylated. Luteoviruses can be divided into two distinct groups (reviewed by Martin et al., 1990). The subject of this paper, the PAV serotype of BYDV (BYDV-PAV, Miller et al., 1988) has a genome organization as shown in Fig. 1A. In contrast, beet western yellows virus (BWYV, Veldt et al., 1988), potato leafroll virus (PLRV; van der Wilk et al., 1989; Mayo et al., 1989; Keese et al., 1990) and the RPV serotype of BYDV (BYDV-RPV, Vincent et al., 1991) contain an
extra 5'-terminal ORF and lack a homolog to the 6.7K ORF. Moreover, the putative polymerase gene of BYDV-PAV shows more similarity to those of the carmoviruses (Miller et al., 1988) than to polymerases of the other sequenced luteoviruses. The polymerases of the other luteoviruses are more closely related to those of the sobemoviruses (Veidt et al., 1988; Habili and Symons, 1989).

Gene expression of BYDV has yet to be characterized. Based on its genome organization, BYDV-PAV likely uses several strategies to express its genes. The first two ORFs located at the 5'-half of the BYDV-PAV genome are translated from genomic RNA. The 60 kilodalton (kDa) protein, a putative RNA-dependent RNA polymerase, appears to be expressed by a translational frameshift event (Brault and Miller, 1992), resulting in a 99 kDa fusion protein. We proposed that the 3' half of the genome is expressed from two subgenomic mRNAs, with the 22K, 17K and 50K ORFs all expressed from the large subgenomic RNA (Miller et al., 1988). The 6.7K ORF of unknown function may be encoded by the small subgenomic RNA. Evidence has been given in support of a similar mechanism for translation of the coat protein and overlapping genes from PLRV (Tacke et al., 1990). However, due to the distinct differences of BYDV-PAV from PLRV-like luteoviruses described above, we set out to determine the strategy used by BYDV-PAV.
MATERIALS AND METHODS

Virus and Viral RNA Purification

The Illinois isolate of BYDV-PAV was kindly provided by Anna Hewings, USDA/ARS, Univ. of Illinois. Virus was propagated in oats (Avena sativa cv. Clintland 64). Virions and viral RNA were extracted as described by Waterhouse et al. (1986).

Electrotransfection of Protoplasts with BYDV RNA

Protoplasts were isolated from an oat (Avena sativa cv. Stout) suspension culture, obtained from H. Rines (USDA/ARS, Univ. of Minnesota), using a method similar to that described for Triticum monococcum by Young et al. (1989) with several modifications. The cells were subcultured weekly in MS medium (Murashige and Skoog, 1962) in a 1:4 ratio of suspension:medium and maintained by shaking at room temperature. Three days after subculture, cells were digested in a 9:1 solution of 0.6 M mannitol:artificial sea water (ASW: 311 mM NaCl, 6.9 mM KCl, 18.8 mM MgSO_4, 16.7 mM MgCl_2, 6.8 mM CaCl_2, 1.75 mM NaHCO_3, 10 mM MES pH 6.0) containing 10 ml of enzyme solution [0.5% Cellulysin (Calbiochem), 0.75% Hemicellulase (Sigma), 0.3% Driselase (Sigma)]. This mixture was incubated for 18 hours at room temperature with shaking at 40 rpm. The protoplasts were then filtered through nylon membranes of 70 mm and 40 mm and sedimented at 100 x g for 5 min. The protoplasts were washed twice in a solution of 0.6 M mannitol:ASW (1:1). After a final wash in electroporation buffer (EPB: 10mM HEPES, 150mM NaCl, 0.2M mannitol,
4mM CaCl2, pH 7.2; Fromm et al., 1985) the pellet was resuspended in EPB containing 0.2 mM spermidine.

Approximately 1-2x10⁶ protoplasts were incubated on ice for 15 min, and 0.5 µg of viral RNA was added to the protoplasts immediately prior to electroporation using 1 pulse at 250 V and 500 µF (Gene Puiser, BioRad). After 15 min on ice, protoplasts were diluted 6-fold in a solution of MS:0.6 M mannitol (1:1). Transfected protoplasts were incubated at 28° in the dark for 0 to 72 hours before harvesting by centrifugation for 5 min at 100 x g. Infection was verified by ELISA of protoplast extracts using BYDV-PAV-specific antibodies kindly provided by Stewart Gray (USDA/ARS, Cornell Univ.).

**Plasmid Construction**

DNA manipulations were performed essentially as described by Sambrook et al. (1989), except where indicated otherwise. All plasmids were cloned in *Escherichia coli* (strain DH5α'). All nucleotide numbering (nt xxxx) refers to the BYDV-PAV genome (Miller et al., 1988).

(i) Plasmids pSP2, pSP3, pSP10 and pSP16 were used to make probes for northern blot analysis (map positions in Fig.1B). To generate pSP2 and pSP3, Accl fragments (nt 2531-2786 and nt 2787-2985, respectively) were isolated from pPA259 (Miller et al., 1988). The termini were converted to blunt ends with T4 DNA polymerase and ligated into HincII-digested pGEM3Zf(+) [Promega, Madison, WI]. cDNA clone pPX25 was prepared (Gubler, 1988) from PAV RNA using the primer: 5' CCCGGGGTTGCGAACTGCTCTTTTCG 3'. This is complementary to the 3' end
of the PAV genome, with the addition of three C residues at the 5' end of the primer to create a unique Smal site. XbaI linkers were ligated to the cDNA, which was then digested with XbaI, and cloned into XbaI-cut pUC18. pPX25 contains the 3'-terminal 1450 bases of BYDV-PAV sequence. pSP10 was constructed by subcloning the complete pPX25 insert into the HindIII and EcoRI sites of pGEM3Zf(+) . To obtain pSP16, the HindIII-SspI fragment of pPA142 (Miller et al., 1988), (nt 1591-2739), was introduced into HindIII-Smal digested pGEM7Zf(+).

(ii) pSP9 : To produce antisense probe for RNase protection studies (Fig.1B), plasmid pSP9 was constructed by subcloning an SspI-SalI fragment (nt 2740-2990) from pPA259 into Smal-SalI digested pGEM3Zf(+).

(iii) pSP17, pSP4, pSP18 and pSP19: These plasmids were used to synthesize subgenomic RNA transcripts for in vitro translation. Positions to which the clones map in the viral genome are shown in Figures 4A and 5A. Plasmid pSP17 was constructed using polymerase chain reaction (PCR) to obtain transcripts containing the precise 5' end of BYDV-PAV subgenomic RNA. First, plasmid pSP1 was generated by cloning the SspI-PstI fragment of pPA259 (nt 2740-3476, PstI site is in the vector) into Smal-PstI digested pGEM3Zf(+). Then, an upstream primer: 5' GGTCTAGATAA7ACGACTCACC 
ATAGATAGGGTTTAGTTAATG3' , containing (5' to 3'): an XbaI site (underlined), the T7 RNA polymerase promoter (italics) and BYDV-PAV sequence from nt 2769 to 2791; and a downstream primer: 5' TCACACAGGAAACAGCTATGAC 3' (pUC/M13 "reverse" primer) were used with circular pSP1 plasmid as a template to amplify an 825 bp fragment by
PCR. The PCR-amplified product was gel-purified, digested with XbaI-PstI and cloned into XbaI-PstI cut pUC118.

To generate pSP4, the sequence upstream of and including the coat protein (CP) start codon was removed from pSPI by digesting with EcoRI. Of the three resulting fragments, the 481bp and 3306bp (vector) fragments were gel purified and religated. The viral sequence in transcripts from pSP4 begins at base 3 of the CP gene (nt 2860 in the genome), and ends at nt 3476 (3' of the CP stop codon).

Plasmid pSP18 was constructed to generate the full-length subgenomic transcripts with precise 5' and 3' termini. First pPA8 (Miller et al., 1988) and pSPI0 were joined at the BsmI site (nt 4600) as follows. The SalI-BsmI fragment (nt 3308 to nt 4600, SalI is in the vector) of pPA8 was cloned into SalI-BsmI cut pSPI0 to generate pSPII. Then the 481bp AccI fragment of pSPI (nt 2985-3466) was inserted into XbaI-cut pSPII to create pSP12. Plasmid pSP12 was digested with XbaI (in the vector), made blunt and digested with SalI (nt 2985). The resulting fragment was inserted into pSPI7, which had been prepared by cleavage with PstI in the vector, converted to blunt ends and subsequently digested with SalI (nt 2985).

To obtain transcripts lacking both the CP and 17K ORF start codons, plasmid pSP19 was generated. The SalI-Smal (nt 2985-5677) fragment of pSPI2 was cloned into SalI-Smal cut pGEM3Zf(+) for in vitro transcription.

In Vitro Transcription

Synthetic transcripts used for in vitro translation were synthesized from HindIII-linearized pSPI7, pSPI4 and SalI-linearized pSPI8, pSPI9
using T7 RNA polymerase as described by the manufacturer (Promega, Inc.,). Briefly, 100 μl reactions containing 40mM Tris-Cl, pH 7.5, 6mM MgCl₂, 2mM spermidine, 10mM NaCl, 10mM dithiothreitol (DTT), 500mM each of the four ribonucleoside triphosphates (rNTPs), 100U RNasin, 8 μg of linearized plasmid DNA template and 100 U of T7 RNA polymerase were incubated for 1h at 37°. To synthesize capped transcript, 50μM GTP, 500μM each ATP, CTP, UTP, and 500μM m7G(5')ppp(5')G (New England Biolabs) were used. Template DNA was removed by digestion with 1U/μg DNA of RQ1 RNase-free DNase I (Promega) for 15 min at 37°. The reaction mixture was extracted twice with phenol-chloroform, and ethanol precipitated. Final RNA concentration was determined spectrophotometrically.

Strand-specific RNA probes used for northern hybridizations and RNase protection assay were synthesized as above by including [α-³²P]UTP. Plasmids pSP2 and pSP3 were XbaI-linearized and pSP9 was SacI-linearized prior to transcription with SP6 RNA polymerase. Plasmids pSP16 and pSP10 were linearized at SacI and PvuI sites, respectively, prior to transcription with T7 RNA polymerase.

RNA Extraction

Total RNA from 200 mg infected or healthy leaves was isolated as described by Wadsworth et al. (1988). Total nucleic acids from protoplasts were isolated by the method of Young et al. (1989). For northern blot analysis, RNA was separated from DNA by LiCl precipitation of the RNA (Sambrook et al., 1989). RNA used in primer extension and RNase protection
was purified using Chromaspin-1000 columns (Clontech Laboratories, Inc., Palo Alto, CA).

**Northern Blot Analysis**

RNA (5-10 µg) was electrophoresed on 1.2% agarose gels containing 1.1% formaldehyde, blotted to nitrocellulose filters in 10X SSC (1x SSC is 0.15 M NaCl and 0.015 M sodium citrate, pH 7.0), and fixed by UV crosslinking. Membranes were prehybridized at 52°C for at least 2h in 50% formamide, 5x SSC, 20mM sodium phosphate, pH6.5, 0.1% SDS and 0.2mg/ml polyanetholesulfonic acid (Sigma). Hybridization was performed in the same buffer containing 1x10^7 cpm of 32P-labeled RNA transcripts for 12-16h at 52°C. Following hybridization, filters were washed once with 1xSSC-0.1% SDS for 20 min at room temperature, then 3 times for 20 min each, at 68°C in 0.1xSSC-0.1%SDS. Blots were dried, and exposed to X-ray film with an intensifying screen at -80°C.

**RNase Protection**

RNase protection was performed as described by Melton *et al.* (1984). Approximately 1x10^5 cpm of labeled RNA transcript was hybridized to 5-10 µg total RNA isolated from protoplasts in 80% formamide, 40mM PIPES, pH6.7, 0.4 M NaCl and 1mM EDTA at 42°C for 1h. Following hybridization, samples were digested with an RNase A/T1 cocktail (Ambion, Austin, TX) for 30 min at 37°C. Digestion was terminated by 0.5%SDS and 100 µg proteinase K at 37°C for 15 min. After phenol-chloroform extraction and ethanol precipitation, protected fragments were denatured in formamide
EDTA buffer at 95° for 5 min and analyzed by electrophoresis on a 6% polyacrylamide, 7M urea gel.

**Primer Extension**

Primer extension was carried out using oligomer P1: 5' TTGCGCGTCTAGGTCTCTA 3', complementary to nt 2875 to nt 2894 of BYDV-PAV genomic RNA. This primer was 5' end-labeled with T4 polynucleotide kinase (New England Biolabs, Beverly, MA) and [γ-32P]ATP (Sambrook et al., 1989). Radiolabeled primer (2x10^6 cpm) was annealed to 5-10 µg of total RNA isolated from protoplasts in 50mM Tris-Cl, pH8.3 and 75mM KCl containing 1U/ml RNasin (Promega) for 30 min at 65°. Annealed primer was then extended by addition of 20mM DTT, 1mM of all four deoxynucleoside triphosphates and 200 units of MMLV H+ reverse transcriptase (Superscript, BRL) and incubating at 48° for 30 min. Extended products were ethanol precipitated and resuspended in 80% formamide-20mM EDTA-0.1% xylene cyanol-0.1% bromophenol blue. After heating for 5 min at 95°, samples were electrophoresed through a 6% polyacrylamide/7M urea sequencing gel and visualized by autoradiography. The sequencing ladder was generated by double stranded sequencing (Sanger, 1977) of pSP9 using the above primer (P1).

**In vitro Translation and Immunoprecipitation**

Approximately 1 µg of transcript was translated in 20 µl reaction volume in the presence of [35S]-methionine (NEN) using nuclease treated rabbit reticulocyte lysate as described by the manufacturer (Promega). In some cases the lysate was supplemented with extra MgCl2 and potassium...
acetate. Translation products were immunoprecipitated (Ball, 1990) using PAV-specific antibodies. One tenth volume of the products was electrophoresed through 5% stacking/12% resolving polyacrylamide-SDS denaturing gels using discontinuous buffer system (Laemmli, 1970). Gels were fixed overnight, stained and either autoradiographed or fluorographed (Hames, 1990). Radioactivity in the gels was quantified using a PhosphorImager 400E (Molecular Dynamics, CA).
RESULTS

Characterization of BYDV-PAV RNAs

Northern blots were employed to determine the number, approximate size and location of subgenomic RNAs generated during BYDV-PAV infection. Total RNA from infected leaves or transfected oat protoplasts was hybridized with labeled, strand-specific RNA probes spanning different regions of the genome (Fig 1B & C). None of these probes hybridized to RNA from healthy leaves or mock-inoculated protoplasts, demonstrating the specificity of these probes for BYDV-PAV RNA. Initial experiments used total RNA from infected plants, but protoplasts were used subsequently due to ease of RNA isolation. The ratio of subgenomic RNAs to genomic RNA was higher in protoplast-derived RNA than plant-derived RNA. (Fig 1C and data not shown). All of the probes hybridized to genomic length RNA of about 5.7kb. Probe pSP10, covering the 3' end of the genome, detected a 2.9kb subgenomic RNA (sgRNA1) and much higher amounts of a 0.8kb subgenomic RNA (sgRNA2) in protoplasts (Fig. 1C) and plants (data not shown), whereas hybridization with probe pSP16, corresponding to an internal part of the putative polymerase gene (60K ORF), did not detect these RNAs. Probe pSP2, covering the 3' end of 60K ORF and part of the following intergenic region up to nt 2786, revealed only one subgenomic RNA of 2.9kb. RNA of this size was also detected with a probe (pSP3) encompassing part of the CP/17K sequence and the intergenic region 3' of nt 2786. The ratio of subgenomic to genomic RNAs was reduced in RNA
FIG. 1. Characterization of BYDV-PAV RNA in infected cells. A. Genome organization of BYDV-PAV indicating sites of proposed abnormal translational events. Boxes represent open reading frames (ORFs). POL = polymerase; CP = coat protein; VPg = possible genome linked viral protein. B. Relative positions of strand-specific RNA probes and oligonucleotide primer. UAA indicates stop codon of 60K ORF. AUGs indicate start codons of CP and 17K ORFs. Positions of (-) strand RNA transcripts used in northern hybridization (open boxes) and RNase protection assay (solid box) are shown below the expanded portions of the genome. Numbers at ends of bars and ORFs indicate positions in BYDV-PAV genome (Miller et al., 1988). Primer P1 (arrowhead) is the 20 base oligomer used for primer extension. C. Northern blot analysis of total RNA from mock (H) and BYDV-PAV infected (I) oat protoplasts (pSP16 and pSP10) and plants (pSP2 and pSP3). RNA was extracted at 36h and 72h after inoculation of protoplasts. Blots were hybridized with 32p-labeled strand-specific probes representing different portions of genome (panel B). Positions of genomic RNA (G), subgenomic RNA1 (sgRNA1), and subgenomic RNA2 (sgRNA2) and mobilities of RNA markers (BRL) are shown beside autoradiographs.
samples extracted from plants. However, the 0.8kb was always in vast excess relative to the 2.9kb sgRNA1 (data not shown). These observations suggest that sgRNA1 is synthesized both in infected leaves and protoplasts and that the 5' end of sgRNA1 is within the 47 nucleotide region upstream of the AccI site (3' end of pSP2) at base 2786 and downstream of the SspI site (3' end of pSP16) at base 2739.

**Identification of the 5' end of sgRNA1**

To more precisely map the 5' end of sgRNA1, RNase protection analysis was used. Labeled antisense probe pSP9 (Fig. 1B), representing the full intergenic region between the 60K ORF and coat protein gene and part of the CP/17K coding region (corresponding to BYDV-PAV sequence 2740 to 2990), was hybridized to total RNA from mock- and BYDV-PAV-inoculated protoplasts. The RNA was then treated with ribonuclease. Major RNA probe fragments of 250 and 221 nt were protected by RNA isolated from protoplasts 36h and 72h after inoculation with BYDV RNA (Fig. 2A). In contrast, the probes were not protected by RNA from BYDV RNA-inoculated protoplasts at 0 h or mock-inoculated protoplasts at any time point. The 250 nt band is equal in size to the portion of the probe containing viral sequence and represents the fragment protected by genomic RNA. The 221 nt protected fragment is in the expected size range for a fragment protected by sgRNA1. This leads to a predicted 5' end of sgRNA1 very near base 2769. A faint band of approximately 165 nt was also protected. However, northern analysis and primer extension (below) provide no evidence for an additional sgRNA of this size.
FIG. 2. Mapping of BYDV-PAV subgenomic RNA1 (sgRNA1) 5' end.

