Control of Start Codon Choice on a Plant Viral RNA Encoding Overlapping Genes

S. P. Dinesh-Kumar and W. Allen Miller

Plant Pathology Department and Molecular, Cellular, and Developmental Biology Program, 351 Bessey Hall, Iowa State University, Ames, Iowa 50011-1020

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, 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 nucleases 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 that 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, 1991a; 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 m^{1}GpppG 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, 1991b). 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, 1991b). 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 Bs7 (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 (RLRs) 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 an 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, 1991b); 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). All of the base changes used in this study 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 transcript. Thus, differences in translatable 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, and in M1/2 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 to 80 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...
Figure 3. Products of Partial Nuclease Digestion of sgRNA1 Transcripts from Sequences of the Wild Type (SP17) and Mutants M1, M2, and M1/2.

After digestion with no nuclease (lanes N) or nucleases T1 (lanes T1), T2 (lanes T2), V1 (lanes V1), or $\Phi$M (lanes $\Phi$) or with mild alkali (lanes L), products were run on 8% polyacrylamide, 7 M urea gels (see Methods). Enzyme digestions were in ionic conditions used for in vitro translation except in lanes marked SEQ, which were under denaturing conditions. The nucleotide sequence of the readable portion of the sgRNA1 transcript is shown at left of the SP17 gel. Relevant sequences of mutant transcripts are at right. Prominent bands in the mutants that are absent in SP17 and M1/2 are marked at left of the M1 and M2 gels. Bottom portions of gels of mutant transcript digestions had patterns identical to the wild type (data not shown).
by nuclease T1. Bases G66, 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 (ΔG = −1.8 kcal/mol) than wild type (ΔG = −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 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 RRLs**

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) used 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 of the mutations shown in Figure 1C are shown in Figure 4B. Histograms of the abundance of each band determined by direct measurement of radioactivity with a Phosphorimager, as well as the 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 ratios 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 an 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

---

Figure 4. In Vitro Translation of in Vitro Transcripts Encoding the CP and 17K 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 the gel in (B) by Phosphorimager and normalized for methionine content. WT, wild type.
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-to-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 six, respectively, nonviral 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 and 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, approximately half as much β-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 the 17K AUG in frame with GUS (Table 1). These constructs gave 63- and 27-fold less GUS activity, respectively, than pAdhGUS that lacked the viral insert, indicating considerable inhibition of GUS expression by the viral insertion. Similar CP-to-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, as shown in Table 1. The constant CP-to-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,
### Table 1. Expression of sgRNA1-GUS Fusions in Protoplasts Using the Adh1 Promoter

<table>
<thead>
<tr>
<th>Sequence Context</th>
<th>Plasmid</th>
<th>Experiment 1</th>
<th>Experiment 2</th>
<th>Experiment 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>GUS Units</td>
<td>GUS Units</td>
<td>GUS Units</td>
<td>CP/17K</td>
</tr>
<tr>
<td>CP AUG 17K AUG</td>
<td>CP/17K</td>
<td>CP/17K</td>
<td>CP/17K</td>
<td>CP/17K</td>
</tr>
<tr>
<td>Salmon sperm pADHGUS</td>
<td>5(11)</td>
<td>13304(16050)</td>
<td>12200</td>
<td>209(458)</td>
</tr>
<tr>
<td>UGAAUGA AAAUGG</td>
<td>pCPWT</td>
<td>486</td>
<td>0.43</td>
<td>0.45</td>
</tr>
<tr>
<td>UGAAUGA AAAUGG</td>
<td>pCPM1</td>
<td>1861</td>
<td>45.00</td>
<td>56.00</td>
</tr>
<tr>
<td>UGAAUGA AAAUGG</td>
<td>pCPM2</td>
<td>1660</td>
<td>0.45</td>
<td>0.50</td>
</tr>
<tr>
<td>UGAAUGA AAAUGG</td>
<td>pCPM1/2</td>
<td>23</td>
<td>29.57</td>
<td>25.57</td>
</tr>
<tr>
<td>UGAAUGA AAAUGG</td>
<td>pCPM3</td>
<td>1151</td>
<td>92.82</td>
<td>62.75</td>
</tr>
<tr>
<td>UGAAUGA AAAUGG</td>
<td>pCPM4</td>
<td>92</td>
<td>0.57</td>
<td>0.52</td>
</tr>
<tr>
<td>UGAAUGA AAAUGG</td>
<td>pCPM5</td>
<td>8</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td>UGAAUGA AAAUGG</td>
<td>pCPM6</td>
<td>18</td>
<td>0.14</td>
<td>0.14</td>
</tr>
<tr>
<td>UGAAUGA AAAUGG</td>
<td>pCPM7</td>
<td>9</td>
<td>3.49</td>
<td>1.50</td>
</tr>
<tr>
<td>UGAAUGA AAAUGG</td>
<td>pCPM8</td>
<td>457</td>
<td>34.65</td>
<td>26.79</td>
</tr>
<tr>
<td>UGAAUGA AAAUGG</td>
<td>pCPM9</td>
<td>13</td>
<td>20</td>
<td>20</td>
</tr>
<tr>
<td>UGAAUGA AAAUGG</td>
<td>pCPM10</td>
<td>34</td>
<td>48</td>
<td>48</td>
</tr>
</tbody>
</table>

