JOURNAL OF VIROLOGY, May 1998, p. 4157–4169 Vol. 72, No. 5
0022-538X/98/$04.00+0


# Forced Evolution Reveals the Importance of Short Open Reading Frame A and Secondary Structure in the Cauliflower Mosaic Virus 35S RNA Leader

MIKHAIL M. POOGGIN,[1,2] THOMAS HOHN,[1]* AND JOHANNES FÜTTERER[3]

*Friedrich Miescher Institute, CH-4002 Basel,[1] and Institute for Plant Sciences, ETH Zentrum, CH-8092 Zurich,[3] Switzerland, and Centre "Bioengineering," Russian Academy of Sciences, 117312 Moscow, Russia[2]*

Received 19 November 1997/Accepted 4 February 1998

**Cauliflower mosaic virus pregenomic 35S RNA begins with a long leader sequence containing an extensive secondary structure and up to nine short open reading frames (sORFs), 2 to 35 codons in length. To test whether any of these sORFs are required for virus viability, their start codons were mutated either individually or in various combinations. The resulting viral mutants were tested for infectivity on mechanically inoculated turnip plants. Viable mutants were passaged several times, and the stability of the introduced mutations was analyzed by PCR amplification and sequencing. Mutations at the 5′-proximal sORF A and in the center of the leader resulted in delayed symptom development and in the appearance of revertants. In the central leader region, the predicted secondary structure, rather than the sORF organization, was restored, while true reversions or second-site substitutions in response to mutations of sORF A restored this sORF. Involvement of sORF A and secondary structure of the leader in the virus replication cycle, and especially in translation of the 35S RNA via ribosome shunting, is discussed.**

Cauliflower mosaic virus (CaMV) is the prototype member of the caulimoviruses, which together with the plant badnaviruses and the animal hepadnaviruses form the family of pararetroviruses (33, 35, 50). The replication of the circular, double-stranded viral DNA involves reverse transcription of a terminally redundant pregenomic RNA which is transcribed from extrachromosomal DNA copies of viral genomes in the nucleus of an infected cell. It is very probable that for all viruses of this family, reverse transcription occurs on RNA molecules that have been packaged into virion or virionlike particles (3, 43). In addition to their function as replicative intermediates, the pregenomic RNAs also serve as complex, polycistronic mRNAs and in some cases as substrates for RNA splicing (for reviews, see references 37 and 50).

For CaMV, the pregenomic RNA (35S RNA) has a length of about 8.2 kb and begins with a highly structured leader sequence of about 600 nucleotides (nt) which contains up to nine short open reading frames (sORFs) with coding potentials of 2 to 35 amino acids. The leader sequence is followed by seven longer ORFs which (with the exception of ORF VI, which is encoded by the subgenomic 19S RNA) are translated either from the 35S RNA or from its spliced derivatives, which also contain most parts of the leader sequence. Translation of these complex RNAs involves a ribosome shunt that allows the bypassing of inhibitory elements in the center of the leader (19, 22) and translation reinitiation that requires the activity of a CaMV-encoded translational transactivator (4, 20, 36, 54).

A number of important steps in the CaMV infection cycle are regulated or influenced by elements located within the leader sequence. The primer binding site (PBS) for reverse transcription is located at the 3′ end of the leader, and template switching during reverse transcription must occur within the terminal redundancy, i.e., the first 190 nt of the leader (3). A sequence near the 5′ end was shown to enhance expression from the 35S promoter at either the transcriptional or the posttranscriptional level (14, 16, 48). The leader greatly influences translation of ORFs downstream of it (2, 18), and sequence regions with an inhibitory effect on translation and regions which at least partially alleviate this inhibition have been recognized (19, 22). The leader also contains the viral polyadenylation signal, which has to be used inefficiently in the leader to allow synthesis of full-length 35S RNA (53). Also, an inefficient splice donor site is located toward the end of the leader (36). The computer-predicted hairpinlike structure of the leader (17) has been confirmed in vitro (30). A sequence located at the top of this hairpin is highly conserved among all caulimoviruses (17) and in rice tungro bacilliform virus (RTBV) (29); its function is unknown at present. Structural analysis of the leader revealed the potential for dimerization (30), which might be important for reverse transcription (cf. true retroviruses). Also, by analogy with related animal viruses, it is likely that at least parts of a packaging signal for encapsidation of the RNA for reverse transcription are located within the leader sequence.