A. RNase protection analysis of total RNA from mock (H) and BYDV-PAV RNA (I) transfected oat protoplasts at 0, 36, and 72h post-inoculation. RNA was hybridized with $^{32}$P-labeled antisense probe pSP9 (Fig 1B) and digested with RNases A and T1. Protected fragments were analysed on a 6% polyacrylamide gel containing 7M urea. Sizes of fragments protected by genomic (G) and subgenomic (sgRNA1) RNA are indicated at left, whereas those on the right indicate sizes of in vitro transcribed RNA markers. N indicates no RNA is included in hybridization mixture. B. Primer extension analysis of RNA from mock (H) and BYDV-PAV RNA (I) transfected oat protoplasts. Total RNA was annealed to end-labeled primer P1 and extended with reverse transcriptase (Fig. 1B). The elongation products were separated on a 6% sequencing gel. The arrow indicates the 5' terminus of the subgenomic RNA1 deduced from the gel. Sequencing ladder shown on the left is generated from double stranded DNA plasmid pSP9 using primer P1 (Fig. 1B). N indicates no RNA included in annealing mixture.
To confirm the 5' end of subgenomic RNA1, primer extension analysis was performed. End-labeled primer (P1, Fig. 1B) was annealed to total RNA from mock- and BYDV-PAV RNA-infected protoplasts and extended by reverse transcriptase. An extension product was detected in the RNA samples prepared from infected protoplasts (Fig. 2B), whose size suggested that it was due to reverse transcriptase reaching the 5' end of sgRNA1. The 5' end at base 2769 is clearly revealed by comparison with the adjacent DNA sequencing ladder generated with the same primer on cDNA clone pSP9. Other, fainter high molecular weight bands were also seen. No band with a 5' end at position 2769 was detected when primer extension was performed on virion RNA (data not shown). These findings verified the results of RNase protection analysis, indicating that the sgRNA1 start site corresponds to the guanosine residue at nucleotide 2769. This corresponds to an 89 nucleotide 5'-untranslated leader upstream of the coat protein start codon AUG at nt 2858.

**In vitro Translation Products of Synthetic Subgenomic RNA**

Once the precise 5' end of the sgRNA1 was determined, cDNA clones were constructed from which sgRNA1 containing perfect 5' ends could be transcribed for *in vitro* translation assays. To test the hypothesis that coat protein and 17K ORFs are translated by initiation at two separate AUG codons in the same subgenomic RNA, transcript pSP17 (containing start codons of both ORFs) and transcript pSP4 (containing only the 17K start codon) were translated in a cell-free rabbit reticulocyte lysate system (Fig. 3A). Translation of the pSP17 transcript resulted in large amounts of a
FIG. 3. *In vitro* translation of BYDV-PAV synthetic subgenomic RNAs containing CP and 17K ORFs.

A. Schematic representation of the pSP17 and pSP4 transcripts used for *in vitro* translation. The '+' or '-' indicates the presence or absence of AUG start codon of the indicated ORF. B. Translation products of transcripts from pSP17 (T, lane 1) and from pSP4 (T, lane 5) were obtained in a cell-free rabbit reticulocyte system. Translation products were immunoprecipitated with BYDV-PAV antiserum (I, lanes 2 and 6) or with preimmune serum (P, lanes 3 and 7). Negative control shows products with no added RNA (N, lane 4). Products were separated by 12% polyacrylamide-SDS gel electrophoresis (Laemmli, 1970) and visualized by autoradiography. On the left, positions of coat protein (CP) and 17K protein bands are indicated. Mobilities of molecular weight standards (in kDa) are at the right.
17 kDa polypeptide and less accumulation of a 22 kDa polypeptide (Fig. 3B, lane 1). In contrast, translation of pSP4 transcript in which the coat protein AUG was deleted, resulted in only the 17 kDa polypeptide (Fig. 3B, lane 5). This shows that the 17K ORF is likely to be translated by initiation at the second AUG (base 2901) in sgRNA1. These translation products were further characterized by immunoprecipitation with BYDV-PAV-specific antiserum. Antiserum reacted with 22 kDa product but not the 17 kDa product, indicating that the former is indeed BYDV-PAV coat protein (Fig. 3B, lanes 2 and 6). None of the products was precipitated by preimmune serum (Fig. 3B, lanes 3 and 7). Translation of capped transcripts of pSP17 and pSP4 gave the same pattern of polypeptides as uncapped transcripts (data not shown).

To study translation of the 50K ORF, a full-length subgenomic transcript (pSP18) and a 5'-truncated transcript (pSP19) were translated in rabbit reticulocyte lysates (Fig. 4A). As with the previous constructs, 17K and CP gene products were produced in the initial translation conditions, yet no product (72 kDa polypeptide) consistent with readthrough of the CP stop codon was detected (Fig. 4B, lane 1). Instead, an unexpected additional polypeptide of about 42 kDa was produced, possibly by initiation at an internal AUG located in the 50K ORF at position 3665 in the genome (Fig. 4A). This is supported by the observation that the 42 kDa polypeptide was also translated from an RNA lacking both the CP and 17K start codons (Fig. 4B, lane 2).
FIG. 4. In vitro translation of synthetic subgenomic RNA1 of BYDV-PAV.
A. Schematic representation of pSP18 and pSP19 transcripts used for in vitro translation. Numbers indicate the nucleotide positions of ends of ORFs and transcripts. UAG at 3460 indicates the leaky stop codon separating CP and 50K proteins. B. Influence of MgCl2 concentration on translation of synthetic subgenomic RNAs. pSP18 and pSP19 transcripts were translated in a rabbit reticulocyte system with the indicated concentrations of MgCl2. Indicated on the left are positions of the 17K and CP translation products, and bands consistent with initiation at AUG 3665 and readthrough of the CP stop codon. C. Influence of MgCl2 and potassium acetate (K Ac) on in vitro translation of synthetic subgenomic RNAs. Translation products of transcripts pSP18 (T, lanes 1 and 3) and pSP19 (T, lanes 2 and 6) in the presence of different concentration of MgCl2 and K Ac as indicated. These products were immunoprecipitated with BYDV-PAV antiserum (lanes 4 and 7) and with preimmune serum (P, lanes 5 and 8).
Magnesium chloride and potassium acetate concentrations were increased in attempts to observe readthrough (Pelham, 1978; Carrington and Morris, 1985). Readthrough-size polypeptide (72 kDa) was observed with addition of MgCl₂ to the translation mix (Fig. 4B, lanes 3 and 5), clearly indicating that synthesis of the 72 kDa polypeptide was sensitive to changes in MgCl₂ concentration. The relative amount of 72 kDa product was 7.3% of CP product at a MgCl₂ concentration of 1.9 mM (Fig. 5). This ratio was further increased to 15-16% of CP when 159 mM potassium acetate was combined with MgCl₂. Overall, the CP, 17 kDa and 42 kDa protein synthesis was decreased with increased potassium acetate and MgCl₂ concentrations.

To determine the identity of the 72 kDa polypeptide, the translation products were immunoprecipitated with BYDV-PAV-specific antiserum. Antiserum reacted with the 72 kDa polypeptide, indicating that this peptide is present in the purified virus preparation (Fig. 4C, lane 4). Lesser amounts of other translation products were immunoprecipitated. They may have arisen as a result of internal initiation or premature termination of translation at many sites along the template (Fig. 4C, lane 4). None of the products was immunoprecipitated with preimmune serum (Fig. 4C, lanes 5 and 8).
FIG. 5. Quantification of effects of MgCl₂ and KAc on relative translation of various polypeptides from pSP18 transcript. Gels in Figs. 5B and 5C were scanned using a Phosphorimager. Log of the molar ratios of 17K, readthrough (RT), and 42K proteins with respect to CP were plotted at the indicated concentrations of MgCl₂ and KAc.
DISCUSSION

Subgenomic RNA Mapping

Due to the ease of extraction and increase in proportion of viral RNAs, mapping of the subgenomic RNAs was facilitated by the use of infected protoplasts. Electroporation and PEG have been used previously to efficiently infect protoplasts from *Triticum monococcum* and barley (Young *et al.*, 1989; Larkin *et al.*, 1991), but oat protoplasts had not been inoculated efficiently (Barnett *et al.*, 1981). The increased ratio of sgRNAs to genomic RNA in protoplast-derived RNA relative to plant-derived RNA probably indicates that more virus particles, which contain only genomic RNA, accumulated in plants from which RNA was isolated 10 days after inoculation. In addition, the nonencapsidated and thus unprotected, sgRNAs are likely to have been degraded preferentially during the more prolonged extraction procedure from whole plants than from protoplasts.

The two sgRNAs identified by northern hybridization are in agreement with previous studies (Miller *et al.*, 1988; Young *et al.*, 1991), which revealed but did not map RNAs of these sizes in infected tissue. Young *et al.* (1991) detected an additional band of about 1.8 kb that is neither present in our gels, nor predicted from the genome organization. The sizes of the sgRNAs are similar to the predominant dsRNAs detected in BYDV-PAV-infected plants (Gildow *et al.*, 1983), which may represent replicative forms of the sgRNAs. No evidence for a separate sgRNA for translation of the 17K ORF was obtained. This supports the hypothesis that the two overlapping
ORFs are translated from a single message (Miller et al., 1988; Veldt et al., 1988; Smith and Harris, 1990; Tacke et al., 1990; Vincent et al., 1990).

Based on sequence similarities, Tacke et al. (1990) predicted that the 5' end of BYDV-PAV sgRNA1 would map to position -40 (nt 2818). More recently, Miller and Mayo (1991) reported the subgenomic RNA of a Scottish isolate of PLRV to be 172 nucleotides longer at its 5' end than that predicted by Tacke et al. The 5' end of the PLRV subgenomic RNA, which is located within the 3' end of the polymerase ORF, has the sequence ACAAAA. This sequence is also present at the 5' terminus of PLRV genomic RNA, suggesting that it is the complement of an origin of (+) strand synthesis on the (-) strand (Miller and Mayo, 1991). BYDV-PAV has this sequence at the same relative position (nt 2729) in the 3' end of the putative polymerase ORF, and at position 19 in the genomic RNA. Yet we present three independent lines of evidence in support of a start of sgRNA1 at nt 2769. sgRNA1 was detected in northern blots with probe pSP2 which cannot hybridize to bases 5' of nt 2786 (-72 relative to CP AUG). Conversely, probe pSP16 which extends 5' from base 2739 detected no sgRNA1. These hybridization patterns suggest that the 5' end of subgenomic RNA1 is located between nt 2739 and nt 2786 (-119 to -72 relative to the CP AUG). RNase protection and run-off reverse transcription both predict that the 5' end initiates at nt 2769. A shadow band at position 2768 (Fig. 2B) may indicate presence of a m7G cap that is not copied well by reverse transcriptase (Ahlquist and Janda, 1984). We detected no bands corresponding to position -40 as predicted by Tacke et al., or at -130 as predicted by Miller and Mayo (1991).
The ACAAAA sequence may be required for polymerase recognition but may not specify a precise initiation site. Because of the unrelatedness of the polymerase genes of PLRV and BYDV-PAV, it is possible that the viral replicases recognize different subgenomic RNA initiation signals. Yet, Vincent et al. (1991) found another region of homology immediately upstream of the PAV start site in which 9 out of 12 bases at nt -102 to -91 match in all five known luteovirus sequences.

We searched for similarities between the two putative subgenomic promoter regions within the BYDV-PAV genome. Although sgRNA2 was not precisely mapped, its size suggests that the 5' end is slightly upstream of the 6.7K ORF start codon. This ORF is required for infection (Young et al., 1991). A short alignment, including a sequence immediately 5' of the sgRNA1 start site is possible (Fig. 6). This includes a subset of the ACAAAA sequence: CAAA at the 5' end of the matching sequences. In the vicinity are identical 10 nt tracts located 20 bases upstream of the CP AUG and 80 bases upstream of the 6.7K AUG. Lack of more significant homology between the two subgenomic promoter regions is not surprising because they are regulated quite differently; sgRNA2 accumulates to much higher levels than sgRNA1 (Fig. 1; and Young et al., 1991). More detailed comparison awaits precise mapping of the 5' end of sgRNA2.

**Translation Initiation at Two AUG Codons.**

A synthetic subgenomic RNA1 containing a precise 5' end was constructed to rule out spurious effects on translation in vitro due to a truncated leader or one containing vector bases. This experiments differs
FIG. 6. Possible alignment between subgenomic promoter regions of BYDV-PAV sgRNAs. Arrow indicates 5' end of sgRNA1 (from Fig. 2). Underlined bases are conserved among known luteoviruses in the intercistronic region between polymerase and CP genes (Vincent et al., 1991). Boxed regions represent 10 base perfect repeats. Italicized numbers indicate number of bases not shown. Start codons of CP and 6.7K ORFs (nt 2858 and 4921, respectively) are shown at right.
from those of other studies (Carrington and Morris, 1985; Tacke et al., 1990) in which extra bases were present at the 5' ends of synthetic subgenomic RNAs. Since we do not know whether sgRNA1 is capped, both capped and uncapped transcripts were translated, but no difference was observed. Translation of a transcript containing the complete 22 kDa CP and 17K ORFs yielded the 17 kDa product in a 7-fold molar excess to the CP. Increasing levels of MgCl₂ and potassium acetate in the translation mixture decreased this ratio to 1:1 (Fig. 5). These higher levels of salts have been reported to improve the fidelity of translation initiation (Kozak, 1990a). Thus the 1:1 ratio may be more correct and the 42 kDa product may be an artifact. The 7:1 ratio is same as that observed when the 5' sequence of the PLRV sgRNA, including either of the two AUGs, was placed in front of a reporter gene in vivo (Tacke et al., 1990). However, simultaneous expression of genes initiating at both AUGs on one mRNA was not possible in that system. The high level of initiation at the 17K start codon is surprising if the 17K ORF encodes the VPg as predicted by its molecular weight (Miller et al., 1988; Murphy et al., 1989). Because a VPg would be needed in only one copy per virion, whereas 180 copies of the CP are required, high levels of CP expression would be expected. However, the levels of expression of these proteins may be regulated differently in the infected cell. More recent evidence suggests that the 17K is a nonspecific, single stranded nucleic acid binding protein (Tacke et al., 1991).

The high level of internal initiation at the second AUG may be explained by the fact that the 17K AUG is flanked by a more frequently used
translation initiation context than the CP start codon (Cavener and Ray, 1991; Kozak, 1986; Lutke et al., 1987). In fact, when the contexts of 10 other plant viruses (including 4 other luteoviruses) which initiate at two nearby AUGs in different reading frames are compared, in all cases the downstream AUG is in a "better" context (Fig. 7A). This supports the leaky scanning model in which ribosomes usually initiate at the first AUG, but sometimes continue on to the second AUG before initiating if the latter is in a more optimal sequence context (Kozak, 1986). Regarding the significance of sequence context, it should be noted that some very highly expressed plant viral genes have start codons in "poor" sequence contexts (e.g. BMV, CarMV) but they are translationally monocistronic mRNAs. In the case of BYDV-PAV, the CP AUG may be sequestered in a stem-loop structure that could make it less accessible to ribosomes (Fig. 7B). The existence of this structure is supported by preliminary studies using partial nuclease digestion of this region under nondenaturing conditions (data not shown). Base-paired regions upstream (Kozak, 1989) and downstream (Kozak, 1990b) from start codons can have negative and positive effects, respectively, on efficiency of initiation. Thus, the upstream secondary structure (Fig. 7B) may also inhibit ribosome binding (Kozak, 1989) for translation initiation at the CP AUG. Confinement of the first AUG in a stem-loop structure was proposed for Kennedya yellow mosaic virus (KYMV) RNA (Ding et al., 1990), but no conserved secondary structures were observed among the other RNAs in Fig. 7A. Site-directed mutagenesis is underway to determine the relative
FIG. 7. A. Sequence contexts of start codons from overlapping ORFs of members of plant virus groups known to initiate translation at two neighboring AUGs in the same RNA. The most frequently used bases at positions -3 through +2 (Cavener and Ray, 1991) are: [A/G]CNAUGGC for monocots and AN[A/C]AUGGC for dicots. A base is defined as consensus if it is present at that position relative to >50% of the known start codons. Two bases are shown at one position (in brackets) if one or the other is present >75% of the time. Those bases that match the consensus are outlined, those which are found at that position flanking <10% of known start codons are in lower case letters. In the case of luteoviruses, the first ORF is the CP gene, the second is the homolog of the 17K ORF in BYDV-PAV. The BYDV-MAV and BYDV-RPV sequences are from Ueng et al. (1992) and Vincent et al. (1991), respectively. TYMV (turnip yellow mosaic virus, Weiland and Dreher, 1989), EMV (eggplant mosaic virus, Osorio-keese et al., 1989), KYMV (Kennedya yellow mosaic virus, Ding et al., 1990) and OYMV (oronis yellow mosaic virus, Ding et al., 1989) are tymoviruses. The first and second AUGs in tymoviral genomes are functional start codons, located five bases apart giving rise to large out-of-frame overlapping genes. CNV (cucumber necrosis virus, Rochon and Johnston, 1991), and TBSV (tomato bushy stunt virus, Hillman et al., 1989) are tombusviruses. The AUGs in the tombusviruses give rise to 21 and 20K out-of-frame ORFs expressed from a subgenomic RNA. B. Possible secondary structure in the 5' end of BYDV-PAV sgRNA1. Free energies (in kcal/mol) shown beside each helix were calculated as in Frier et al. (1986) using the program RNASE (Cedergren et al., 1988). Start codons of the CP and 17K ORFs are in bold, underlined text. Boxed region is the 10 base conserved sequence in Fig. 6.
### A

<table>
<thead>
<tr>
<th>Virus</th>
<th>1st ORF</th>
<th>2nd ORF</th>
</tr>
</thead>
<tbody>
<tr>
<td>BYDV-PAV</td>
<td>uGA AUG AA</td>
<td>AAA AUG GC</td>
</tr>
<tr>
<td>BYDV-MAV</td>
<td>uGA AUG AA</td>
<td>GGA ALG GC</td>
</tr>
<tr>
<td>BYDV-RPV</td>
<td>uA AUG AG</td>
<td>cGA AUG GC</td>
</tr>
<tr>
<td>BWYV</td>
<td>uUA AUG AA</td>
<td>uCA ALG GA</td>
</tr>
<tr>
<td>PLRV</td>
<td>uUA AUG AG</td>
<td>GA ALG UC</td>
</tr>
<tr>
<td>TYMV</td>
<td>cAAA AUG AG</td>
<td>GU ALG GC</td>
</tr>
<tr>
<td>EMV</td>
<td>uCA AUG C</td>
<td>uCA ALG GC</td>
</tr>
<tr>
<td>KYMV</td>
<td>cUG AUG UC</td>
<td>cA ALG GC</td>
</tr>
<tr>
<td>OYMV</td>
<td>uGAUG UC</td>
<td>cUA ALG GC</td>
</tr>
<tr>
<td>CNV</td>
<td>uGAUG GA</td>
<td>ACC ALG GA</td>
</tr>
<tr>
<td>TBSV</td>
<td>uGAUG GA</td>
<td>ACC ALG GA</td>
</tr>
</tbody>
</table>

### B

![RNA secondary structure diagram](image)

- ΔG = -14.2
- CP ORF Start
- Transcription Start
- 17K ORF Start
contributions of sequence context and secondary structure to the translational activity of the two start codons.