- **Underlining indicates mutations.**
- **CP or 17K indicates which AUG is in frame with the GUS coding region. Plasmid maps and sequences of sgRNA1 inserts are shown in Figure 5. All plasmids have the Adh1 promoter and first intron.**
- **GUS units are defined as nanomoles of 4-methylumbelliferone produced per minute per milligram of protein. Background counts were defined as GUS units obtained by electroporation with salmon sperm DNA only (top line). These background units were subtracted from all subsequent samples. All GUS units represent averages from two identically treated samples.**

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 three- 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 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-to-17K AUG initiation ratio was about the same as the 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 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-to-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
The 17K AUG that reduce initiation at that site. Yet, in all of these constructs CPWT, CPM4, CPM7, and CPM9 have identical that reduced translation initiation at the 17K start site also re-
duced initiation of translation at the CP AUG. For example, trans-
lations were greater in vivo. Protoplast expression differed from
obtained in vitro, but the magnitudes of the effects of muta-
tions in the 17K AUG context on initiation differed (discussed below).

### Table 2. Expression of sgRNA1-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&lt;sup&gt;b&lt;/sup&gt;</th>
<th>GUS Units&lt;sup&gt;c&lt;/sup&gt;</th>
<th>CP/17K</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Salmon sperm pAGUS1</td>
<td>4</td>
<td>3506</td>
</tr>
<tr>
<td>CGAGUGA&lt;sup&gt;a&lt;/sup&gt;</td>
<td>17K AUG</td>
<td>p35CPWT</td>
<td>179</td>
<td>0.51</td>
<td></td>
</tr>
<tr>
<td></td>
<td>17K AUG</td>
<td>p3517KWT</td>
<td>348</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AAGUGG&lt;sup&gt;a&lt;/sup&gt;</td>
<td>17K AUG</td>
<td>p35CPM1</td>
<td>875</td>
<td>94.59</td>
<td></td>
</tr>
<tr>
<td></td>
<td>17K AUG</td>
<td>p3517KM1</td>
<td>9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GGGUGG&lt;sup&gt;a&lt;/sup&gt;</td>
<td>17K AUG</td>
<td>p35CPM2</td>
<td>786</td>
<td>0.35</td>
<td></td>
</tr>
<tr>
<td></td>
<td>17K AUG</td>
<td>p3517KM2</td>
<td>2234</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AAGUGG&lt;sup&gt;a&lt;/sup&gt;</td>
<td>17K AUG</td>
<td>p35CPM1/2</td>
<td>581</td>
<td>36.17</td>
<td></td>
</tr>
<tr>
<td></td>
<td>17K AUG</td>
<td>p3517KM1/2</td>
<td>16</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup> Underlining indicates mutations.
<sup>b</sup> CP or 17K indicates which AUG is in frame with the GUS coding region. Plasmid maps and sequences of sgRNA1 inserts are given in Figure 5. The 35 indicates that viral sequence was inserted in pAGUS1, which contains a 35S promoter.
<sup>c</sup> GUS units are defined as nanomoles of 4-methylumbelliferone produced per minute per milligram of protein. Background counts were defined as GUS units obtained by electroporation with salmon sperm DNA only (top line). These background units were subtracted from all subsequent samples. All GUS units represent averages from two identically treated samples.
<sup>d</sup> Bases at positions –20 and –26 relative to the CP AUG are changed as shown in Figure 1C.