The importance of some leader regions for isolated steps like translation (19, 22) and polyadenylation (51, 52) has been analyzed in transient expression systems, but these studies did not clarify whether structural features of the leader or the primary sequence were responsible for the observed effects. Such experiments also do not allow investigation of the correct interplay of the different functions in the viral infection cycle. An alternative approach to elucidate structure and sequence requirements is the in vivo analysis of mutant viruses. Even though the lack of a reliable single-cell infection system for caulimoviruses precludes the determination of the precise step influenced by a given mutation, the whole-plant system with several subsequent infection cycles allows the study of the evolution of the best-adapted sequences in a small-scale experiment. Evolution is enhanced by the error-prone process of replication by reverse transcription, which generates sequence variants as soon as even a limited amount of replication occurs (49). In fact, such analyses provided the first data indicating

* Corresponding author. Mailing address: Friedrich Miescher Institute, P.O. Box 2543, CH-4002 Basel, Switzerland. Phone: 41 61 697 72 66. Fax: 41 61 697 39 76. E-mail: thomas.hohn@fmi.ch.

TABLE 1. Individual sORF mutations

| sORF | Original start-codon context[a] (strength[b]) | Mutation | | Restriction site created |
|---|---|---|---|---|
| | | Designation | Position[c] | |
| | 61 72 | | | |
| A | ATA ATA **ATG** TGT GAG TAG (M) | $M_{A(TTG)}$ | A61T | *Ssp*I |
| | | $M_{A(TTT)}$ | ATG61TTT | *Ssp*I |
| | | $M_{A(ACG)}$ | T62C | *Afl*III |
| | | $M_{A(ACA)}$ | TG62CA | *Nsp*I |
| | | $M_{A(gcT)}$[d] | TG64GC | |
| | | $M_{A(TGg)}$[d] | T66G | |
| | | $M_{A(stop)}$[d] | G72C | *Sca*I |
| B | TCG CTC **ATG** TGT (W) | $M_{B(GAG)}$ | AT114GA | *Xho*I |
| C | CTT AGT **ATG** TAT (M) | $M_{C(ACT)}$ | TG144CT | *Sca*I |
| D′ | GCC CAG **ATG** CCG (W) | $M_{D'(CGG)}$ | AT292CG | *Eag*I |
| D | GAA AAG **ATG** CTA (M) | $M_{D(GCG)}$ | AT344GC | *Hae*II |
| E | AAT ATC **ATG** AAG (M) | $M_{E(CTT)}$ | ATG398CTT | *Afl*II |
| E′ | GAA GAC **ATG** GAA (S) | $M_{E'(AAG)}$ | T428A[e] | |
| E″ | AAG ACG **ATG** GAA (S) | $M_{E''(TCG)}$ | AT467TC | *Aat*II |
| F | AAG GGA **ATG** CTT (M) | $M_{F(AAG)}$ | T519A | *Hin*dIII |

[a] Start codon in boldface.
[b] Strength of context according to Kozak (38). W, weak; M, moderate; S, strong.
[c] Numbering is from the 5′ end of the strain CM4-184 35S RNA (see Fig. 4); when a doublet or triplet is mutated, the number of the 5′-proximal nucleotide is given.
[d] The sORF A ATG remains intact.
[e] Occurs in strain S and in other CaMV strains lacking sORF E′ (Fig. 1).

that translation of CaMV follows rules that are highly unusual for eukaryotic cells (9, 10, 55).

Here we describe an analysis of mutations of the sORFs in the leader. In a number of systems, fine regulation of gene expression is accomplished by sORFs (for reviews, see references 26, 27, and 31), and it also has been suggested that RNA usage for packaging or translation is regulated by sORFs (12, 13, 57). Testing mutations of all sORFs in the leader of the CaMV 35S RNA, either alone or in various combinations, for their effects on infectivity in repeated infection cycles revealed that the structure in the center of the leader and a translation event at sORF A are major determinants of viral competitiveness. These data provide new insight into the process of ribosome shunting, which appears to be a general mechanism utilized by different viruses, including RTBV (23), adenovirus (59), Sendai virus (7), and budgerigar fledgling disease virus (41).