Readthrough

A commercial rabbit reticulocyte translation mix was satisfactory for translation of the CP and 17K ORFs, but translation of the 72K ORF could not be observed under standard conditions. By changing the concentrations of MgCl₂ and potassium acetate in the translation mix, synthesis of the 72 kDa readthrough polypeptide increased greatly (Fig. 4). Similar increases in stop codon readthrough due to addition of MgCl₂ and/or potassium acetate have been reported previously (Pelham, 1978; Carrington and Morris, 1985). These observations from BYDV-PAV synthetic sgRNA1 indicate that readthrough of the CP stop codon occurs at least in vitro. Evidence consistent with readthrough has also been observed in vitro for BWYV luteovirus (Veldt et al., 1988). The level of readthrough in vivo has been quantitated for only a few plant viruses and it is much lower than that found in vitro. Using a reporter gene, readthrough of the TMV 126K ORF stop codon was shown to be 5% (Skuzeski et al., 1990) and in the case of PLRV it was 0.9-1.3% (Tacke et al., 1990). The lower percentage readthrough in vivo may be due to a different population of suppressor tRNAs from than present in vitro (Beier et al., 1984).

The 72 kDa readthrough polypeptide was bound specifically by BYDV-PAV antiserum, indicating it is likely fused to the CP. This is supported by studies of PLRV (Bahner et al., 1990) and BYDV-PAV (Waterhouse et al., 1989), in which antibody made against readthrough protein alone reacted
only with the readthrough fusion protein and not with the CP, whereas antibodies against CP detected both proteins.

The mechanism involved in the readthrough process is not clearly understood. TMV readthrough seems to be mediated by a tRNA which reads the UAG as a tyrosine codon (Beler et al., 1984) and in case of Moloney Murine leukemia virus (Mo-MuLV) as glutamine (Yoshinaka et al., 1985). Extensive mutagenesis showed that the 6 bases 3' of the UAG are required for efficient readthrough of the 126K ORF of TMV (Skuzeski et al., 1991). Because this region shows little similarity with that of luteoviruses, all of which have the sequence AAAUAGGUAGA (the underlined UAG is the CP stop codon; Vincent et al., 1991), it is likely that luteoviruses use different signals than TMV. In addition, yet another cis-acting 3' sequence has been implicated for readthrough of a UAG in Mo-MuLV (Honigman et al., 1991). Site-directed mutagenesis is underway to understand the roles if any of the flanking bases on CP stop codon readthrough.
ACKNOWLEDGMENTS

The authors are grateful for technical assistance of Amy Morrow and Jean-Francois Bonnet. We thank Richard Lister and Mike Mayo for providing results prior to publication, and Adrianna Hewings and Stewart Gray for generously providing source material and valuable advice. This work was funded by a McKnight Foundation Individual Research Award in Plant Biology and USDA/CSRS grant no. 8900627. This is paper no. J-14666 of the Iowa Agriculture and Home Economics Experiment Station. Project numbers 2904 and 2936.
LITERATURE CITED


PAPER II  CONTROL OF START CODON CHOICE ON A PLANT VIRAL RNA
ENCODING OVERLAPPING GENES
Control of start codon choice on a plant viral RNA encoding overlapping genes.

Dinesh-Kumar, S.P., M.S.
Miller, W.A., Ph. D.

Plant Pathology Department and
Molecular, Cellular, and Developmental Biology Program
Iowa State University
Ames, Iowa 50011-1020
ABSTRACT

The signals that control initiation of translation in plants are not well understood. To dissect some of these signals, we used a plant viral mRNA on which protein synthesis initiates at two, out-of-frame start codons. On the large subgenomic RNA (sgRNA1) of barley yellow dwarf virus-PAV serotype (BYDV-PAV), the coat protein (CP) and overlapping 17K open reading frames (ORFs) are translated beginning at the first and second AUG codons, respectively. The roles of bases at positions -3 and +4 relative to the AUG codons in efficiency of translation initiation were investigated by translation of sgRNA1 mutants in a cell-free extract and by expression of a reporter gene from mutant sgRNA1 leaders in protoplasts. The effects of mutations that disrupted and restored secondary structure encompassing the CP AUG independently of, and in combination with, changes to bases -3 and +4 were also examined. Partial digestion of the 5' end of the sgRNA1 leader with structure-sensitive nuclease gave products that were consistent with the predicted secondary structure. Secondary structure had an overall inhibitory effect on translation of both ORFs. In general, the "Kozak rules" of start codon preference predominate in determining start codon choice. Unexpectedly, for a given CP AUG sequence context, changes which decreased initiation at the downstream 17K AUG also reduced initiation at the CP AUG. To explain this observation, we propose a new model in which pausing of the ribosome at the second AUG allows increased initiation at the first AUG. This detailed analysis of the roles of primary and secondary
structure in controlling translation initiation should be of value for understanding expression of any plant gene and in the design of artificial constructs.
INTRODUCTION

Translation can be a major step in the control of eukaryotic gene expression (Kozak, 1991; Merrick, 1992). In plants the control of translation has not been well studied (reviewed by Gallie, 1993). Because most plant viruses have RNA genomes, translation is a particularly important step in regulating gene expression (Sleat and Wilson, 1992). Regulation usually occurs during initiation of translation. According to the scanning model of eukaryotic translation initiation, the 40S ribosomal subunit, with associated factors, binds the m7GpppG cap structure at the 5' end of the mRNA and then scans in the 3' direction until the first AUG is encountered; at this point, translation initiation factor 2 (eIF2) escorts the tRNA^{Met} to the AUG codon, the 60S ribosomal subunit binds, and polypeptide synthesis ensues (Kozak, 1989a; Merrick, 1992). Here we define initiation as the steps at which the 60S ribosomal subunit binds and synthesis of the polypeptide chain begins. This is distinct from initiation of 40S subunit binding the mRNA.

Translation of 95% of eukaryotic mRNAs initiates at the 5' proximal AUG (Kozak, 1989a). However, certain bases flanking the AUG affect its efficiency as an initiation codon. The most efficient start codons are flanked by a purine, usually adenine, at the -3 position, and a guanine at position +4, in which the A of the AUG is numbered +1 and bases 5' of this are numbered negatively. These are the most common bases flanking start codons in plant mRNAs (Lutcke, et al., 1987; Cavener and Ray, 1991). If
the 5' proximal AUG lacks both of these features, a portion of the scanning 40S subunits may bypass that AUG and initiate at the second AUG if the latter lies in an appropriate context. This process has been dubbed leaky scanning (Kozak, 1986a; 1986b). Leaky scanning has been exploited by simian virus 40 (Sedman and Mertz, 1988), reovirus (Munemitsu and Samuel, 1988), possibly hepatitis B virus (Lin and Lo, 1992), and other vertebrate viruses to regulate translation of truncated or overlapping genes (reviewed by Kozak, 1992). This strategy allows viruses to simultaneously maximize coding capacity from minimal genetic material and to translationally regulate the molar ratios of the gene products. Alteration of the bases at positions -3 and +4 affected relative translation of overlapping reovirus genes as predicted by the leaky scanning model (Munemitsu and Samuel, 1988). Leaky scanning has not been demonstrated in plants.

The genome of the PAV serotype of barley yellow dwarf virus (BYDV-PAV) consists of a 5.7-kb RNA encoding at least six open reading frames (ORFs), as shown in Figure 1A. Three ORFs are expressed from a subgenomic mRNA (sgRNA1), comprising the 3' half of the genome, that is generated in infected cells (Dinesh-Kumar et al., 1992). sgRNA1 encodes the coat protein (CP) with a molecular weight of 22,047, an ORF with a molecular weight of 17,147 (17K ORF) that is completely nested within the CP gene, but in a different reading frame, and an ORF with a molecular weight of 49,797 (50K) that is expressed by read-through of the CP stop codon. We showed that the CP and 17K ORFs can be translated in vitro from a single mRNA by initiation at the first two AUG codons (Dinesh-Kumar
et al., 1992). The RNAs of other plant viruses, including tymoviruses (Weiland and Dreher, 1989), tombusviruses (Johnston and Rochon, 1990), maize chlorotic mottle virus (Nutter et al., 1989), and other luteoviruses (Veidt et al., 1988; Tacke et al., 1990) also contain overlapping reading frames that are translated by initiation at the first two AUGs of the mRNA. In addition, the 5' proximal AUG is bypassed in favor of a downstream AUG in a more optimal context during translation initiation on RNAs of plum pox virus (Riechmann et al., 1991), maize retroid element Bsl (Jin and Bennetzen, 1989), and some artificial constructs (Hensgens et al., 1992). In all known or suspected cases of initiation at the first two AUGs of a plant viral RNA, the second (5' distal) AUG is in a "better" context (purine at position -3, and/or guanine at +4) than the first AUG (Dinesh-Kumar et al., 1992). This observation supports the notion that leaky scanning occurs. However, the effect of sequence context on start codon choice has not been rigorously examined in plants.

In addition to primary structure, several examples exist in which secondary structure 5' of, or including, an AUG inhibits initiation (Kozak, 1989b; Grens and Scheffler, 1990; Fu et al., 1991; Liebhaber et al., 1992). Computer analyses predict that the first (CP) AUG of BYDV-PAV sgRNA1 is sequestered in a stem-loop, whereas the second (17K) AUG is in a single-stranded region (Dinesh-Kumar et al., 1992). This observation led us to propose that the secondary structure may inhibit translation initiation at the CP AUG relative to that at the start of the 17K ORF. In this study, we provide evidence for the secondary structure of the 5' end of sgRNA1 and
its effect on gene expression. We also performed systematic mutagenesis of the -3 and +4 positions flanking both the CP and 17K AUGs and altered the AUG codons themselves. The effects of these mutations on initiation at each AUG were examined in rabbit reticulocyte lysates (RRL) and by transient expression in oat protoplasts of the *Escherichia coli uidA (GUS)* gene in frame with either the CP or 17K AUG. Thus, we provide simultaneous comparisons of the effects of primary and secondary structure contexts on translation initiation at competing AUGs on an mRNA.
RESULTS

Figure 1B shows the map of previously constructed plasmid pSP17 from which the mRNA encoding the CP and 17K ORFs can be transcribed (Dinesh-Kumar et al., 1992). The resulting transcript has the precise 5' end of the actual sgRNA1 that was mapped to base 2769 of the BYDV-PAV genomic RNA. Upon translation in a RRL, both the CP and 17K ORF products accumulated (Dinesh-Kumar et al., 1992).

Effects of mutations near the CP start codon on secondary structure

We altered the bases at positions -3 and +4 because those have been shown to have the most influence on start codon efficiency in vertebrates (reviewed by Kozak, 1991), and a purine at position -3 and a guanine at +4 are the most common bases flanking plant start codons (Lutcke, et al., 1987; Cavener and Ray, 1991). The base changes are shown in Figure 1C. We refer to the presence of an A residue at position -3 and a G at +4 as the optimal context, and U and A at these positions, as is the case with the CP start codon, as the suboptimal context. In mutants M1, M1/2, M3, and M8 the bases around the CP AUG were changed from suboptimal to optimal (bases 87 and 93, respectively, when numbered from the 5' end of sgRNA1). With the exception of M1/2, these changes are also expected to disrupt base pairing around the CP AUG, based on the structure shown in Figure 2. This structure was predicted by the programs STAR (Abrahams et al., 1990) and RNASE (Cedergren et al., 1988) on the full length pSP17
Figure 1. Organization of BYDV-PAV sgRNA1 and sequences flanking its first two AUG codons.

(A) Genome organization of BYDV-PAV RNA. Locations of unusual translational events (Brault and Miller, 1992; Dinesh-Kumar et al., 1992) are shown. Subgenomic RNAs (sgRNA1 and sgRNA2) detected in infected cells but not virions have been mapped previously (Dinesh-Kumar et al., 1992).

(B) Map of pSP17 (Dinesh-Kumar et al., 1992). Transcription initiates at the sgRNA1 start site immediately adjacent to the T7 polymerase promoter. Bases are numbered from the 5' end of sgRNA1.

(C) Mutations flanking CP and 17K ORF start codons. Mutations are underlined. pSP17 is the wild-type in vitro transcription plasmid. Mutations at bases U65 and A71 in M2 and M1/2 are shown in Figure 2.
Figure 2. Nuclease cleavage sites on the predicted secondary structures of the 5' end of sgRNA1.

Nuclease-sensitive sites from Figure 3 were superimposed on computer-predicted secondary structures of the 5' ends of mutant and wild-type sgRNA1. Arrows identify sites cleaved by nuclease T1 or T2 (single-stranded). Filled circles show V1 cleavage sites (double-stranded). The size of the symbol is proportional to the frequency of cleavage (intensity of band in Figure 3). Only the portions of the mutant transcripts that gave different folding patterns from the wild-type are shown. Bases 1-63 and >98 in all mutants gave patterns identical to the wild-type.
transcript. Thus, differences in translational efficiency in mutant M1 (and M3 and M8) could be due to the effect on immediate sequence context and/or the predicted disruption of the base-paired regions that flank the CP AUG. To distinguish between these two possibilities, a second mutant, M2, was constructed with nucleotides 65 and 71 changed to C and U, respectively. These changes are on the opposite side of the stem from those in M1, thereby causing a similar weakening of secondary structure as in M1, but leaving the wild-type bases at -3 and +4 around the CP AUG. Mutant M1/2, which combines all the mutations in M1 and M2, would have the mutations at positions -3 and +4 as in M1 but is expected to have the wild-type secondary structure. Thus, if the alterations to secondary structure occur as predicted, differences from wild-type start codon efficiency in M1 would be due to primary and/or secondary structure, in M2 it would be due to secondary structure only, and in M1/2 it would be due to primary structure only.

To determine whether the wild-type and mutant transcripts folded as predicted, we treated end-labeled transcripts with structure-sensitive nucleases. The wild-type transcript and mutants M1, M2, and M1/2 were digested partially with single strand-specific ribonucleases T1 (cuts after guanosine residues) and T2 (cuts after A→U under our conditions) as well as with nuclease V1 which preferentially cuts some but not all phosphodiester bonds of double-stranded regions. To ensure that the structures "seen" by the nucleases were the same as those "seen" by ribosomes, transcripts were digested under the same salt conditions used in the in vitro translation assay.
(Schulz and Reznikoff, 1990). Comparison of the digestion products with those obtained by nuclease digestion under denaturing conditions, as shown in Figure 3, revealed wide variation in sensitivity of phosphodiester bonds to nuclease. Many bases were not cut by any enzymes, presumably due to steric hindrance by the folded RNA. In Figure 2 the nuclease-sensitive sites are plotted on the predicted structures of the various constructs, indicating that the nuclease-sensitivity correlated closely with the predicted structures.

In all transcripts, bases 77-81 were extremely sensitive to single strand-specific nucleases, consistent with their being in a loop. Other predicted single stranded regions also showed sensitivity to nucleases T1 and T2. All bases cut by V1 are in predicted helical regions except for bases 38, 134, and 137, which were cut weakly. The clearest structural differences were reflected in the digestion patterns of G residues by nuclease T1. Bases G68, G92, and G93 became more sensitive to nuclease T1 in M1, as predicted (arrows, Figure 3). Although G92 and G93 are predicted to be in a helix, it is weaker in M1 ($\Delta G = -1.8$ kcal/mol) than wild-type ($\Delta G = -3.2$ kcal/mol) because of the U65-G93 base-pair substitution for the U65-A93 base pair. Therefore, both G residues in M1 are more accessible to T1 nuclease, suggesting that the helix "breathes" or does not exist. In mutant M2, a somewhat greater change is predicted. When compared to the wild-type, this structural change is reflected in greater sensitivity of bases G68 and G92 to T1 (arrows, Figures 2 and 3). Importantly, G92 (the G residue of the CP AUG) was more T1 sensitive in M1 and M2 than in the wild-type or M1/2. In the regions outside of the predicted changes, M1 and M2 gave
Figure 3. Products of partial nuclease digestion of transcripts of wild-type (SP17) and mutants M1, M2 and M1/2 sgRNA1 sequences. After digestion with no nuclease (lanes N), or nuclease T1 (lanes T1), T2 (lanes T2), V1 (lanes V1) or fM (lanes f), or with mild alkali (lanes L), products were run on 8% polyacrylamide, 7M urea gels (see Methods). Enzyme digestions were in ionic conditions used for in vitro translation except in lanes marked SEQ, which under denaturing conditions. The nucleotide sequence of the readable portion of the sgRNA1 transcript is shown at left of SP17 gel. Relevant sequences of mutant transcripts are at right. Prominant bands in the mutants that are absent in SP17 and M1/2 are marked at left of M1 and M2 gels. Bottom portions of gels of mutant transcript digestions had patterns identical to the wild-type (data not shown).
nuclease digestion patterns identical to that of the wild-type transcript. This shows that the mutations introduced no unexpected, long-distance structural changes. The double mutant, M1/2 behaved in a manner identical to the wild-type, which was exactly as predicted. Thus, the structural data obtained on the above mutants agree well with the computer predictions, and we expect that M3 and M8 would also disrupt the secondary structure like M1. Possible alterations to secondary structure by mutations around the 17K AUG cannot be ruled out, as evidenced by the unexpected V1 sensitivity of two bases in and near this codon.