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 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 M5. There was little or no initiation at GCG (M9). All of the above results agree with those obtained in vitro, but the magnitudes of the effects of mutations were greater in vitro. 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).

### Discussion

#### 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, 1989). The wild-type CP-to-17K ratio varies significantly with salt conditions (Dinesh-Kumar et al., 1992). The conditions used here were defined as optimal because they permitted efficient read-through 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 observed simultaneously in vitro, but initiation is detected at only one at a time 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 overlapping ORFs can cause different elongation rates in the 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; that is, less efficient elongation in the 17K ORF than the CP ORF would reduce relative levels of the 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) 3 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 reduce 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 pAGUS1 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 (ΔG = −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 kenneNYA yellow mosaic tymovirus RNA (Ding et al., 1990) in which initiation most 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 effects of A and G at these positions have been demonstrated in transgenic plants (Taylor et al., 1987). Alteration of bases −1 and −2 (McEiroy et al., 1991) or −2, −4, and −5 (Guerrineau 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. Pacing 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, because 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 are in agreement with experiments by Kozak (1991b) and with other studies on 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 (Peabody, 1987; Mehdi et al., 1990) 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 product, because the single base change from 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 initiation at the CP AUG 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 where the 60S subunit binds and protein synthesis proceeds. The CP AUG would be bypassed for two reasons: first, it is in a suboptimal context, and second, the tendency of the 5' end of sgRNA1 to form a 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), thus causing the 40S subunit to dissociate from the sgRNA. Hence, mutations that reduce secondary structure permit more frequent 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 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 (Wolin and Walter, 1988; Doohan and Samuel, 1992). The first 40S subunit behind the paused 80S complex on the 17K AUG would be located near the CP AUG (43 bases upstream). This stalled 40S subunit would have more time to interact with the suboptimal context CP AUG and would initiate translation there more frequently. 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, 1990). In support of our interpretation, Doohan and Samuel (1992) observed a pause site 25 to 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 approximately twice as many bases as the 80S subunit (Kozak and Shatkin, 1977; Doohan and Samuel, 1992). Our results led 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.

**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 (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 the time allowed for it to be bound by the 60S subunit, after which protein synthesis proceeds.

**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 virus (PLRV; luteovirus group) was sevenfold more efficient 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 al-
low 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 be-
tween 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 muta-
genic primers and their templates are listed in Table 3. 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 mul-
tiple 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 (Adh) gene followed by the first intron (Callis
et al., 1987). The Adh initiation codon in pAT13 at base 100 of the
Adh mRNA (Dennis et al., 1984; Nick et al., 1988) was changed to
AGG in pAdhGUS by PCR mutagenesis using mutagenic primer
SPDK7: 5'-GGGCAAGGGCGACCGCGG-3', which anneals to bases
95 to 112 from the transcription start site. This resulted in the ATG in
the introduced Ncol site (with or without) or any ATG introduced into
the multiple cloning site (with concomitant removal of the Ncol site)
encoding the first AUG on the Escherichia coli uidA (GUS) mRNA.

Construction of CP/17K-GUS Fusions

The wild-type and mutant coat protein (CP) or 17K open reading frames
(ORFs) containing the first 170 to 172 bases of subgenomic mRNA
(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 to 20, and
downstream primer SPDK9 (3'-TCTTGTGAAACGCTGCCCATGCCCCGGAT-5'),
which contained sgRNA1 bases 156 to 172 and an Apal site (italics)
with constructs used in vitro studies as templates. The 17K AUG
in-frame series was constructed using the upstream primer SPDK8
with the downstream primer SPDK10 (3'-TCTTGTGAAACGCTGCCCATG
CCCCGGAT-5'). This primer differs from SPDK9 only in lacking two
bases complementary to the end of the viral insert (nucleotides 171
to 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 was
translated, and the products were analyzed electrophoretically, as de-
scribed previously (Laemmli, 1970; Dinesh-Kumar et al., 1992), using