## MATERIALS AND METHODS

**Virus and plants.** CaMV strain Ca540, which was used as a reference wild-type virus in this study, is not a natural isolate. Its genome consists of the *Bam*HI (1927)-*Bst*EII (127) portion of the strain S genome (which includes the leader and which was modified by the insertion of a 9-bp *Bam*HI-*Sma*I linker [GATC CCGGG] behind the first ATG of ORF V [3634 to 3636]) and the complementing *Bst*EII (127)-*Bam*HI (1927) portion of the genome of strain CM4-184 (25, 34) bearing a natural deletion of 420 nt (spanning position 1387 to 1806) within the aphid transmission factor gene. (The numbering in this section is in accordance with that of the strain S sequence [15].) Neither the artificial insertion nor the natural deletion interferes with viral infectivity. However, the latter feature destroys aphid transmissibility. Therefore, the possibility of cross-contamination by insect-transmitted viruses during parallel analyses of different mutants could be excluded. In all mutant viruses, the S leader of Ca540 has been replaced with the strain CM4-184 leader containing the sORF mutations. In a control experiment, it was established that a wild-type Ca540 virus and its derivative with an unmodified CM4-184 leader had the same infection characteristics (i.e., latency period and severity of symptoms). The sequence differences between the leaders (see Fig. 1) allowed a clear differentiation between mutated or reverted viruses and the wild-type control virus.

Plasmid pCa540 contains the strain Ca540 DNA genome, linearized at a unique *Sal*I site (4837) and cloned into *Sal*I of pUC8V209, a derivative of pUC8 lacking the *Sma*I site (the *Sal*I linker GGTCGACC was inserted between the *Sma*I and *Hin*dIII sites of pUC8).

The wild-type and mutant viruses were propagated in turnip (*Brassica rapa-rapifera* cv. Just Right) in a phytobox with illumination for 16 h/day at 20 or 26°C. To mechanically inoculate plants, viral DNA was amplified in *Escherichia coli* DH5α and excised from the plasmid vector with *Sal*I. Then, the digested DNA was ethanol precipitated and dissolved in sterile water to a concentration of 1 μg/μl (unless otherwise stated). In the presence of a small amount of celite abrasive, 25 μl of inoculum was gently rubbed with gloved fingers onto the first and second true leaves of three-leaf-stage seedlings. Passages of infectious viruses to new plants were performed by homogenizing 0.2 g of infected tissue of young leaves in 0.5 ml of 10 mM Tris-HCl (pH 7.5)–1 mM EDTA in a sterile mortar, clarifying the extract by centrifugation, and applying 50 μl of the supernatant onto seedlings as described above. Special care was taken to avoid cross-contamination by different isolates during all procedures. Typically, symptoms appeared after 20 days for infections with cloned viral DNA and after 14 days for infections with sap from infected plants. Observed delays are expressed as a percentage of the time required for infection by the wild-type virus.

**Construction of virus mutants.** All point mutations were introduced by the method of PCR-mediated site-directed mutagenesis described by Galvez and DeLumen (24). Briefly, with a pair of primers flanking the leader and a pair of complementary overlapping primers containing a mutation, two overlapping PCR products were produced. Subsequent PCR ligation of these products, in the presence of the flanking primers, yielded the whole leader sequence containing the desired mutation. The latter PCR product either was used as a template for further mutation or was cloned back into the original template plasmid, pV322. pV322 contains the fragment of the strain CM4-184 genome comprising the 35S RNA promoter, the leader, ORF VII, and the beginning of ORF I. (pV322 is described as pCAT1 in reference 4.) Flanking primers 1 and 2 corresponded, respectively, to positions 7331 to 7355, including the *Eco*RV site within the 35S RNA promoter, and positions 137 to 114 of the complementary strand, including the *Bst*EII site within ORF VII. The *Eco*RV and *Bst*EII sites, both unique in pV322, were used for the subcloning of PCR products. All mutations created as described above are shown in Table 1. Combinations of different mutations, used to produce the ATG-less leader mutant $M_{ALL}$ and some other mutants shown in Table 2, were constructed by taking advantage of two restriction sites, *Bsm*BI (7689) and *Alw*NI (7870), which divide the leader into three parts.