Effects of mutations on translation of CP and 17K ORFs in reticulocyte lysates

With the secondary structures around the 5' ends of mutant sgRNAs established, we first tested the effects of the base changes on translation in vitro. To facilitate accurate comparisons between reactions, equal amounts of completely intact RNA, shown in Figure 4A, were used as template in all RRL assays. The amount of template (200 ng) should ensure that the mRNA and not some undefined factor was rate limiting in the translation mix. Because the presence of a 5' cap on the transcripts had little effect on translation product amounts or ratios (Dinesh-Kumar et al., 1992), uncapped transcripts were used. The magnesium and potassium concentrations were optimal as determined previously (Dinesh-Kumar et al., 1992). The translation products of the transcripts containing all the mutations shown in Figure 1C are shown in Figure 4B. Histograms of abundance of each band,
Figure 4. In vitro translation of in vitro transcripts encoding the CP and AUG ORFs.
(A) Ethidium bromide-stained 1% agarose gel of T7 polymerase transcripts obtained from the indicated mutants used for translation in RRL.
(B) Fluorograph of RRL translation products of transcripts shown in (A), following electrophoresis on a 12% polyacrylamide-SDS gel. Mobilities of CP and 17K ORF products are indicated at right.
(C) Relative radioactivity measured in bands on gel in (B) by Phosphorimager and normalized for methionine content. WT, wild-type.
determined by direct measurement of radioactivity with a Phosphorimager, as well as ratios of CP-to-17K product, are shown in Figure 4C. Two additional repetitions of this experiment gave similar results (data not shown). Alteration of the CP AUG context to optimal resulted in increased ratio of CP to 17K products (mutants M1, M1/2, M3, and M8). Mutant M2, in which both AUGs were in wild-type contexts but which weakened base pairing around the CP AUG, gave a similar increase in CP, but an insignificant decrease in 17K protein production relative to the wild-type. In the converse experiment, with the wild-type secondary structure restored but the CP AUG in the optimal context, M1/2 gave the same drastic reduction in 17K product as did M1 alone but gave only wild-type levels of CP. These results suggest that a greater proportion of scanning ribosomes initiated protein synthesis at the first (CP) AUG when it was in a good context, regardless of secondary structure, but that disruption of the secondary structure increased initiation of CP translation regardless of sequence context.

With both AUGs in a suboptimal context (M4), translation of both ORFs was halved, maintaining the wild-type ratios of CP/17K products. Changing the CP AUG to ACG eliminated initiation at this codon and increased initiation at the 17K AUG (M5). However, combining this mutation with placement of the 17K AUG in a suboptimal context gave no translation products at all (M6). Changing the 17K AUG to ACG, while leaving it and the CP AUG in native contexts (M7), reduced translation of the 17K ORF and the CP ORF by 50%. This mutation, combined with placing the CP AUG in an optimal context (M8), gave higher levels of CP and very little 17K.
Finally, to test if 17K initiation could occur at a more extreme deviation from AUG, it was changed to GCG (M9) at which little or no initiation occurred.

Effects of mutations on start codon choice in protoplasts

To determine the control of start codon choice in plant cells, the sgRNA1 5' leader, from bases 1 to 170 or 172, was inserted in the *E. coli uidA* (*GUS*) gene with either the CP or the 17K AUG in frame with the coding region of the *GUS* gene, as shown in Figure 5. Expression was driven by a maize alcohol dehydrogenase (*Adh1*) promoter and first intron (Callis et al., 1987), or by a duplicated cauliflower mosaic virus (CaMV) 35S promoter (Kay et al., 1987), resulting in transcripts containing 141 or 6, respectively, non-viral bases 5' of the start of the sgRNA1 leader. For each mutation, two constructs were made, one with the CP AUG in frame with *GUS*, the other with the 17K AUG in frame. They differ by two bases at the virus-*GUS* sequence fusion that are absent in the 17K constructs (Figure 5B). To verify reproducibility of results, the constructs were tested in duplicate and in more than one experiment, and the insertless vector was used as a positive control for transformation and *GUS* expression in all experiments, as shown in Table 1.

When the wild-type sequence was inserted in pAdhGUS, about half as much b-glucuronidase (*GUS*) activity was detected in cells containing pCPWT, which has the CP AUG in frame with *GUS*, as in cells containing pCP17K, which has with the 17K AUG in frame with *GUS* (Table 1). These constructs gave 63- and 27-fold less GUS activity, respectively, than
Figure 5. Vectors for expression of GUS from sgRNA1 start codons.

(A) Maps of plant expression vectors into which the 5' end of sgRNA1 (panel B) was inserted.

Black bars (GUS) represent the GUS ORF. 35S and NOS boxes indicate tandemly repeated CaMV 35S promoter and nos transcription termination sites, respectively. Adh1 P and Adh1 I indicate the Adh1 promoter and first intron, respectively. Arrows below the maps show transcription start sites. The Adh1 start codon (ATG) of pAT13 was changed to AGG in pAdhGUS. The multiple cloning site from pAGUS1 (Skuzeski et al., 1990) that was inserted into pAT13 contains BamHI (B), Sall (S), Ncol (N), HindIII (H) and Apal (A) sites.

(B) Sequence of 5' end of sgRNA1 inserted in GUS expression vectors between BamHI and Apal sites in pAGUS1 or pAdhGUS.

Vector-derived bases are in italics. Amino acid sequences of the portions of CP and 17K ORFs inserted in the vectors are shown, as are the first four amino acids of the GUS ORF (long arrow at bottom). The sequence is shown with CP AUG in frame with the GUS coding region (pCPWT or p35CPWT). The viral bases 171-172 (GU) are absent (D) in plasmids with the 17K AUG in frame with GUS (e.g., p17KWT or p3517KWT).
A

<table>
<thead>
<tr>
<th>Frameshift</th>
<th>Readthrough</th>
</tr>
</thead>
<tbody>
<tr>
<td>39K</td>
<td>22K/50K</td>
</tr>
<tr>
<td>6.7K</td>
<td></td>
</tr>
</tbody>
</table>

B

- pSP17
- sgRNA1 start
- PstI HindIII
- CP
- 17K

C

| SP17: GUGAAUGAA-34 nt-CAAAUGGC |
| M1: GAGAAUGA-------CAAAUGGC    |
| M2*: GAGAAUGA-------CAAAUGGC   |
| M1/2*: GAGAAUGA-----CAAAUGGC   |
| M3: GAGAAUGA-------CUAAUGAC    |
| M4: GUGAAUGA--------CUAAUGAC   |
| M5: GUGAACGAA-------CAAAUGGC   |
| M6: GUGAACGAA-------CUAAUGAC   |
| M7: GUGAAUGA--------CAAAACGGC  |
| M8: GAGAAUGA--------CAAAACGGC  |
| M9: GUGAAUGA--------CAAAACGGC  |

* Mutations not shown: U65→C and A71→U
Table 1. Expression of sgRNA1-GUS fusions in protoplasts using the Adh1 promoter.

<table>
<thead>
<tr>
<th>CP AUG</th>
<th>17K AUG</th>
<th>Plasmid</th>
<th>GUS units&lt;sup&gt;a&lt;/sup&gt;</th>
<th>CP/17K</th>
<th>GUS units</th>
<th>CP/17K</th>
<th>GUS units</th>
<th>CP/17K</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Experiment 1</td>
<td></td>
<td>Experiment 2</td>
<td></td>
<td>Experiment 3</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>GUS units&lt;sup&gt;b&lt;/sup&gt;</td>
<td></td>
<td>GUS units&lt;sup&gt;b&lt;/sup&gt;</td>
<td></td>
<td>GUS units&lt;sup&gt;b&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Salmon Sperm</td>
<td></td>
<td>5</td>
<td></td>
<td>11</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>pADHGUS</td>
<td></td>
<td>13304</td>
<td></td>
<td>16050</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>p17KWT</td>
<td></td>
<td>486</td>
<td></td>
<td>458</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>pCPWT</td>
<td></td>
<td>212</td>
<td></td>
<td>209</td>
<td></td>
</tr>
<tr>
<td>UGAAUGA</td>
<td>AAAAUGG</td>
<td>pCPM1</td>
<td>1813</td>
<td>40</td>
<td>45.00</td>
<td>32</td>
<td>56.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>p17KM1</td>
<td>739</td>
<td>0.43</td>
<td>786</td>
<td>32</td>
<td>0.50</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>pCPM2</td>
<td>1660</td>
<td>0.45</td>
<td>1563</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AGAAUGG</td>
<td>AAAAUGG</td>
<td>pCPM1/2</td>
<td>680</td>
<td>23</td>
<td>29.57</td>
<td>23</td>
<td>25.57</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>p17KM1/2</td>
<td>1151</td>
<td>12</td>
<td>92.82</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AGAAUGG</td>
<td>UAAAAUGG</td>
<td>pCPM3</td>
<td>1151</td>
<td>12</td>
<td>92.82</td>
<td></td>
<td>1004</td>
<td>16</td>
</tr>
<tr>
<td>AGAAUGG</td>
<td>UAAAAUGG</td>
<td>pCPM4</td>
<td>52</td>
<td>92</td>
<td>0.57</td>
<td>45</td>
<td>0.52</td>
<td></td>
</tr>
<tr>
<td>AGAAUGG</td>
<td>UAAAAUGG</td>
<td>pCPM5</td>
<td>8</td>
<td>906</td>
<td>0.01</td>
<td>11</td>
<td>0.01</td>
<td>819</td>
</tr>
<tr>
<td>AGAAUGG</td>
<td>UAAAAUGG</td>
<td>pCPM6</td>
<td>18</td>
<td>128</td>
<td>0.14</td>
<td>14</td>
<td>0.14</td>
<td>105</td>
</tr>
<tr>
<td>AGAAUGG</td>
<td>UAAAAUGG</td>
<td>pCPM7</td>
<td>32</td>
<td>9</td>
<td>3.49</td>
<td>40</td>
<td>1.50</td>
<td></td>
</tr>
<tr>
<td>AGAAUGG</td>
<td>UAAAAUGG</td>
<td>pCPM8</td>
<td>457</td>
<td>13</td>
<td>34.65</td>
<td>539</td>
<td>26.79</td>
<td></td>
</tr>
<tr>
<td>AGAAUGG</td>
<td>UAAAAUGG</td>
<td>pCPM9</td>
<td>34</td>
<td>7</td>
<td>5.10</td>
<td>48</td>
<td>9.41</td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup>CP or 17K indicates which AUG is in-frame with the GUS coding region. Plasmid maps and sequences of sgRNA1 inserts are in Figure 5. All plasmids have the Adh1 promoter and first intron.

<sup>b</sup>Gus units are defined as nanomoles of 4-methyl umbelliferone produced per min per mg protein. All samples except Salmon sperm have had the GUS units obtained in the presence of salmon sperm DNA in that experiment subtracted. All GUS units represent averages from two identically-treated samples.

<sup>c</sup>Bases at positions -20 and -26 relative to the CP AUG are changed as in Fig. 1C.
pAdhGUS that lacked the viral insert, indicating considerable inhibition of GUS expression by the viral insertion. Similar CP/17K AUG initiation ratios (0.4 to 0.5) were obtained whether expression was driven by the Adh1 promoter and first intron or the CaMV 35S promoter, although the latter was inhibited less by the insertion of the viral sequence (Table 2). The constant CP/17K ratio, regardless of promoter, indicates that the length of the leader upstream of the viral sequence has little effect on start codon choice.

When the CP AUG was in an optimal context and in frame with GUS (pCPM1, p35CPM1, pCPM1/2, p35CPM1/2, pCPM3 and pCPM8, Tables 1 and 2), much higher levels of GUS were obtained than with the wild-type, suboptimal context. Initiation at the CP AUG in mutants M1, M2 and M1/2 was five- to ninefold greater than with the wild-type, unlike the RRL results in which a one- to twofold increase was observed. As in RRLs, when the 17K AUG was in frame with GUS in constructs containing the M1 or M1/2 mutations, the amount of GUS was reduced drastically. The three- to fourfold increase in expression from the CP or 17K AUGs in the M2 mutants shows that relaxation of secondary structure around the CP AUG, with both AUGs in their native contexts, increased expression of GUS initiated at either start codon. The CP/17K AUG initiation ratio was the same as wild-type.

Because the effects of mutations on start codon efficiency were similar with both promoters, effects of further mutations were studied in only one vector (pAdhGUS).

When the CP AUG was in an optimal context, and the 17K was suboptimal (the opposite of the wild-type situation: mutant M3), initiation
Table 2. Expression of sgRNA1-\textit{GUS} fusions in protoplasts using the CaMV 35S promoter.

<table>
<thead>
<tr>
<th>Sequence context</th>
<th>CP AUG</th>
<th>17K AUG</th>
<th>Plasmid(^a)</th>
<th>GUS units(^b)</th>
<th>CP/17K</th>
</tr>
</thead>
<tbody>
<tr>
<td>UGAAUGA AAAAUGG</td>
<td>17K</td>
<td>p35CPWT</td>
<td>p3517KWT</td>
<td>179</td>
<td>0.51</td>
</tr>
<tr>
<td>ΔGAUGG AAAAAUGG</td>
<td>17K</td>
<td>p35CPM1</td>
<td>p3517KM1</td>
<td>875</td>
<td>94.59</td>
</tr>
<tr>
<td>UGAAUGA(^c)</td>
<td>17K</td>
<td>p35CPM2</td>
<td>p3517KM2</td>
<td>786</td>
<td>0.35</td>
</tr>
<tr>
<td>ΔGAUGG(^c)</td>
<td>17K</td>
<td>p35CPM1/2</td>
<td>p3517KM1/2</td>
<td>581</td>
<td>36.17</td>
</tr>
</tbody>
</table>

\(^a\)CP or 17K indicates which AUG is in frame with the \textit{GUS} coding region. Plasmid maps and sequences of sgRNA1 inserts are in Figure 5. The 35 indicates that viral sequence was inserted in pAGUS which contains a 35S promoter.

\(^b\)GUS units are defined as nanomoles of 4-methylumbelliferone produced per min per mg of protein. All samples except salmon sperm have had the GUS units obtained in the presence of salmon sperm DNA in that experiment subtracted. All GUS units represent averages from two identically treated samples.

\(^c\)Bases at positions -20 and -26 relative to the CP AUG are changed as in Fig. 1C.
from the 17K AUG was negligible (Table 1), while the CP AUG was fivefold more efficient than it was in the wild-type. With both AUGs in a suboptimal context (M4), the ratios of CP/17K AUG initiation were unchanged from the wild-type and overall translation from both AUGs was reduced four- to fivefold. Changing the CP AUG to ACG eliminated initiation at that site and doubled levels of GUS activity initiating from the 17K AUG (M5). Combining the CP AUG-to-ACG mutation with placing the 17K AUG in a suboptimal context (M6) reduced initiation at the 17K AUG to one-eighth the level obtained in with M5. There was little or no initiation at GCG (M9). All the above results agree with those obtained in vitro, but the magnitudes of the effects of mutations were greater in vivo. Protoplast expression differed from RRLs in two other ways: initiation at ACG failed to occur even in the optimal context of the 17K ORF (M7), and (in some cases) the effects of mutations in the 17K AUG context on initiation at the CP AUG differed (discussed below).

**Efficient initiation at the 17K AUG is needed for maximum initiation at the CP AUG**

In all in vivo assays and in some in vitro assays, any mutation that reduced translation initiation at the 17K start site also reduced initiation of translation at the CP AUG. For example, constructs CPWT, CPM4, CPM7, and CPM9 all have identical (wild-type) contexts and predicted secondary structure around the CP start codon, differing only in mutations in or flanking the 17K AUG that reduce initiation at that site. Yet, in all these mutants,
translation initiated at the CP start codon four to sixfold less efficiently than wild-type in vivo, and twofold less efficiently in vitro for all but M9. Similarly, in CPM1, CPM3 and CPM8, the CP AUG is in the optimal context (with corresponding disruption in secondary structure), yet GUS expression when initiated by the CP AUG is substantially lower in M3 and M8, than M1 in vivo. A model to explain these observations is discussed below.
Differences between cell-free and protoplast assay systems.

RRL and transient expression assays each have certain advantages. In RRLs, the full-length CP and 17K genes are translated from a single message and only translational events are observed, but the method suffers from the risk of artifacts induced by salt conditions or loss of important factors during isolation of the cell-free system (Kozak, 1989c). The wild-type CP/17K ratio varies significantly with salt conditions (Dinesh-Kumar et al., 1992). The conditions used here were defined previously as optimal because they permitted efficient readthrough of the CP stop codon as expected, and they minimized initiation at other internal AUGs. Because it may not reflect the natural conditions, we concluded that the in vitro system shows the general trends of the effects of mutations on translation, and what may happen in vivo, but may not be quantitatively accurate.

The results of two AUGs competing on a single mRNA are being observed simultaneously in vitro, but initiation at only one at a time is detected in vivo. Presumably, initiation occurs at the out-of-frame AUGs in vivo, resulting in truncated products. In protoplasts, changes in GUS activity due to mutagenesis could result from changes in gene expression at levels other than translation. However, transcriptional effects are unlikely due to the similarity of results obtained with different promoters. Other post-transcriptional variations are unlikely due to the consistency with in vitro results in most constructs.
Although only a portion of the viral sequence is present in vivo, this serves as an advantage for studying initiation because the same protein (GUS) is being translated whether the CP or 17K AUG is in frame with GUS. Because differential codon usage in the two ORFs can affect elongation rates of the overlapping ORFs (Fajardo and Shatkin, 1990), these effects would be seen in the cell-free translation of the in vitro transcript but not in GUS expression in vivo. This may explain why the 17K AUG appears to function more efficiently than the CP AUG in vivo but equal to the CP AUG in vitro; i.e., less efficient elongation in the 17K ORF than the CP ORF would reduce relative levels of 17K product in the RRL. Alternatively, the CP derived amino acids that are fused to GUS may reduce GUS activity more than those derived from the 17K ORF.