---

**Table 3. Primers and Templates Used for Mutant Construction**

<table>
<thead>
<tr>
<th>Primer Name</th>
<th>Sequence&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Bases (Numbered from 5' End of sgRNA1&lt;sup&gt;b&lt;/sup&gt;)</th>
<th>Mutant Generated (Figure 1C)</th>
<th>Template Used</th>
</tr>
</thead>
<tbody>
<tr>
<td>SPDK1</td>
<td>CTGAATTCATTCCACCTCC</td>
<td>99 to 80</td>
<td>M1</td>
<td>pSP17</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>M8</td>
<td>pM7</td>
</tr>
<tr>
<td>SPDK2</td>
<td>CTAGTGGAAGTCTGGAACT</td>
<td>79 to 61</td>
<td>M2</td>
<td>pSP17</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>M1/2</td>
<td>pM1</td>
</tr>
<tr>
<td>SPDK3</td>
<td>CTTGTCATTTAGATTGGCGGT</td>
<td>142 to 119</td>
<td>M3</td>
<td>pM1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>M4</td>
<td>pSP17</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>M6</td>
<td>pM5</td>
</tr>
<tr>
<td>SPDK4</td>
<td>CTGAATTCGTTACCACC</td>
<td>99 to 82</td>
<td>M5</td>
<td>pSP17</td>
</tr>
<tr>
<td>SPDK5</td>
<td>CTTGTCGCCTTTGATTTGC</td>
<td>142 to 124</td>
<td>M7</td>
<td>pSP17</td>
</tr>
<tr>
<td>SPDK6</td>
<td>CTTGTCGCCTTTGATTTGC</td>
<td>142 to 124</td>
<td>M9</td>
<td>pSP17</td>
</tr>
</tbody>
</table>

<sup>a</sup> Sequence is in the 5' to 3' direction with mismatches to the wild-type sequence underlined.

<sup>b</sup> All primers are complementary to the viral coding sequence. Numbering is from the 5' end of sgRNA1 (base 2769 in the viral genome).
nuclease-treated rabbit reticulocyte lysate (RRL) (Promega) in the presence of 35S-labeled 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 Phosphorimager (model 400E; Molecular Dynamics, Sunnyvale, CA). The ratios of CP-to-17K products were determined after adjusting for the number of methionine residues in the products (four in CP, three in the 17K ORF).

### Secondary Structure Mapping

Unlabeled sgRNA1 transcripts containing nucleotides 1 to 217 were transcribed from plasmids pWT, pM1, pM2, and pM1/2 after linearizing with Sall. 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 ~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 7 M urea, 30 mM EDTA, 0.05% bromophenol blue, and 0.05% xylene cyanol solution and freezing at −80°C. Products were electrophoresed on 8% polyacrylamide, 7 M urea gels. For size markers, transcripts were sequenced under denaturing conditions using 1 unit of T1 or 0.5 units of V1 (cuts A>B) or 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 ~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 7 M urea, 30 mM EDTA, 0.05% bromophenol blue, and 0.05% xylene cyanol solution and freezing at −80°C. Products were electrophoresed on 8% polyacrylamide, 7 M urea gels. For size markers, transcripts were sequenced under denaturing conditions using 1 unit of T1 or 0.5 units of V1 (cuts A>B) nucleases, 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 105 protoplasts and exactly 0.1 µg of plasmid DNA or carrier DNA (salmon sperm) were mixed in electroporation buffer (Fromm et al., 1987). Samples were electrophoresed on 8% polyacrylamide, 7 M urea gels. For size markers, transcripts were sequenced under denaturing conditions using 1 unit of T1 or 0.5 units of V1 (cuts A>B) nucleases, as described previously (Miller and Silver, 1991), or they were partially hydrolyzed with sodium carbonate.

### REFERENCES


Control of start codon choice on a plant viral RNA encoding overlapping genes.
S P Dinesh-Kumar and W A Miller
Plant Cell 1993;5:679-692
DOI 10.1105/tpc.5.6.679

This information is current as of June 22, 2017

Permissions


eTOCs

Sign up for eTOCs at:
http://www.plantcell.org/cgi/alerts/ctmain

CiteTrack Alerts

Sign up for CiteTrack Alerts at:
http://www.plantcell.org/cgi/alerts/ctmain

Subscription Information

Subscription Information for The Plant Cell and Plant Physiology is available at:
http://www.aspb.org/publications/subscriptions.cfm

© American Society of Plant Biologists
ADVANCING THE SCIENCE OF PLANT BIOLOGY