Mutant versions of the leader were transferred to the viral genome by replacing the *Eco*RV (7343)-*Bst*EII (127) fragment of pCa540 with the corresponding one from the derivatives of pV322.

**DNA preparation and PCR, cloning, and sequencing of viral progeny from infected plants.** Total DNA from infected plants was prepared by a fast method described by Padegimas et al. (45). Small pieces of young infected leaves were ground in liquid nitrogen, and about 50 to 100 mg of plant tissue powder was transferred to an Eppendorf tube containing 850 μl of TELT extraction buffer (50 mM Tris-HCl [pH 8.0], 10 mM EDTA, 4 M LiCl, 4% [wt/vol] Triton X-100). Extraction was carried out by shaking the tube for 15 min at 37°C. Then RNA was precipitated by incubation on ice for 5 min followed by centrifugation in an Eppendorf centrifuge at 12,000 rpm for 5 min. To 700 μl of supernatant was added an equal volume of phenol-chloroform (1:1), and then this mixture was shaken for 15 min at 37°C. The centrifugation step was repeated, and 500 μl of the resulting upper phase was transferred to a new tube. The DNA was precip-

TABLE 2. Combinations of individual sORF mutations[a]

| Mutant | Individual sORF mutations combined | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | A | B | C | D′ | D | E | E′ | E″ | F |
| $M_{AB}$ | $M_{A(TTG)}$ | × | | | | | | | |
| $M_{AC}$ | $M_{A(TTG)}$ | | × | | | | | | |
| $M_{ABC}$ | $M_{A(TTG)}$ | × | × | | | | | | |
| $M_{DE}$[b] | | | | | × | × | × | | |
| $M_{ABCDE}$[b] | $M_{A(TTG)}$ | × | × | | × | × | × | | |
| $M_{D'DE}$[b] | | | | × | × | × | × | | |
| $M_{E''F}$ | | | | | | | | × | × |
| $M_{ALL+D'}$ | $M_{A(TTG)}$ | × | × | | × | × | × | × | × |
| $M_{ALL}$ | $M_{A(TTG)}$ | × | × | × | × | × | × | × | × |
| $M_{ALL(TTT)}$ | $M_{A(TTT)}$ | × | × | × | × | × | × | × | × |
| $M_{ALL+A(ACA)}$ | $M_{A(ACA)}$ | × | × | × | × | × | × | × | × |
| $M_{ALL+A(gcT)}$ | $M_{A(gcT)}$[c] | × | × | × | × | × | × | × | × |
| $M_{ALL+A(TGg)}$ | $M_{A(TGg)}$[c] | × | × | × | × | × | × | × | × |
| $M_{ALL+A(stop)}$ | $M_{A(stop)}$[c] | × | × | × | × | × | × | × | × |

[a] Individual sORF mutations are described in Table 1.
[b] Carries an in-frame ATG within sORF F (GTA545ATG), originating from strain S.
[c] The sORF A ATG remains intact.

itated with 2 volumes of ethanol, washed with 1 ml of 70% ethanol, air dried, and dissolved in 100 μl of sterile bidistilled water.

A 0.4- to 2-μl volume of total DNA (about 20 to 100 ng) was taken for PCR analysis. PCR was carried out in a 25-μl reaction mixture (1× Vent polymerase buffer [New England Biolabs] supplemented with additional $MgSO_4$ [to 6 mM], 200 μM each deoxynucleoside triphosphate, 30 pmol of each flanking primer [no. 1 and no. 2; see above], and 0.125 U of Vent polymerase [New England Biolabs]) for 35 cycles of 15 s at 95°C, 30 s at 65°C, and 30 s at 72°C in a DNA thermal cycler (Perkin-Elmer Cetus). The resulting PCR product of the expected size (834 nt) was extracted from low-melting-point agarose, trimmed with *Eco*RV and *Bst*EII (the sites located within primers 1 and 2, respectively), purified by agarose gel electrophoresis, and cloned into the *Eco*RV and *Bst*EII sites of pV322. For the introduction of the point mutations (see above), the same PCR conditions and cloning procedure were used. Sequencing of mutant versions of the leader showed that no additional substitutions occurred during these manipulations.