The fact that the fold increases in initiation at the CP AUG are greater in vivo than in vitro may reflect differences in rate limiting steps, with initiation being rate limiting in vivo and elongation being more limiting in RRLs. These differences are similar to those observed by Kozak (1989b), who showed that effects of alterations to bases flanking start codons were greater in a transient assay in vertebrate cells than in RRL and by Gallie et al. (1987) and Sleat and Wilson (1992) who found greater enhancement of translation by the tobacco mosaic virus (TMV) W sequence in transient expression assays than in RRL. In summary, the in vivo system probably reflects the control of initiation more accurately than RRL (Kozak, 1989c).
Effect of secondary structure

The nuclease sensitivity results were consistent with the predicted secondary structure of the 5' end of sgRNA1. The fact that the mutations altered the structure only where predicted was important, because unpredicted changes to structure can occur (e.g. Miller and Silver, 1991). Slight weakening of the base pairing without altering the bases immediately flanking the CP AUG (mutant M2) increased translation initiation at that AUG. This agrees with previous observations that stem loops preceding or encompassing a start codon reduces translation efficiency (Kozak, 1988; Rao et al., 1988; Rouault et al., 1988; Yager and Coen, 1988; Liebhaber et al., 1992). In those cases, the stem loops were more stable than those on sgRNA1, but de Smit and van Duin (1990) showed that increasing stabilities of stem loops by as little as 1.4 kcal/mol could drastically decrease initiation of translation in E. coli. In plants, 5' leaders that give highly efficient translation, have little or no secondary structure (Jobling and Gehrke, 1987; Sleat and Wilson, 1992). Thus, the stem loop at the very 5' end of sgRNA1 would be expected to result in a poor template. Consistent with this, insertion of the viral leader in pAdhGUS and pAGUS reduced GUS expression (initiated at the CP AUG) substantially. However, the high-expressing mutant p3517KM2 reduced translation only 1.6-fold relative to the insertless control, even though it has only seven single-stranded bases at the 5' end of the mRNA, prior to the first strong stem-loop. The smaller reduction in GUS activity when driven by the 35S promoter compared to the Adh1 promoter differs from observations of Kozak (1989b), who showed that the inhibition
of chloramphenicol acetyltransferase gene expression by an artificial stem loop ($\Delta G = -30$ kcal/mol) was much greater when it was located 12 bases from the 5' end compared to 54 bases away. This may be due to the much more stable stem loop used in that study. The effect of disrupting the 5' proximal stem loop of sgRNA1 remains to be tested.

Regulation of expression of the CP and 17K ORFs by sequestering of the first AUG in secondary structure does not seem to be a general rule for luteoviral subgenomic RNAs, because such structures are not conserved among luteoviruses (W.A. Miller, unpublished observation). However, such sequestering has also been proposed for the first AUG of kennedy yellow mosaic tymovirus RNA (Ding et al., 1990) in which initiation likely occurs at the first two AUGs. The apparent resemblance of the sgRNA1 secondary structure encompassing the CP AUG to a tRNA cloverleaf is of unknown significance. tRNA-like structures serve as origins of replication at the 3' termini of RNAs of viruses in several groups (e.g. Miller et al., 1986). A tRNA-like structure at the 5' end of the *E. coli* threonyl tRNA synthetase mRNA serves as a signal for feedback inhibition of its own expression (Moine et al., 1990). Because the mRNA mimics tRNA$^{Thr}$, the synthetase binds its own mRNA, shutting off translation.

**Roles of bases at positions -3 and +4**

A purine exists at position -3 in 87%, and a guanine at +4 in 70% of known plant start codons (Cavener and Ray, 1991). The positive effect of A and G at these positions has been demonstrated in transgenic plants (Taylor
et al., 1987). Alteration of bases -1 and -2 (McElroy et al., 1991) or -2, -4, and -5 (Guerineau et al., 1992) in the presence of an A at -3 and a G at +4 had little effect on translation in plants. In the case of BYDV-PAV sgRNA1, the bases at -3 and +4 are much more important than secondary structure in start codon selection. Placing the CP AUG in the optimal context greatly increased the proportion of initiation events at that AUG, regardless of structure. Relative proximity to the 5' end is also important, as significant initiation occurred at the 5' proximal (CP) AUG in the suboptimal context, but only negligibly at the 5' distal 17K AUG in the suboptimal context. These results agree with numerous experiments by Kozak, and on other viral RNAs such as reovirus s1 RNA which, like BYDV sgRNA1, has two out-of-frame ORFs initiating at the first two AUGs of the RNA (Munemitsu and Samuel, 1988).

Initiation at ACG

Initiation at ACG on M7 RNA in vitro is consistent with other observations in animal (Mehdi et al., 1990; Peabody, 1987) and plant (Schultze et al., 1990) cells. In all examples, initiation at an ACG requires an optimal sequence context and occurs more efficiently in vitro than in vivo. We know of no reports of initiation at GCG.

Ribosome pausing model

One unforeseen result was that for a given CP AUG context, the level of initiation at this AUG was reduced if the 17K ORF AUG was in a poor
context or eliminated. This reduction is not due to alterations in the CP-GUS fusion, because the single base change: AUG to ACG that eliminated the 17K start codon, did not alter the amino acid sequence of the CP ORF. In the CP ORF, it resulted in a change from AAU to AAC, both of which encode asparagine. Thus we concluded that it is initiation at the CP AUG that is being reduced by negative mutations in and around the 17K AUG. We are unaware of a precedent for initiation at a downstream AUG positively affecting initiation at an upstream AUG.

To explain this phenomenon, we propose a model, diagrammed in Figure 6, in which ribosomes pausing at the second AUG enhance initiation at the first AUG. According to this model, the first 40S subunit often scans past the CP AUG until it reaches the 17K start codon, at which the 60S subunit binds and protein synthesis proceeds. The CP AUG would be bypassed for two reasons, first it is in a suboptimal context, second the tendency of the 5' end of sgRNA1 to form secondary structure (which is known to reduce initiation: Kozak, 1989b; Grens and Scheffler, 1990; Fu et al., 1991; Liebhaber et al., 1992), may out-compete the unwinding activity of the 40S subunit and associated factors (eIFs 4A, 4B and 4F; Jaramillo, et al., 1991), causing the 40S subunit to dissociate from the sgRNA. Hence, mutations that reduce secondary structure permit more initiation at the CP AUG. Formation of the 80S complex, or simply pausing of the 40S subunit at the 17K AUG on wild-type sgRNA1, would be expected to melt some of the base pairing upstream of the 17K AUG. Fourteen to 20 bases on either side of the AUG are melted by the 80S complex and are inaccessible to
Figure 6. Model for enhancement of translation initiation at the CP AUG by pausing of the ribosome at the 17K AUG. sgRNA1 leader structure (Figure 2) is shown as a line with the CP and 17K AUGs in their relative positions.

(A) The 40S subunit binds the 5' end of the RNA and scans toward the CP AUG melting secondary structure as it goes.

(B) When the 40S subunit reaches the CP AUG it either (1) dissociates from the RNA due to competing secondary structure (arrow to left), (2) continues scanning due to poor context (arrow to right), or (3) is bound by the 60S subunit and CP synthesis begins (not shown).

(C) When a 40S subunit reaches the 17K AUG, it pauses as the 60S subunit binds and the first peptide bond is formed. This keeps the adjacent upstream structure melted, giving the trailing 40S subunit better access to the CP AUG, and momentarily prevents it from continued scanning (striped arrow and vertical bars).

(D) The stalled 40S subunit at the CP AUG increases time allowed for it to be bound by the 60S subunit, after which protein synthesis proceeds.
A

B

C

D
nuclease, during stable monosome formation (Wolin and Walter, 1988; Kozak, 1989c; Liebhaber et al., 1992). Thus, at least some of the base pairing in the vicinity of the CP AUG would be melted, making it more accessible to the next 40S subunit.

Scanning ribosomes pause at start codons and trailing ribosomes can stack up behind a paused ribosome (Doohan and Samuel, 1992; Wolin and Walter, 1988). The first 40S subunit behind the paused 80S complex on the 17K AUG would be located near the CP AUG (43 bases upstream). This slowed-down 40S subunit would have more time to "discover" that it is resting near the CP AUG (which it would have bypassed more frequently, given the poor context) and initiate translation. This is analogous to Kozak's observation that a strong stem loop 14 bases downstream of an AUG, or even GUG or UUG, can increase initiation at that codon, presumably because of the pause in scanning induced by the stem loop (Kozak, 1990a). In support of our interpretation, Doohan and Samuel (1992) observed a pause site 25-30 nucleotides upstream of the second start codon of the dual ORF reovirus s1 RNA, in addition to the major pause site at the start codon. Pausing at both sites was eliminated when the second start codon was removed. Because the CP AUG is 43 bases upstream of the 17K AUG, perhaps the pausing by the 40S subunit at the 17K AUG, before binding of the 60S subunit, is responsible for increased initiation at the CP AUG because the 40S subunit alone spans about twice as many bases as the 80S subunit (Kozak and Shatkin, 1977; Doohan and Samuel, 1992). Our results lead us to predict that the presence of optimal flanking bases should increase
pausing at that AUG. To test this model, ribosomal pause assays need to be performed on the BYDV-PAV sgRNA1 mutant and wild-type transcripts.

Regulation of CP and 17K levels in infected cells

The results we have described do not necessarily reflect the ultimate regulation of viral protein levels in infected cells. The relative levels of CP and 17K products during infection are unknown for any luteovirus. In constructs similar to our p35CPWT and p3517KWT, Tacke et al. (1990) found that GUS synthesis initiated from the 17K AUG of potato leafroll luteovirus (PLRV) sevenfold more efficiently than from the PLRV CP AUG, in contrast to the twofold difference we observed. However, the translation efficiencies of the full genes and the stability of their products may differ. Trans-acting viral or virus-induced proteins could also regulate synthesis of these proteins. Because of its function, the CP would be expected to be highly expressed late in infection. Although the function of the 17K ORF is unknown, the homologous protein of PLRV has some properties of the cell-to-cell movement protein of TMV, including the ability to bind single-stranded nucleic acid nonspecifically and a tendency to aggregate (Tacke et al., 1991). Other evidence suggests that it is a 5' genome-linked protein (Keese and Gibbs, 1992). Elucidation of its function would allow better interpretation of the regulation of its expression. Even without this knowledge, our analysis of viral gene expression has allowed a better understanding of translation initiation in plants. From this we uncovered
unexpected interactions between start codons, resulting in a new, testable modification of current models of start codon efficiency in eukaryotes.
METHODS

Construction of mutant plasmids for in vitro studies

The construction of plasmid pSP17 (Figure 1B) has been described previously (Dinesh-Kumar et al., 1992). All mutations listed in Figure 1C were generated by the two-step polymerase chain reaction (PCR) mutagenesis method of Landt et al. (1990) using a mutagenic primer and M13 forward and reverse primers as flanking primers. The mutagenic primers and their templates are listed in Table 2. In all cases the PCR-amplified, 825-bp mutagenized fragment was gel purified, digested with XbaI-PstI, and cloned into XbaI-PstI-cut pUC118. All point mutations were confirmed by plasmid sequencing using a Taq Track sequencing kit (Promega).

Plasmids pAGUS1 and pAdhGUS

The plasmid pAdhGUS (Figure 5A), was constructed by inserting a multiple cloning site from pAGUS1 (a gift from J. Skuzeski, University of Nebraska-Lincoln; Skuzeski et al., 1990) into pAT13 (Figure 5A, a gift from M. Fromm, Plant Gene Expression Center, Albany, CA). pAdhGUS is similar to pAGUS1 except that it contains the promoter from a maize alcohol dehydrogenase (Adh1) gene followed by the first intron (Callis et al., 1987). The Adh1 initiation codon in pAT13 at base 100 of the Adh1 mRNA (Dennis et al., 1984; Nick et al., 1986) was changed to AGG in pAdhGUS by PCR mutagenesis using mutagenic primer SPDK7: 5' GGGCAAGGCGACCGCGG
Table 3. Primers and templates used for mutant construction.

<table>
<thead>
<tr>
<th>Primer name</th>
<th>Sequence^</th>
<th>Bases (numbered from 5' end of sgRNA1b)</th>
<th>Mutant generated (Figure 1C)</th>
<th>Template used</th>
</tr>
</thead>
<tbody>
<tr>
<td>SPDK1:</td>
<td>CTGAATCCATTCTCCACCTC</td>
<td>99-80</td>
<td>M1</td>
<td>pSP17</td>
</tr>
<tr>
<td>SPDK2:</td>
<td>CTAGTGAGTCTGGAACT</td>
<td>79-61</td>
<td>M2</td>
<td>pM7</td>
</tr>
<tr>
<td>SPDK3:</td>
<td>CTTGTGICTTATTAGATTGCGCGT</td>
<td>142-119</td>
<td>M1/2</td>
<td>pM1</td>
</tr>
<tr>
<td>SPDK4:</td>
<td>CTGAATTCTTCACCACC</td>
<td>99-82</td>
<td>M5</td>
<td>pSP17</td>
</tr>
<tr>
<td>SPDK5:</td>
<td>CTTGTCGGTTTTTGATTTGC</td>
<td>142-124</td>
<td>M7</td>
<td>pSP17</td>
</tr>
<tr>
<td>SPDK6:</td>
<td>CTTGTCGCCCTTTGATTTGC</td>
<td>142-124</td>
<td>M9</td>
<td>pSP17</td>
</tr>
</tbody>
</table>

^Sequence is in 5' to 3' direction with mismatches to the wild-type sequence underlined.

bAll primers are complementary to the viral coding sequence. Numbering is from the 5' end of sgRNA1 (base 2769 in viral genome).
3', which anneals to bases 95-112 from the transcription start site. This resulted in the ATG in the introduced Ncol site (with no insert) or any ATG introduced into the multiple cloning site (with concomitant removal of the Ncol site) encoding the first AUG on the GUS mRNA.

Construction of CP/17K-GUS fusions

The wild-type and mutant CP or 17K ORFs containing the first 170-172 bases of sgRNA1 were fused to GUS in pAdhGUS or pAGUS1 so that either the CP AUG or the 17K AUG was in frame with GUS. The CP AUG in-frame series of plasmids was generated by PCR using upstream primer SPDK8 (5'-CGGGATCCGATAGGGTTTATAGTTAGTA-3'), which contained a BamHI site (italics) and sgRNA1 bases 1-20, and downstream primer SPDK9 (3'-TCTTGTCAAGCCGGTCCCGGGAT-5') which contained sgRNA1 bases 156-172 and an Apal site (italics) with constructs used in in vitro studies as templates. The 17K AUG in-frame series was constructed using the upstream primer SPDK8 with the downstream primer SPDK10 (3'-TCTTGTCAAGCCGGTCCCGGGAT-5'). This primer differs from SPDK9 only in lacking two bases complementary to the end of the viral insert (nucleotides 171-172 of sgRNA1) adjacent to the Apal site (italics). In both cases the PCR-amplified fragments were digested with BamHI and Apal and cloned into BamHI-Apal-cut pAdhGUS or pAGUS1. All constructs were verified by sequencing the entire insert.
In vitro transcription and translation

RNA was transcribed from HindIII-linearized wild-type and mutant plasmids using T7 RNA polymerase as described previously (Dinesh-Kumar et al., 1992). Two hundred nanograms of intact transcripts were translated, and the products were analyzed electrophoretically as described previously (Laemmli, 1970; Dinesh-Kumar et al., 1992) using nuclease-treated rabbit reticulocyte lysate (RRL) (Promega) in the presence of 35S-labelled methionine. The concentrations of magnesium chloride and potassium acetate were adjusted to 1.9 and 159 mM, respectively. Relative amounts of translation products were quantified using a Molecular Dynamics Phosphorimager 400E (Sunnyvale, CA). The ratios of CP to 17K products were determined after adjusting for number of methionine residues in the products (4 in CP, 3 in the 17K ORF).

Secondary structure mapping

Unlabeled, sgRNA1 transcripts containing nucleotides 1-217 were transcribed from plasmids pWT, pM1, pM2 and pM1/2 after linearizing with Sail. These transcripts were 5' end-labeled with γ-32P-ATP and gel purified as described by Miller and Silver (1991). Partial digestions with 0.05 units of T1 (Gibco-BRL), 0.05 units of T2 (Gibco-BRL), or 0.05 units of V1 (Pharmacia) nucleases were performed in 5 µl reactions containing approximately 10,000 cpm of gel-purified, end-labeled RNA, 5 µg yeast tRNA, 50 mM Tris, pH 8.0, 1.9 mM MgCl2, 159 mM potassium acetate. Reactions were incubated for 10 min at 37°C and terminated by adding an
equal volume of 7M urea, 30mM EDTA, 0.05% bromophenol blue, and 0.05% xylene cyanol solution and freezing at -80°C. Products were electrophoresed on 8% polyacrylamide, 7M urea gels. For size markers, transcripts were sequenced under denaturing conditions using 1 unit of T1 or 0.5 units of fM (cuts A>U) nuclease, as described previously (Miller and Silver, 1991), or they were partially hydrolyzed with sodium carbonate.