Plasmid DNA was sequenced either manually, using a Sequenase 2.0 kit (U.S. Biochemicals), or automatically, with AmpliTaq FS DNA polymerase and dye terminators, on an ABI 377 DNA sequencer (Perkin-Elmer Applied Biosystems). In the latter case, one run of the reaction with primer 1, annealing 80 nt upstream of the 35S RNA leader, allowed us to read reliably the whole leader sequence (612 nt in length).

**Computer prediction of RNA secondary structure.** RNA secondary structure was predicted with the MFold program (Wisconsin Package, version 9.0; Genetics Computer Group, Madison, Wis.) based on the method of Zuker and Stiegler (61). The first 912 nt of the 35S RNA, comprising the leader and ORF VII, was folded at 20 or 26°C to fit our in planta experimental conditions (see the legend to Fig. 2 for more details).

## RESULTS

**Comparison of sORFs present in the 35S RNA leaders of CaMV isolates.** The 612-nt 35S RNA leader of strain CM4-184 that was mutated in this study harbors nine sORFs, designated A, B, C, D′, D, E, E′, E″, and F (the start codons of sORFs E′ and E″ are located within sORF E). Comparison of 11 different CaMV isolates clearly showed that most of these sORFs are well conserved (Fig. 1). Alignment of the leader sequences revealed nucleotide substitutions, short insertions, and deletions at 80 positions (13%); 36 of them fell in the sORFs. However, in only 10 cases did two or more strains have the same change at a given position within a sORF.

sORFs A and E″, both three codons in length, are present in all strains and are not affected by any substitution. sORFs B, C, D, and E appeared to be less conserved, and some sORFs (including D′ and E′) are present in only some strains (Fig. 1).

sORF F is of particular interest since it is located immediately upstream of the first long ORF (VII) in a region required for ribosome shunting (22), close to the PBS. Translation of this sORF might influence the function of these latter elements and translation of ORF VII. However, sORF F turned out to be the most variable. It overlaps ORF VII by either 14 or 8 nt or not at all. In one strain (S), it contains an internal ATG codon, and in another strain (CMV1), it is missing altogether (Fig. 1). This variability suggests that sORF F is not very important for virus viability.

**Strategy for studying the roles of the sORFs.** To study the requirement for translational events at the sORFs in the leader and/or their potential products, we mutated their ATG start codons. To account for a possible redundancy of sORF functions, we not only mutated every sORF individually but removed them in different combinations from the 5′ region (A, B, and C), the central region (D′, D, E, and E′), and the 3′ region (E″ and F) of the leader. Furthermore, an ATG-less leader lacking all sORFs was also constructed. All individual point mutations and combinations thereof are summarized in Tables 1 and 2, respectively. In most cases, the sORF muta-

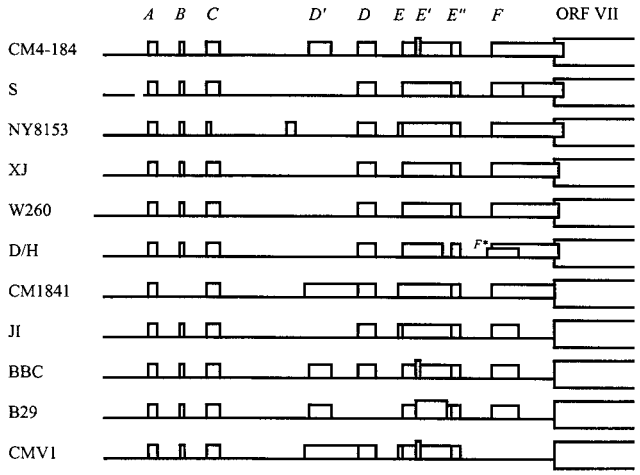


FIG. 1. Comparison of sORFs in the 35S RNA leaders of different CaMV strains and isolates found in the GenEMBL database. The leader sequence preceding the first large ORF (ORF VII) is depicted as a thick line; the sORFs (A to F) are indicated by boxes. Vertical lines define positions of sORF start and stop codons.

| Mutant | Inoculation with DNA at 20°C: Delay | Inoculation with DNA at 20°C: Reversions | Temp | 1st passage: Delay | 2nd passage: Delay | 2nd passage: Reversions and second site substitutions | 3rd passage: Delay | 3rd passage: Reversions and second site substitutions | 4th passage: Delay | 4th passage: Reversions and second site substitutions |
|---|---|---|---|---|---|---|---|---|---|---|
| $M_{ALL}$ | 50% (2/2)** | 13* none****<br>1 G63C ($M_{A(TTG)}$***) | 20°C | 50% (2/2) | 29% (2/2) | 3* c292T****<br>2 t61A+c292A<br>1 G249A+c292A<br>1 G249A+c292A+A597T<br>1 G249A+c292T+C4G<br>1 c292A+g344A | 14% (2/2) | 2* A60G+G249A+c292T****<br>2 A60G+G249A+c292T+T511C<br>1 A60G+G249A+c292T+C13G<br>1 A60G+G249A+c292A+g344A<br>1 A60G+G249A+c292A+g344A+A597(4A)<br>1 A60G+G249A+c292A+G566A<br>1 t61A+c292A | 0% (2/2) | 4* t61A+T247G+c292A+g344A****<br>1 t61A+g114A+T247G+c292A+g344A+G-2A<br>2 t61A+T247G+c292A+g344A+tc467AT<br>1 t61A+T247G+c292A+g344A+tc467AT+TT590CA |
| | | | 26°C | 71% (2/2) | 36% (2/2) | 5 G249A+c292A<br>1 G249A+c292A+G-46A<br>1 t61A+G249A+c292A<br>1 G249A+c292T | 14% (2/2) | 2 A60G+T247G+c292T<br>1 A60G+T247G+c292T+C209T<br>1 A60G+T247G+c292A+G343A<br>1 A60G+G249A+c292A<br>1 A60G+G249A+c292A+g344A<br>1 t61A+G249A+c292A+g344A<br>1 t61A+G249A+c292A+t467A<br>1 t61A+G249A+c292A<br>1 T59(2A)+t61A+g114A+G249A+c292A<br>1 A60G+del(190-492) | 14% (2/2) | 3 A60G+G249A+c292A+g344A<br>1 A60G+G249A+c292T+g344A<br>1 A60G+T247G+c292T+G343A<br>1 A60G+T247G+c292T+G343A+del(456-460)<br>1 A60G+T247G+G249A+c292T+g344A<br>1 A60G+T247G+G249A+c292T<br>1 t61A+T247G+c292A |
| $M_{ABC}$ | 35% (1/1) | 12 none | 20°C | 36% (2/2) | 14% (2/2) | 1 A60G+C508T<br>1 A60G+T119A<br>1 A60G+T156C<br>1 A60G+C6T<br>1 A60G<br>1 A57G All: +A391T | 0% (2/2) | 3 A60G+C508T<br>1 A60G+T119A+C508T<br>1 A60G+T119A+C508T+del(28-45)<br>1 A60G+T119A<br>1 A60G+T119A+T156G<br>2 A60G+G118A+C508T<br>1 t61A+G118A+C508T<br>1 A57G All: +A391T | 0% (2/2) | 5 A57G+c144G<br>1 A57G+T64A+c144G<br>2 A60G+c144G<br>2 A60G All: +A391T |
| | | | 26°C | 36% (2/2) | 0% (2/2) | 4 A60G<br>1 A60G+T-77A<br>1 A57G All: +A391T | 14% (2/2) | 2 A60G<br>1 A60G+c144T+T175(2A)<br>1 A60G+c144T+T175A<br>1 A57G+c144G All: +A391T | 36% (2/2) | 2 A60G+T119A<br>1 A60G+T119A+C508T<br>1 A60G+del(503-507)+C508T+T528C<br>1 A60G+G118A<br>1 A60G+G118A+TG399CA($M_E$***)<br>1 A60G<br>1 A60G+T590A<br>1 A57G+g144A All: +A391T |
| $M_{A(TTG)}$ | 10% (1/1) | 11 none | 20°C | 14% (2/2) | 0% (2/2) | 7 A60G<br>1 A60G+C-55T<br>2 A60ATGGTA | 0% (2/2) | 4 A60G<br>1 t61A | not analysed | |
| | | | 26°C | 0% (2/2) | 0% (2/2) | 6 A60G<br>2 A60ATGGTA<br>1 A57G<br>1 none | 0% (2/2) | 6 t61G<br>1 A60G | not analysed | |

* number of clones with a given sequence
** infected per inoculated plants
*** mutation in a sORF start codon
**** nucleotide changes: (i) numbering is from the 5'-end of the strain CM4-184 35S RNA (see Fig. 4); (ii) "-" indicates positions upstream of the transcription start site; (iii) small case letters indicate nucleotides introduced with the original sORF mutations; (iv) "del" stands for a deletion.