Protoplast transformation and GUS assay

Protoplasts were isolated from *Avena sativa* (cv. Stout) suspension culture (cell line S226 obtained from H. Rines, U.S. Department of Agriculture Agricultural Research Service, University of Minnesota, St. Paul) using the procedure described previously (Dinesh-Kumar et al., 1992) as improved by Higgs and Colbert (1993). Approximately 1x10^6 protoplasts and exactly 80 mg of plasmid DNA or carrier DNA (salmon sperm) were mixed in electroporation buffer (Fromm, et al., 1987). Samples were electroporated by delivering one pulse at 450V and 500μF capacitance using a Bio-Rad Gene Pulser. Electroporated protoplasts were incubated in the dark at room temperature for 24 hr before collection for analysis. GUS activity was determined as described by Jefferson (1987) using a Hitachi F-2000 fluorescence spectrophotometer. Protein concentrations were determined using the Bio-Rad (Richmond, CA) Bradford Protein Assay kit. Each construct was assayed in duplicate at least in two separate experiments with different batches of protoplasts.
ACKNOWLEDGMENTS

The authors thank David Higgs for advice on electroporation of oat cells, Jim Skuzeski for providing pAGUS1, and Mike Fromm and the U.S. Department of Agriculture, Plant Gene Expression Center, Albany, CA for pAT13. This work was funded by USDA National Research Initiative Grant no. 91-37303-6424 and a McKnight Foundation Individual Investigator Research Award in Plant Biology (to WAM). This is paper no. J-15299 of the Iowa Agriculture and Home Economics Experiment Station. Project number 2936.
REFERENCES


GENERAL SUMMARY

Barley yellow dwarf virus (BYDV-PAV serotype) has a positive sense, 5.7 kb RNA genome encoding at least six open reading frames (ORFs). BYDV uses variety of unusual mechanisms for translation of its genes. These include frameshifting in the polymerase gene, translation of internal gene via subgenomic RNAs, leaky scanning at the AUGs of the overlapping coat protein (CP) and 17K ORFs, and readthrough of the CP stop codon.

In course of this research I have addressed several questions relating to the gene expression strategies of luteoviruses, with specific emphasis on BYDV-PAV. In the first part of my work showed that the BYDV uses subgenomic RNAs to express its internal genes. It generates two subgenomic RNAs in infected plants and protoplasts. I mapped the 5' end of the large subgenomic RNA (sgRNA1) to base 2769 using northern hybridization, primer extension and RNase protection methods. Using synthetic transcripts generated from in vitro transcription I first demonstrated that the sgRNA1 serves as a tricistronic message, to express CP, 17K and readthrough proteins in vitro. The CP and overlapping 17K ORFs are translated by initiation at two out-of-frame AUGs, and the 50K ORF by in-frame readthrough of the CP stop codon.

In the second part of my research, I used CP/17K overlapping reading frames to understand the role of sequences and structures on translation initiation process in plants. In all luteoviruses the 17K open reading frame is embedded with in the coat protein reading frame. The 17K start codon of
BYDV-PAV is 43 bases downstream of the CP start codon. The 17K ORF start codon is in the better sequence context compared to the CP start codon and the CP start codon is sequestered within the stem loop structure at the 5' end of the sgRNA1. I performed systematic mutagenesis to study the effect of start codon flanking bases and secondary structure on expression of coat protein and 17K ORF. Both \textit{in vitro} and \textit{in vivo}, changing the CP start codon context to the optimal resulted in increased initiation at the CP AUG compared to wildtype sequence. Disruption of secondary structure increased expression of both CP and 17K ORFs irrespective of the sequence context and the ratio of the two products was unaltered. These results indicated that the Kozak's leaky scanning mechanism operates also in plants. This is also the first study of simultaneous comparisons of the effects of primary and secondary structure contexts on translation initiation at competing AUGs on an mRNA.

Surprisingly any mutations that changed the 17K start codon or its flanking bases into suboptimal context greatly reduced initiation at the CP AUG irrespective of its context. To explain this unusual phenomenon we proposed a new model, in which pausing of ribosomes at the 17K AUG promotes initiation at the CP AUG (see Figure 6 of Paper II of this thesis for details). Recent studies on initiation of translation of the overlapping reading frames on reovirus s1 RNA support our model (Doohan and Samuel, 1993). Both \textit{in vitro} and \textit{in vivo}, ribosomes paused at the start codon and more pausing was observed at the start codon flanked by optimal bases than the AUG in the suboptimal context. Secondary pause sites between the two
start codons were observed only in case of overlapping reading frames, whereas in the monocistronic s4 message ribosomes were uniformly distributed. Application of these ribosome pause assays to luteoviral mRNAs will shed light on regulation of synthesis of CP and 17K ORF in luteovirus and on initiation of translation in plants in general.

The above observations of BYDV-PAV have implications on expression of CP in transgenic plants to achieve coat-protein mediated resistance against luteoviruses. In transgenic potato plants expressing an mRNA encoding the overlapping CP and 17K ORFs, coat protein expression was so low as to be barely detected if at all (Kawchuk et al., 1990; Tumer, et al., 1991; Barker 1992). In some cases, the plants were virus-resistant, but it is not clear whether the resistance was due to the CP or 17K products or both (Miller and Young, 1994). To assure expression of CP only, transgenic potato plants were engineered which expressed a mutant in which the 17K start codon was deleted (Tumer et al., 1991). Even with this construct, the level of expression of CP was low. Based on my results (paper II of this thesis), any change to the 17K start codon or its context has a direct effect on the expression of CP. Hence, we predict that changing the CP start codon context into optimal sequence while keeping the 17K start codon in its native (optimal) context is necessary to achieve maximum expression of coat protein in transgenic plants.

It is important to determine the efficiency of CP and 17K translation in the natural viral context. Such a study has not been done in any luteoviruses. I introduced all the mutations discussed in the paper II of this
thesis into full-length infectious clone of BYDV-PAV. I have also generated the antibody to 17K protein, which was expressed in E.coli. These tools will facilitate further understanding of expression of these proteins during infection.

I constructed a full-length infectious clone of BYDV-PAV. This clone is highly infectious compared to the one which is previously published (Young et al., 1991 and pers. communication). The full-length infectious clone was used to demonstrate the requirement of the 39K ORF stop codon at the 3' end of the shifty heptanucleotide. Several deletion mutants were generated to elucidate the function of different ORFs of BYDV-PAV genome. Current studies indicate that the ORF1, ORF2 and 1150 nucleotide at the 3' end of the genome or ORF6 are necessary for BYDV-PAV replication. I have created several mutants of BYDV-PAV genome, this will facilitate understanding replication of BYDV-PAV.

Sequences involved in the luteovirus group of readthrough signal has not been studied so far. In my recent experiments I was able to show readthrough of the CP stop codon in vivo, by inserting GUS reporter gene into the full-length BYDV-PAV genome. Such a strategy has been used in other plant viruses to study replication and movement. The constructs containing GUS will provide a guide for mutagenesis studies aimed at elucidating the mechanism of readthrough for the luteovirus class. In addition, these constructs can be used to study luteovirus movement in the plant cells. This will be interesting because luteoviruses are confined to phloem tissues in the infected plants. This can also be used as a vector to
express useful products. Such strategy has been demonstrated using TMV
readthrough signal to produce Angiotensin-I-converting enzyme (ACEI)
peptide, which is found in the tryptic hydrolysate of milk casein (Hamamoto
et. al., 1993).
REFERENCES


Kemper, B., and Stolarsky, L. (1977). Dependence on potassium concentration of the inhibition of the translation of messenger
ribonucleic acid by 7-methylguanosine 5'-phosphate. *Biochemistry*, 16, 5676-5680.


ACKNOWLEDGMENTS

I am indebted to Dr. W. Allen Miller, my major professor and mentor, for his continuous help, support and encouragement during the course of my graduate studies.

I thank my program of study committee members; Dr. John H. Hill, Dr. James T. Colbert, Dr. Eric Henderson and Dr. Robert Benbow for helpful discussion and advice. I appreciate Dr. Steven R. Rodermel, Dr. Susan L. Carpenter, Dr. John H. Hill and Dr. James T. Colbert, for critical reading of the manuscripts before submission to the journals. I thank Dr. Rich DeScenzo for advice and help with computers.

I would like to thank my colleagues and friends; Dr. Veronique Brault, Dr. Stanley L. Silver, Dr. Di Rong, Dr. Jayaram, Dr. Cindy Paul, Mohan, Shanping Wang, Lada Rasochova, Alan Eggenberger, Michael Omunyin, Amy Marrow, Randy Beckett, Jan Seibel, Holly Sipes, David Higgs, Vijay Thatiparthi and Manzoor Mohideen, for helpful discussions and collaborations.

I extend my appreciation to my parents, sisters (Swarna and Ranjini), brother (Prashantha), granny, uncles (Gururaj and Sthyanarayana), aunts (Shashikala and Ambuja), and my relatives for their patience and encouragement during this study.
APPENDIX I  IN VITRO SYNTHESIS OF FULL-LENGTH INFECTIOUS TRANSCRIPTS FROM A cDNA CLONE OF BARLEY YELLOW DWARF VIRUS
Production of infectious RNA transcripts from full-length cDNA clones has become an essential tool for molecular genetic studies of RNA viruses. This approach has been utilized for several RNA viruses including a range of plant viruses (Ahlquist et al., 1984; Dawson et al., 1986; Mesi et al., 1986; Allison et al., 1988; Domier et al., 1989; Hemenway et al., 1990; Rochon and Johnston, 1991; Weiland and Dreher, 1989), animal viruses (van der Werf et al., 1986; Rice et al., 1987) and bacteriophage Qb (Shaklee et al., 1988). This technique provides an unlimited, homogeneous supply of inoculum. In addition, it provides means to examine gene expression and functions of various viral gene products by modifying biologically active cDNA clones using site-directed mutagenesis.

Construction of infectious cDNA clone is very useful in luteoviruses, because they are phloem-limited and attain very low concentration in the infected tissue and they are not mechanically transmissible. Two infectious clones have been reported for members of the luteoviruses, namely barley yellow dwarf virus (BYDV-PAV serotype) (Young et al., 1991) and beet western yellows virus (BWYV) (Veidt et al., 1992). In this appendix, I describe construction of a full-length cDNA clone of BYDV-PAV to understand functions of viral genes, and the roles of specific sequences in translation, replication and encapsidation. The in vitro synthesized transcripts were infectious when introduced into oat protoplasts. Using several site-specific and deletion mutants we show that the ORF 1, ORF 2 and either ORF6 or 1157 nucleotides at the 3' end of the genome are
required for efficient replication in protoplasts. These results contradict the published results from Young et al., (1991).

Construction of full-length cDNA clone of BYDV-PAV RNA

A full-length clone of BYDV-PAV RNA was constructed for in vitro transcription (Figure 1). The protocol of (Frohman et al., 1988) was used to amplify the 5' end of PAV-IL RNA. The oligonucleotide: 3' - CCAGTGTCCTGCGCTCGGGCCGA^GG^CCCJ^CCCJCG/K^JCCCCGGGAAC- 5', containing sequence complementary to bases 1225-1246 (italics) and added BspHI, HindIII and Apal restriction sites (underlined) was used to prime first-strand cDNA synthesis using MMLV RNase H" reverse transcriptase (Superscript, BRL) (Sambrook et al., 1989). The upstream primer, 5'-ACGGCGCCGCTAATACGACTCACTATTGCATACCATCTTCACA-3' was similar to that used by Young et al., (1991). It contained (5' to 3') a NotI site (underlined), the bacteriophage T7 DNA-dependent RNA polymerase promoter sequence (italics), followed by a G residue and the first 23 nucleotides of the 5' end of the virion RNA. The amplified product was gel-purified and cloned into NotI-Apal-cut pGEM5Zf(+) (Promega), resulting in plasmid pPAV1 (Figure 1). To obtain pPAV2, the Clal-HindIII fragment of pSP15 (bases 1044-1591, Dinesh-Kumar et. al., 1992) was cloned into pPAV1 that had been digested with the same enzymes. To generate pPAV3, the HindIII-Apal fragment of pSP15 comprising 1591-2860 (Apal site is in the vector) was cloned into HindIII-Apal-cut pPAV2. Plasmid pPAV4 was constructed by cloning the
Figure 1 Genome organization of BYDV-PAV RNA and schematic diagram of full-length clone assembly. Large ORFs are indicated above relevant restriction sites, numbered as in Miller, et al., (1988a). Maps of inserts in intermediate plasmids used to assemble full-length clone (see text) are diagrammed below the genome organization. Boxes at left end of the maps represent the T7 RNA polymerase promoter; solid bars depict the portion from BYDV-PAV-IL, the remainder (open bars) is from the Australian isolate of BYDV-PAV (Miller et al., 1988a). Abbreviations: A, Aval; C, Clal; E, EcoRl; H, Hindlll; Sa, Sail; Sm, Smal; Ss, Sspl.
481 base pair (bp) EcoRI fragment (bases 2860-2985) of pSP17 (Dinesh-Kumar et al., 1992) into EcoRI-cut pPAV3.

Intermediate cloning vector pUC118-1180-H3 was constructed by cloning the ScaI-Sall fragment of pSL1180 (Pharmacia, Milwaukee, WI) into pUC118 from which the HindIII site had been deleted. Then the Sall-Smal fragment of pSP18 (Dinesh-Kumar et al., 1992) was cloned into pUC 118-1180-H3, resulting in pPAV5. Finally, the NotI-Sall fragment of pPAV4 was cloned into NotI-Sall-cut pPAV5 to obtain pPAV6, which contains a full length cDNA copy of BYDV-PAV (Figure 1). All the cloning junctions were confirmed by double-stranded DNA sequencing (Taq Track sequencing kit, Promega).

Construction of plasmid pFLFSM4

The BYDV-PAV sequence from nucleotide 1044 (ClaI site) to 1591 (HindIII site) was cloned into pUC118-1180-H3 to create plasmid pPAV(C-H). Mutations were made by the method of Herlitze and Koenen, (1990), using M13 forward and reverse primers and mutagenic primer DIFS1: 5'-AGAGCCCCTCTGAAAAAAACCCAC- 3' (inserted base underlined). Primer DIFS1 changed the 39K stop codon from UAG to UCAG, creating in-frame 99K fusion of the 39K and 60K ORFs. After mutagenesis, the ClaI-HindIII fragment was subcloned into pPAV2, then the NotI-HindIII fragment was cloned back into the full-length clone to create plasmid pFLFSM4.
Sequence of 5' terminal 1044 nucleotides of BYDV-PAV-Illinois strain and comparison with the other BYDV-PAV strains

The full-length clone of BYDV-PAV described here is a chimera with the 5'-terminal 1044 nucleotides derived from Illinois strain of BYDV-PAV (PAV-IL), while the remainder is from the sequenced Australian isolate (Miller et al., 1988b). To determine the sequence of the 5' end of PAV-IL strain, four subclones were generated from pPAV1. These subclones were sequenced by dideoxy method (Sanger et al., 1980) using Taq Track double stranded sequencing kit (Promega).

Figure 2 shows the 5' terminal 1044 nucleotide sequence of BYDV-PAV-IL and its alignment with sequenced Australian (BYDV-PAV-Aust), and Purdue (BYDV-PAV-P) strains. The sequences were analyzed using computer programs GCG sequence analysis software package version 6 and an IBM-compatible program by W.R. Bottomley (CSIRO, Division of Plant Industry, Canberra, Australia). The nucleotide sequence of PAV-IL has 93.1% identity to the Australian strain and 93.6% identity to the Purdue strain (Figure 2). At the amino acid sequence level, PAV-IL was 97.7% identical to Australian strain and 98.7% identical to the Purdue strain (data not shown). Despite the differences, the full-length chimeric transcript was highly infectious in oat protoplasts (see below).

In vitro translation of full-length BYDV-PAV RNA

The translation properties of the full-length transcripts and virion RNA were analyzed in eukaryotic cell-free translation systems. Full-length and
<table>
<thead>
<tr>
<th>Position</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-50</td>
<td>AGTGAAGATT GACCATCTCA CAAAAGCTGT TACG-TGCTT GTAACACACT</td>
</tr>
<tr>
<td>51-100</td>
<td>GCACGCGCTG TTTGTATTC GGAAGTAGTT GGCAAAGCG TCCCCCTATT</td>
</tr>
<tr>
<td>101-150</td>
<td>GCCGTGACAA CCTAAGGAGCC CCACTTCCCA CCAAGCCCA TCAAGTTCCTT</td>
</tr>
<tr>
<td>151-200</td>
<td>GACCGGCGCCA GGGCGCAGG GGTCAAGAG CTCATTTGAC</td>
</tr>
<tr>
<td>201-250</td>
<td>ATGGTCTATT TAGATTGAAA TCTATATTAC ATGCTTTGAA</td>
</tr>
<tr>
<td>251-300</td>
<td>ATGGGATATT CAGGGCCAAT TTGAGGACTT CGAGGGGAAG CTCGCTGAGG</td>
</tr>
<tr>
<td>301-350</td>
<td>AGTGTGCCGA GAGGGAGGAT GAGGTGGAAG AGGCGAGGAG CCTCTTGAAA</td>
</tr>
<tr>
<td>351-400</td>
<td>CTGTGCGTCG CCCAAAAATC TAAATCTGGG GTGACCGAGG CTTGGACCGA</td>
</tr>
<tr>
<td>401-450</td>
<td>CTGCCGGGAA CCGAGAGGAT GAGGTGGAAG AGGCGAGGAG CCCCCTTCCAA</td>
</tr>
<tr>
<td>451-500</td>
<td>CTGTGCGTCG CCCAAAAATC TAAATCTGGG GTGACCGAGG CTTGGACCGA</td>
</tr>
<tr>
<td>501-550</td>
<td>CTGCAGGCA AGAGCGAGG GTGTTGTTTA GCCACACCCT TCTGCGGAGC</td>
</tr>
<tr>
<td>551-600</td>
<td>GAGAGCGAGC ACCAATTGGA GGGCGAGGG GCCAAAAAT AATCATCGGA</td>
</tr>
<tr>
<td>601-650</td>
<td>AAGGGGCGCC GGCTCGCACT AAGCTGGAAA CGACTTGGG AGTCTCTCA</td>
</tr>
</tbody>
</table>
Figure 2 Alignment of 5'-terminal 1044 nucleotide sequence of Australian (Aust), Illinois (IL), and Purdue (P) isolates of barley yellow dwarf virus. Complete sequence of Australian isolate is shown. In Illinois and Purdue isolates only the bases that are different from Australian isolate are shown.
mutant RNA transcripts were synthesized by in vitro transcription of Smal-linearized pPAV6 and mutant clones with the T7 Megascript kit (Ambion, TX). *In vitro* translation was carried out with 200 ng of uncapped transcripts or virion RNA. One-fifth of the *in vitro* translation reaction mix was electrophoresed through 5% stacking and 10% resolving polyacrylamide-SDS gels using a discontinuous buffer system (Laemmli, 1970). Gels were fixed and fluorographed (Hames, 1990). The relative radioactivity in each gel band was measured using Imagequant 3.0 software on a Phosphorimager 400E (Molecular Dynamics, Sunnyvale, CA). Frameshift rate was calculated as: (relative radioactivity in 99K band/28) X 100/[(relative radioactivity in 39K band/10) + (relative radioactivity in 99K band/28)]. This normalized the radioactivity for the 28 methionine residues in the (39K + 60K) = 99K transframe ORF and the 10 methionine residues in the 39K ORF alone.

When BYDV-PAV virion RNA or full-length transcripts were translated under optimal ionic conditions (below) in a wheat germ extract, an abundant peptide was produced that migrated at an apparent molecular weight of 47 kDa, using BMV translation products as markers (Figure 3, lanes 2, 3). We believe that this is the product of the 39K ORF. A less abundant product that migrated as a 95 kDa polypeptide was also visible. This is close to the size and abundance expected for the transframe protein (39K + 60K = 99K). To verify this, a deoxycytidilate residue was inserted in the stop codon of the 39K ORF (mutant pFLFSM4) to destroy the stop codon and place the 60K ORF in the same frame as the 39K ORF, resulting in a single 99K ORF. As expected, the translation product of this RNA comigrated with
Figure 3 Fluorograph of *in vitro* translation products of BMV RNA (lanes 1, 6), BYDV-PAV virion RNA (lanes 2, 7), transcripts of Smal-linearized pPAV6 (lanes 3, 8) and pFLFSM4 (lanes 4, 9) and no added RNA (lanes 5, 10), following SDS-PAGE. Lanes 1-5: wheat germ translation products (2.5 µl aliquots of BMV RNA reaction, 5 µl aliquots of BYDV-PAV RNA reactions). Lanes 6-10: rabbit reticulocyte lysate products (2.5 µl aliquot of BMV RNA reaction, 10 µl aliquots of BYDV-PAV RNA reactions). All reactions contained 0.2 mg of RNA template in 25 µl reaction mixture with 1,071 Ci/mmol 35S-methionine. Details are in the text. Molbilities (in kilodaltons) of BMV products are shown on the left; expected sizes of BYDV-PAV products are on the right.
the approximately 95 kDa product from wild-type RNA but was present in much greater abundance (Figure 3, lane 4). Concomitantly, little or no 47 kDa-migrating translation product was present among translation products of pFLFSM4 RNA, supporting the notion that it is the product of the 39K ORF.

In contrast to the wheat germ translation products, in the reticulocyte extract, using optimal ionic conditions (Dinesh-Kumar et al., 1992), much smaller amounts of putative 39K ORF and 99K products were detected (Figure 3, lanes 7, 8). A faint band migrated at about 99K, but trace amounts of a similar band were present in the extracts to which no RNA was added (Figure 3, lane 10). Thus any frameshifting that occurs in reticulocyte lysates, is at a much lower efficiency than in wheat germ and is partially obscured by the endogenous band. Translation of the 39K ORF was two- to threefold less efficient in reticulocyte lysates than in wheat germ. Whereas BMV RNA translated more efficiently in reticulocyte lysates. The relative amounts of 39 and 99 kDa polypeptides were measured using ImageQuant software on Phosphorimager scans of the gels. The rate of frameshift of transcripts in wheat germ extracts ranged from 2.5 to 5%, but the virion RNA frameshifted at a rate of 12%. When the background band of approximately 99 kDa in the no RNA lane was subtracted from the reticulocyte lysate translation products, the full-length and virion RNA were found to frameshift at rates of 1.2 and 1.5%, respectively.

Because optimal ionic conditions can vary between mRNAs and between batches of cell-free systems (Kemper and Stolarsky, 1977; Kozak, 1990), the ionic conditions of the translation systems were varied to see if
this would affect the amount of frameshift product. Increasing the potassium acetate concentration from 54 mM, supplied by the manufacturer, to 154 mM dramatically improved translation efficiency. Wheat germ translation shown in Figure 3 was performed in 154 mM potassium acetate. Despite the extensive optimization of the magnesium and potassium concentrations in the reticulocyte lysate, the rate of frameshifting remained substantially below that in wheat germ extracts (Figure 3).

Biological activity of BYDV-PAV transcripts

Oat protoplasts were isolated and electroporated with viral RNA or full-length transcripts (Dinesh-Kumar et al., 1992; Dinesh-Kumar and Miller, 1994). Infection was verified by ELISA of protoplast extracts using BYDV-PAV-specific antibodies.

Synthesis of viral coat protein from protoplasts electroporated with 5 mg of full-length transcripts or 100 ng of virion RNA could be readily detected 24 hr. after inoculation and continued to accumulate to at least 48 hr (Figure 4). The coat protein levels revealed by high ELISA readings indicate that subgenomic RNA (message for CP) is synthesized as a result of virus replication and the CP is not from the inoculum RNA. Five mg of transcript gave approximately the same amount of coat protein expression after 48 hr as 100 ng of virion RNA (Figure 4). This indicates that the full-length infectious transcripts were approximately 50-fold less infectious compare to equivalent amount of virion RNA. This full-length transcript is more highly infectious than that reported
Figure 4 Expression of BYDV-PAV coat protein in oat protoplasts electroporated with 500 ng of virion RNA (viral RNA), 5 µg of capped and uncapped full-length transcripts derived from Smal-linearized pPAV6, and no RNA. ELISA values were determined at 0, 24 and 48 hr after inoculation.
previously (Young et al., 1991), in which coat protein was first detected 72 hr after inoculation in *Triticum monococcum* cells.

Coat protein was not detected from protoplast extracts when capped full-length transcripts were used for infection (Figure 4). Most viral RNAs contain either cap structure or genome-linked viral protein (VPg) at the 5' end. In both cases capping enhances infectivity. Failure of capped transcripts to replicate in case of BYDV-PAV may indicate the lack of a cap or cap-like structure at the 5' end of the RNA. This is supported by the observation that the BYDV-PAV genome translates in a cap-independent manner in a wheat germ translation system (Wang and Miller, unpublished). Furthermore, the RNA-dependent RNA polymerase of BYDV-PAV is more similar to the polymerases of tombus-, carmo- and dianthoviruses that lack VPg's than to those which have VPg's (BYDV-RPV and other subgroup II luteoviruses). In contrast to our findings, Young et al., (1991) reported two-fold higher infection with capped BYDV-PAV transcripts compared to uncapped transcripts, but later said capping made no difference (pers. communication).

**Effect of deletions in ORFs 1, 5 and 6 on BYDV-PAV infection**

Deletion mutants created from pPAV6 are shown in Figure 5. To determine whether the 39 kDa protein expressed from ORF1 is required for replication, two clones pPAV19 and pPAV26 were generated. To create plasmid pPAV19, first the plasmid pPAV6 was cut with Csp45I (nt 148), and
Figure 5 Infectivity of BYDV-PAV RNA transcripts containing deletions in ORFs 1, 5 and 6 in oat protoplasts. A. Genome organization of BYDV-PAV. Restriction sites used in creating deletion mutants with corresponding position in the BYDV-PAV genome (Miller et al., 1988; and Figure 2 of this chapter) are shown below the genomic RNA. B. Accumulation of BYDV-PAV coat protein in oat protoplasts electroporated with 5-10 mg of transcripts derived from Smal-linearized full-length and mutant clones. Gap in mutants indicates deletion sites. ELISA values are average of two replicates.
filled in with Klenow fragment of DNA polymerase I, and then cut with Eco72I (1135) prior to recircularization. The resulting plasmid pPAV19 contain a deletion in the ORF1 (nts 148-1135) (Figure 5). When transcript derived from pPAV19 was translated in wheat germ extract, expression of ORF2 (60 kDa polymerase protein) was also affected in addition to ORF1 (data not shown). To overcome this effect, plasmid pPAV26, containing a smaller deletion in the ORF1 (nts 578-959), was generated by digesting the plasmid pPAV6 with Ball and religation.

Transcripts derived from PAV19 and PAV26 were inoculated into oat protoplasts. The ELISA readings indicate that these transcripts fail to accumulate coat protein (Figure 5B). Thus 39 kDa product of ORF1 appears to be indispensable for replication of BYDV-PAV RNA in protoplasts. This result was expected because 39 kDa protein and the frameshift protein (99 kDa) contain conserved sequence motifs characteristic of viral and cellular helicases and RNA-dependent RNA polymerases (Habili and Symons, 1989). This contrasts with the findings of Young et al., (1991), in which frameshift mutation in ORF 1 did not affected the accumulation of BYDV-PAV coat protein.

To address the role of the readthrough protein (ORF5) in viral replication, plasmids pPAV13, pPAV22 were generated (Figure 5B). pPAV13, which carries a deletion in the ORF5 from nt 3785-4513 was constructed by inserting the Scal-Xmal fragment (nts 4513-5677) from pPAV6 between Hpal-Xmal (nt 3785-5677) sites of pPAV6. pPAV22 was obtained by cloning the klenow filled in BamHI- Xmal fragment (nts 4837-
5677) from pPAV6 into the Hpal-Xmal cut (nts 3785-5677) pPAV6. This plasmid carries a deletion in ORF5 from nt 3785-4813 and an additional 23 nt 3' of ORF5.

The coat protein accumulated to the wildtype level in protoplasts electroporated with transcripts derived from pPAV13 (Figure 5B). This shows that the readthrough protein is not involved in replication or coat protein synthesis or virus assembly. This is in agreement with the findings in BWYV, in which frameshift mutations or deletion mutants in the readthrough region did not affect viral RNA replication (Reutenauer et al., 1993). In contrast, Young, et al., (1991) claimed that frameshift mutations in the BYDV-PAV readthrough region eliminated infectivity. Subsequently he found they did not (pers. communication).

When transcripts derived from pPAV22 were inoculated into oat protoplasts, there was no accumulation of coat protein (Figure 5B). This indicate that the region between Scal and BamHI is required for BYDV-PAV replication. Furthermore, translation of pPAV22 transcript in wheat germ extract resulted in no 39 kDa protein or 99 kDa frameshift product (data not shown). This region is also required for cap-independent translation in wheat germ extracts (Wang and Miller, unpublished). Since uncapped transcripts are used for inoculation of the protoplasts, lack of synthesis of ORF 1 and ORF 2 products \textit{in vivo} might be the reason for pPAV22 transcript being non-infectious.

To determine whether ORF 6 is required for infection, plasmid pPAV24 containing a deletion in ORF 6 from nt 4837-5190 was generated (Figure.
5B). First, plasmid pPAV6 was cut with BamHI and BclI. Of the three resulting fragments two large fragments of size 3097 (in the PAV genome) and 4967 (PAV genome and the vector sequence) were gel-purified and religated. Transcripts derived from this plasmid fail to infect oat protoplasts (Figure 5B), which implies that expression of ORF 6 may be required for replication of BYDV-PAV RNA. But it is not clear whether the protein or the nucleotide sequence is required for replication, because the deletion of this region in the full-length clone may inhibit the synthesis of the second subgenomic RNA (message for ORF 6). Findings reported by (Young, et al., 1991) indicate that the expression of this protein is required for replication because frameshift mutation introduced in this protein abolished infectivity in protoplasts, but many of our previous results contradict their's. We believe that the cis-sequence in this region is required for efficient replication, because when these uncapped transcripts are translated in wheat germ extract, they fail to express ORF1 and ORF2 products. Furthermore, part of this deletion up to PstI site (nt 5009) is required for cap-independent translation in wheat germ extract (Wang and Miller, unpublished). It is interesting to see whether capping of pPAV22 and pPAV24 transcripts will rescue the replication properties in oat protoplasts.

**Future work**

Several additional deletions in the different part of the BYDV-PAV genome were generated (Figure 6). This will facilitate in identification of different cis-sequences and viral trans-factors that are required for BYDV-PAV replication.
Figure 6 Diagram of various point and deletion mutants of BYDV-PAV genome. A. Restriction sites used in creating deletion mutants with corresponding position in the BYDV-PAV genome (Miller et al., 1988a; and Figure 2 of this chapter). B. Structure of point and deletion mutants. PAV6 represents the full-length infectious clone. Gap in mutants indicate deletions in the full-length genome. In PAV30, ORFs 3, 4 and 5 are deleted and the start codon of ORF6 is changed to ACG. The start codon of ORF3 (CP) is modified to ACG in PAV31, and start codon of ORF4 is changed to GCG in PAV32.
APPENDIX II

ROLE OF 39K ORF STOP CODON ON BYDV-PAV FRAMESHIFTING AND REPLICATION
A shifty heptanucleotide sequence followed by a stem-loop structure or pseudoknot structure on the mRNA are two most important cis-signals found to play a major role in the translational frameshift event (reviewed in Atkins et al., 1990; Hatfield et al., 1992). Another signal which might play a role in frameshifting is stop codon of the zero reading frame (Weiss et al., 1987; Brault and Miller, 1992). In BYDV-PAV changing the stop codon (UAG) of 39K ORF into UCG completely abolished frameshifting in plant cells. In synthetic constructs, insertion of a stop codon next to shifty heptanucleotide sequence resulted in enhancement of frameshift efficiency in Escherichia coli (Weiss et al., 1987; Weiss et al., 1990). In some eukaryotic viruses including BYDV-PAV, red clover necrotic mosaic virus (RCNMV, Xiong and Lommel, 1989), rous sarcoma virus (RSV, Jacks et al., 1988) and mouse mammary tumor virus (MMTV, Hizi et al., 1987), the stop codon is located immediately 3' to the shifty heptanucleotide sequence (Figure 1). The stop codon might cause pausing of ribosomes, which may give more time for ribosomes to shift at the slippery site. In E. coli, the gene for polypeptide release factor 2 (RF-2) contains an in-frame stop codon. To produce full-length RF-2 product, it uses +1 frameshift event. In this case, ribosomal pausing at the in-frame stop codon plays a crucial role in directing the efficient frameshift event (Donly et al., 1990). Recently it has been shown that in case of reovirus monocistronic s4 RNA translation, the ribosomes pause more at start and stop codons both in vitro and in vivo (Doohan and Samuel, 1993).
Figure. 1 Eukaryotic viruses that contain upstream reading frame stop codon immediately 3' to the shifty heptanucleotide. *pol*, reverse transcriptase (in RSV and MMTV), or RNA-dependent RNA polymerase in (BYDV-PAV and RCNMV). *gag*, virus core protein.
To understand the role of stop codon on frameshift event in full-length natural context of BYDV-PAV RNA, I introduced several mutations in the 39 K stop codon (see below).

**Construction of 39K ORF stop codon mutants**

The UAG stop codon of 39K ORF was replaced by several sense and other stop codons by two step polymerase chain reaction (PCR) mutagenesis method of (Landt et al., 1990) using a subclone pPAV(C-H) (see Appendix I). Mutations were made by using a mutagenic primer (see table 1) and M13 forward and reverse primers as flanking primers. After mutagenesis, the Clal-HindIII fragment was subcloned into pPAV2 (see Appendix I), then the NotI-HindIII fragment was cloned back into the full-length clone pPAV6 (see Appendix I; Di et al., 1993). All mutations were confirmed by double stranded sequence analysis.

**Table 1. Primers used for 39K ORF stop codon mutant construction**

<table>
<thead>
<tr>
<th>Primer name</th>
<th>mutagenic primer (underlined bases are modified)</th>
<th>mutant generated</th>
<th>expected change in 39K stop codon</th>
</tr>
</thead>
<tbody>
<tr>
<td>ks3696</td>
<td>AGAGCCCCTT(TC)AAAAACCCAC</td>
<td>pFLFSM1</td>
<td>UAA</td>
</tr>
<tr>
<td></td>
<td></td>
<td>pFLFSM2</td>
<td>UGA</td>
</tr>
<tr>
<td>ks5077</td>
<td>ACAGAGCCCCCTCGAAAAACCCAC</td>
<td>pFLFSM3</td>
<td>UCG</td>
</tr>
<tr>
<td>ks3697</td>
<td>AGAGCCCCTC(CGA)AAAAACCCAC</td>
<td>pFLFSM5</td>
<td>CGG</td>
</tr>
<tr>
<td></td>
<td></td>
<td>pFLFSM6</td>
<td>UGG</td>
</tr>
</tbody>
</table>

^ All primers are complementary to the viral bases 1170 to 1150
Role of 39K ORF stop codon on frameshifting and BYDV-PAV infection

Transcripts were generated from Smal-linearized full-length infectious clone pPAV6 and clones containing mutations in the 39K ORF stop codon using T7 megascript kit (Ambion, TX). Two hundred nanograms of transcripts were translated in wheat germ extract in the presence of 35S-labeled methionine, and frameshift percentage was calculated as described previously (see Appendix I). Changing the stop codon of the 39K ORF to sense codons (UCG, CGG, UGG) or another stop codon (UAA or UAG) had no effect on frameshifting efficiency (Figure 2). The 39K ORF reading frame is extended by 132 nucleotides in those mutations that convert the stop codon to a sense codon. This resulted in production of a 42 kDa protein instead of 39 kDa (Figure 2). These in vitro results indicate that the stop codon of the 39K ORF is not required for frameshifting. This is similar to RSV, in which mutations in the gag terminator did not affect frameshift efficiency in reticulocyte lysate (Jacks et al., 1988). But it differs from the in vivo reporter gene (GUS) expression in protoplasts, in which the frameshift was abolished when the 39K ORF stop codon was changed to UCG (Brault and Miller, 1992) or other sense codons (Paul and Miller, unpublished).

To investigate the role of 39 K stop codon on BYDV-PAV infection, 5-10 mg of full-length wildtype and mutant transcripts were electroporated into oat protoplasts (Dinesh-Kumar et al., 1992; Dinesh-Kumar and Miller, 1993). Infection was analyzed by ELISA using BYDV-PAV specific antiserum. Accumulation of coat protein was similar to that of wild type level when the
Figure. 2  *In vitro* translation of full-length and 39K stop codon mutant transcripts. Two hundred nanograms of intact transcripts were translated in wheat germ extract in the presence of $^{35}$S methionine. Products were separated on 10% polyacrylamide-SDS gel and visualized after fluorography. Mobilities (in kDa) of BMV translation products are shown on the left; expected sizes of BYDV-PAV products are shown on the right. Calculated frameshift rate are shown below.
UAG stop codon of 39K ORF was changed to UAA or UGA stop codons (Figure 3A). Mutant transcripts in which the 39K ORF stop codon was changed to the sense codons UCG, CGG or UGG or UCAG abolished infectivity (Figure 3A). Replication of mutant and wild type transcripts were determined by Northern blots of total RNA extracts using viral RNA probes (Figure 3B). Only those transcripts with 39K ORF stop codon replicated in oat protoplasts (Figure 3B). The results of replication and infection studies are consistent with the in vivo GUS expression assay (Brault and Miller, 1992)

Conclusions

It is difficult to explain the differences between in vitro and in vivo results. Since any change of the stop codon of 39K ORF to a sense codon abolishes infection indicate that the stop codon may be required for efficient frameshifting in vivo to express the RNA-dependent RNA polymerase product (P1/P2). Alternatively frameshifting may occur normally in infected cells but addition of extra C-terminal amino acids into the 39kDa protein may inactivate its function. Deletion analysis reported in Appendix I shows that the 39 kDa protein is essential for infection. However, when same sense mutations were introduced into 39K ORF stop codon, they failed to express GUS in oat protoplast system (Paul and Miller, unpublished). This suggests that the stop codon may play an essential role in frameshifting in vivo in BYDV-PAV. It is interesting to see whether these mutants can be rescued for replication by providing functional polymerase in trans.
Figure. 3 Infection and replication of full-length and 39K stop codon mutant transcripts. Approximately 5-10 µg of transcripts derived from Smal-linearized full-length PAV6 and mutant clones were electroporated into oat protoplasts. A. Accumulation of coat protein 24 and 48 hr after inoculation was determined using BYDV-PAV specific antiserum. B. Replication was detected by Northern hybridization of total RNA, extracted 48 hr after inoculation using ^32P-labeled RNA probe complementary to the 5' end of the genome. G, position of genomic RNA.
APPENDIX III

SIGNALS THAT CONTROL TRANSLATIONAL READTHROUGH OF COAT PROTEIN STOP CODON IN BARLEY YELLOW DWARF VIRUS IN VIVO
Signals involved in luteovirus coat protein (CP) stop codon readthrough have not been investigated. In *in vitro* translation experiments with full-length synthetic subgenomic RNA of BYDV-PAV in reticulocyte lysates, the CP stop codon is suppressed at about 7 to 15% (Dinesh-Kumar et al., 1992). However, *in vitro* readthrough efficiency was highly susceptible to changes in salt concentration (magnesium chloride and potassium acetate). Hence I decided to determine the signals that are required for efficient readthrough *in vivo* using the oat protoplast system.

**Effect of different amounts of viral sequence spanning CP stop codon on readthrough using GUS expression system**

To monitor readthrough, we inserted various amounts of sequence spanning the CP stop codon region between the start codon and the rest of the coding region of *Escherichia coli* uidA (*GUS*) gene in pAdh-GUS (Dinesh-Kumar and Miller, 1993) and pAGUS1 (Skuzeski et al., 1990) expression vectors. Expression of *GUS* gene requires translational readthrough of the CP stop codon. Plasmid pAdh-GUS contains the promoter from maize alcohol dehydrogenase (*Adh1*) gene and first intron, whereas pAGUS1 contains a duplicated cauliflower mosaic virus 35S promoter. Two types of CP stop codon downstream sequences were included in various constructs. One with the C-rich sequence which is present downstream of the CP stop codon in all luteoviruses and another without the C-rich sequence. The C-rich sequence results in protein product in which every other amino acid is proline. This proline-rich domain has been suggested to play a role in
separating the coat protein domain from the readthrough domain (Bahner et al., 1990).

Plasmids pRT1 to pRT4 (Figure 1) were obtained by PCR, using pPAV6 full-length BYDV-PAV plasmid as template and different pairs of upstream and downstream primers (see Table 1). In all cases the Ncol-Apal cut PCR amplified fragment was gel purified and cloned into pAdh-GUS that had been cut with the same enzymes. The pairs of primers used were ks 3598 and ks3597 for pRT1; ks3598 and ks1644 for pRT2; ks 3596 and ks3597 for pRT3; ks 3596 and ks1644 for pRT4. Plasmids pRT1 and pRT3 include the proline rich sequence downstream of CP stop codon. To determine the readthrough percentage, plasmids pRT5 to pRT8 were generated, in which the CP stop codon was changed into a sense codon (GAG, glutamate). This was achieved by the two step PCR method of Landt et al. (1990) using pRT1 to pRT4 as templates and respective pairs of upstream and downstream primers and mutagenic primer ks5079, 5' GAGGAGTCTACCTCTTTTGCCG 3', (modified nucleotide is underlined).

Table 1. Primers used for construction of readthrough plasmids

<table>
<thead>
<tr>
<th>Primer Name</th>
<th>Sequence</th>
<th>bases in BYDV-PAV</th>
</tr>
</thead>
<tbody>
<tr>
<td>ks3596</td>
<td>CGCCATGGTCATTACGATGAGTGTCACTGATGTCAGT</td>
<td>3423 to 3442</td>
</tr>
<tr>
<td>ks3598</td>
<td>CGCCATGGTCCAGGAATCAACGATA</td>
<td>3345 to 3361</td>
</tr>
<tr>
<td>ks3597</td>
<td>ATGGGCCCCGATATACCTCGAAGAA</td>
<td>3591 to 3575</td>
</tr>
<tr>
<td>ks1644</td>
<td>TAGGGGCCGGTGTGAGAGTCTCA</td>
<td>3477 to 3482</td>
</tr>
</tbody>
</table>

a Ncol site is underlined
b Apal site is underlined. Primers are complementary to the viral RNA.
Figure. 1 GUS activity from plasmids with various amounts of coat protein stop codon spanning region. Approximately 80 µg of plasmid was electroporated into oat protoplasts and GUS activity was measured after 24 hr. Each GUS activity value is the mean of two replicates. Percent readthrough was calculated as the percentage of GUS activity of each stop codon containing constructs with its control construct that differed only by the replacement of CP UAG with GAG. Calculations were made after subtracting background value in cells electroporated with salmon sperm DNA.
<table>
<thead>
<tr>
<th>Adh Promoter</th>
<th>GUS activity (fluorescence units)</th>
<th>% Read-through</th>
</tr>
</thead>
<tbody>
<tr>
<td>3345 3591</td>
<td>pRT1 17.0</td>
<td>0.15</td>
</tr>
<tr>
<td>3345 3477</td>
<td>pRT2 15.6</td>
<td>0.09</td>
</tr>
<tr>
<td>3345 3591</td>
<td>pRT5 10926.2</td>
<td></td>
</tr>
<tr>
<td>3345 3477</td>
<td>pRT6 17536.2</td>
<td></td>
</tr>
<tr>
<td>3423 3591</td>
<td>pRT3 22.0</td>
<td>0.08</td>
</tr>
<tr>
<td>3423 3591</td>
<td>pRT7 25541.2</td>
<td></td>
</tr>
<tr>
<td>3423 3477</td>
<td>pRT4 22.4</td>
<td>0.09</td>
</tr>
<tr>
<td>3423 3591</td>
<td>pRT8 23461.2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>pAGUS1 20710.0</td>
<td></td>
</tr>
<tr>
<td>Salmon DNA</td>
<td></td>
<td>23.8</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>35S Promoter</th>
<th>GUS activity (fluorescence units)</th>
<th>% Read-through</th>
</tr>
</thead>
<tbody>
<tr>
<td>3345 3591</td>
<td>pRT9 10.0</td>
<td>0.28</td>
</tr>
<tr>
<td>3345 3477</td>
<td>pRT13 3505.0</td>
<td></td>
</tr>
<tr>
<td>3345 3477</td>
<td>pRT10 11.6</td>
<td>0.25</td>
</tr>
<tr>
<td>3423 3591</td>
<td>pRT14 4615.6</td>
<td></td>
</tr>
<tr>
<td>3423 3477</td>
<td>pRT11 9.6</td>
<td>0.23</td>
</tr>
<tr>
<td>3423 3477</td>
<td>pRT15 4121.6</td>
<td></td>
</tr>
<tr>
<td>3423 3477</td>
<td>pRT12 4.4</td>
<td>0.1</td>
</tr>
<tr>
<td>3423 3477</td>
<td>pRT16 4384.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>pAGUS1 4335.6</td>
<td></td>
</tr>
<tr>
<td>Salmon DNA</td>
<td></td>
<td>19.5</td>
</tr>
</tbody>
</table>

- coat protein region
- CP-rich region
- gene
- CP stop codon
- GUS gene
- readthrough region
- % Read-through

Coomassie blue staining of total protein extracts from transgenic tobacco grown in soil. (A) Photomicrograph of Adh promoter region. (B) Photomicrograph of 35S promoter region. (C) Photomicrograph of 35S promoter region. (D) Photomicrograph of 35S promoter region.
Approximately $10^6$ oat protoplasts were electroporated with 80 mg of plasmid DNA (Dinesh-Kumar and Miller, 1993) and GUS activity was determined after 24 hr. of incubation in the dark (Jefferson, 1987). The results are summarized in Figure 1. In none of the constructs containing the CP stop codon was termination suppressed, because observed GUS activity was only about 50% higher than the background. Rate of readthrough observed was about 0.1 to 0.15 %, which is 70 to 100 fold less than that observed in vitro (Dinesh-Kumar et al., 1992). A similar approach was used to study tobacco mosaic virus (TMV) class of readthrough signals and the efficiency observed was about 5% in vivo (Skuzeski et al., 1990). However they failed to detect readthrough of maize chlorotic mottle virus (MCMV) and carnation mottle virus (CarMV) stop codons which belong to the luteovirus group of readthrough signals. In potato leaf roll luteovirus (PLRV), the construct similar to our pRT4 resulted in suppression of about 1% in tobacco and potato protoplasts (Tacke et al., 1990). Failure to observe readthrough in our experiments is not due inefficient expression level in oat protoplast system because our positive control values (Figure 1) were comparable with PLRV and TMV experiments.

Transcripts from constructs pRT1 to pRT8 contain 141 non viral bases from the Adh1 promoter. To avoid this, Ncol-Apal fragment from pRT1 to pRT8 were inserted into pAGUS1 to generate plasmids pRT9 to pRT16 which used the CaMV 35S promoter (Figure 1). Transcripts from these plasmids contain only 11 non viral bases. These constructs were also unable to suppress the CP stop codon (Figure 1).
Effect of the entire CP coding sequence on efficiency of suppression

To see whether more extensive sequence elements are required than those used in the above constructs, we cloned the entire subgenomic RNA leader, CP coding region and part of readthrough region into pAdh-GUS and pAGUS1 (Figure 2). Plasmids pRT17 and pRT19 were generated by PCR using pPAV6 as template and upstream primer SPDK8 (5' CGGGATCCGATAGGGTTTATAGTTAGTA 3'), which contained a BamHI site (underlined) and sgRNA1 bases 1 to 20, and ks3597 as downstream primer (table 1). Plasmids pRT18 and pRT20 were generated using SPDK8 as upstream primer and ks1644 as downstream primer. Positive control plasmids pRT21 to pRT24 were generated using the mutagenic primer ks5079 as explained above. In all cases the PCR-amplified fragments were digested with BamHI-Apal and cloned into pAdh-GUS or pAGUS1 that had been cut with the same enzymes.

These plasmids were electroporated into oat protoplasts and GUS activity was measured after 24 hr. Results are summarized in Figure 3. Although percent readthrough observed is higher in this experiment, it is difficult to accept these values because the GUS activity due to the CP suppression is only about one-third higher than that of the background. Higher readthrough percentage in case of pRT18 is a result of very low GUS activity from positive control construct (pRT22). The positive control values from the construct pRT22 is approximately 100 fold less compared to the positive control values obtained from constructs pRT4 to pRT8 in Figure 1. It indicates that the longer sequence at the N-terminal GUS coding region is
Figure 2  GUS activity from plasmids containing subgenomic leader, entire coat protein coding region and part of readthrough. Approximately 80 mg of plasmid was electroporated into oat protoplasts and GUS activity was measured after 24 hr. Each GUS activity value is the mean of two replicates. Percent readthrough was calculated as in Figure 1.
detrimental to its expression. However, in 35S constructs, pRT23 and pRT24 the reduction in the expression was only about 0.5 to 1 fold. The difference between Adh and 35S constructs is Adh constructs include more non viral bases, and requires processing of an intron.

Insertion of GUS gene downstream of CP stop codon in BYDV-PAV genome to study the signals involved in readthrough

Hybrid viruses consisting of the complete viral genome and extra genes like GUS or chloramphenicol acetyltransferase (CAT) have been shown to replicate efficiently and assemble into virions (Dawson et al., 1989; Doija et al., 1992). To study readthrough signals of luteoviruses in the full viral context I inserted the GUS gene downstream of the CP stop codon in BYDV-PAV full-length infectious clone. Expression of GUS in these constructs requires the synthesis of large subgenomic RNA through virus replication process. From the large subgenomic RNA it can be expressed only by translational readthrough of the CP stop codon.

Plasmids pPAVGUSRT1, pPAVGUSRT2, pPAVGUSRT3, and pPAVGUSRT4 (Figure 3) were generated by ligating AflIII-blunt-Sall fragment of pRT19, pRT20, pRT23, and pRT24 respectively into Sall-Hpal cut pPAV6. The Sall-AflIII fragment contains the viral CP and part of readthrough and the entire GUS coding region. Plasmid pPAVGUSRT2 includes the C-rich sequence. pPAVGUSRT3 and pPAVGUSRT4 are control plasmids for pPAVGUSRT1 and pPAVGUSRT2 respectively. Transcripts from these
Figure. 3 Suppression of coat protein stop codon in vivo. A. Genome organization of BYDV-PAV. GUS gene was inserted downstream of the CP stop codon in full-length infectious clone pPAV6 (see text for details). B. Approximately 5-10 mg of transcripts derived from Smal-linearized full-length clone pPAV6 or clones containing GUS gene were electroporated into oat protoplasts, and GUS activity was measured after 24 hr. GUS activity is expressed as pmoi of methylumbelliferone (MU) synthesized per mg of total protein per min. Each GUS activity value is the mean of two replicates. Percent readthrough was calculated as in Figure 1.
plasmids were synthesized using T7 Megascript kit (Ambion, TX). About 5-10 mg of transcripts were electroporated into oat protoplasts and GUS activity was determined.

Transcripts derived from pPAVGUSRT2 which contained the C-rich sequence downstream of CP stop codon, expressed GUS at a level of 8.27% of that of positive control transcript derived from pPAVGUSRT4, in which CP stop codon is changed to sense codon (GAG) (Figure 3). The percentage suppression observed from the transcripts derived from the plasmid pPAVGUSRT1, in which GUS inserted upstream of the C-rich sequence was about 0.23% (Figure 3). The level of GUS activity in cells infected with pPAVGUSRT2 transcript was about 100 fold higher than the protoplasts inoculated with pPAV6 full-length transcript without the GUS gene. The level of GUS activity observed in plasmid pPAVGUSRT1 was only about 3 fold higher than the control. These results indicate that the CP stop codon is suppressed in oat protoplasts at the level of 8%, and the C rich sequence present downstream of the CP stop codon is necessary for efficient readthrough. The efficiency of readthrough observed here is about the same as observed in reticulocyte lysate (7-15%; Dinesh-Kumar et al., 1992). These results could indicate possible role of viral or viral induced trans-acting factors (e.g. specific suppresser tRNAs). However, this seems very unlikely because of the fact that the readthrough occurs in rabbit reticulocyte lysates.
Possible readthrough signals

The luteovirus CP stop codon is followed by a guanosine residue (Figure 4A), which is the most favored nucleotide in the eukaryotic tetranucleotide stop signal (Brown, et al., 1990). Thus it should be a strong rather than a leaky stop signal. This implies that sequences beyond this position are required for efficient readthrough like in gag gene stop codon suppression of mammalian type C retroviruses (reviewed in Hatfield et al., 1992). Mammalian type C retroviruses also have a guanosine residue following the stop codon (Figure 4B), and the readthrough signal is located in nucleotide sequence downstream of the stop codon (Feng et al., 1992). The six C residues beginning 19 nucleotides downstream of the stop codon and a pseudoknot 3' to the stop codon are required for efficient readthrough (Feng et al., 1992; Wills et al., 1991). Similarly in BYDV-PAV, the C-rich sequence downstream of CP stop codon which is conserved in all luteoviruses and PEMV is required for efficient readthrough (Figure 4A). This C-rich nucleotide sequence results in translation product in which every other amino acid is proline. It is not clear whether the nucleotide sequence per se, or involvement of proline tRNAs are required for readthrough. The most striking downstream sequence conserved sequence in all the viruses of this group, is the pentanucleotide: CCCCA (Figure 4A). Upon extensive analysis we were unable to find any kind of secondary structure or pseudoknot structure like in case of type C retroviruses. These conserved sequence provide a guide for
Figure 4 Possible signals of stop codon readthrough in various plant and animal viruses (boxed region).
mutagenesis studies aimed at elucidating the mechanism of readthrough for the luteovirus class.

The signals required for readthrough in other classes seems to be very simple. The stop codons of TMV, TRV and alpha virus group of readthrough is followed immediately by a cytosine residue (Figure 4C and 4D), the rarest base in eukaryotic tetranucleotide stop signals (Brown et al., 1990). Hence, this may favor slow decoding, and enhanced suppression. In case of TMV (Skuzeski et al., 1991) and Sindbis virus (Li and Rice, 1993) changing this C resulted in complete abolishment of the readthrough efficiency. In Sindbis virus this C alone is sufficient to mediate the readthrough level of 10% (Li and Rice, 1993). In TMV the sequence CAR-YYA (R = purine, Y = pyrimidine) adjacent to the UAG codon is required for efficient readthrough (Skuzeski et al., 1991).