Colours show first and second site reversions in: sORF A (see Fig. 5) sORF B stem section 2 stem section 3 (see Fig. 4) bowl sORF E'

FIG. 2. Restoration of leader features on passage of sORF mutants in turnip plants.

tions introduced new restriction sites (Table 1), allowing them to be followed at different steps. Our previous results had suggested that sORF A might play an important role in the ribosome shunt mechanism (22, 48a). The conservation of this sORF among all known CaMV strains supports this hypothesis. To evaluate critical parameters of a translational event at this sORF and to determine if its 3-amino-acid product is required, several additional mutations to the sORF A sequence were created (Table 1). Mutant leaders were introduced into a non-aphid-transmissible recombinant strain of CaMV.

Almost all of the mutant viruses, including the one with a sORF-less leader, were infectious to turnip plants, resulting eventually in the development of symptoms identical to those caused by wild-type CaMV. However, a delay in symptom appearance of between 10 and 200% (with respect to the time required for symptom appearance with the wild-type virus) was observed in some cases, suggesting a deleterious effect of the respective mutations. Because suboptimal viral genomes should be outcompeted by better-adapted genomes generated during error-prone transcription or reverse transcription, virus populations from plants infected with mutant viruses were propagated by serial passage to new plants. Since the mutations might also influence structural features of the leader and thus be affected by temperature, infections were performed at both 20 and 26°C.

Routinely, on every passage, a mixture of saps from two plants infected with the same mutant virus was applied onto two new plants. The same mixture was used for analysis of the viral progeny by PCR amplification, cloning, and sequencing of the leader. In most cases, any originally observed delay of symptom appearance vanished within a few passages, suggesting that more-efficient revertants had outcompeted the original mutants. The appearance of revertants was confirmed by sequencing of individual clones from the progeny of mutant viruses. The results of infectivity studies and sequencing analyses are summarized in Fig. 2 and 3. In the following sections, we describe in detail how the sORF mutations affected the viability of the virus and which types of reversions (if any) evolved in response to the individual and combined mutations in different regions of the leader.

**Analysis of an ATG-free leader.** Removal of all nine sORFs from the leader (mutant $M_{ALL}$) resulted in a delay of symptom appearance of 50% in the primary infection cycle. The leader sequences of 14 clones isolated from the infected plants were determined, but a sequence change (from G to C) was observed in only one case at one position (nt 63). The latter mutation was not found in any later generation obtained by further inoculation of plants with this virus pool and therefore seems to be rather accidental and not to confer a selective advantage. This result indicates that CaMV without any ATG codons in the leader sequence of the 35S RNA is viable. All *cis*-acting signals required for replication must still be functional in this mutant. It is less certain that all signals required for proper gene expression are also still functional, since it is

| Mutants | Inoculation with DNA at 20°C | | Temp | 1st passage | 2nd passage | |
|---|---|---|---|---|---|---|
| | Delay | Reversions | | Delay | Delay | Reversions and second site substitutions |
| $M_{E''}$ | 0% (2/2)** | not sequenced | 20°C | 14% (2/2) | 14% (2/2) | 2* none**** |
| | | | 26°C | 0% (2/2) | 0% (2/2) | 2 none |
| $M_F$ | 0% (2/2) | not sequenced | 20°C | 0% (2/2) | 0% (2/2) | 2 none |
| | | | 26°C | 0% (2/2) | 0% (2/2) | 2 none |
| $M_{E''F}$ | 0% (2/2) | not sequenced | 20°C | 0% (2/2) | 0% (2/2) | 1 none<br>1 a519C($M_F$***) |
| | | | 26°C | 0% (2/2) | 0% (2/2) | 2 none |
| $M_{D'DE}$ | 25% (2/2) | not sequenced | 20°C | 14% (2/2) | 0% (2/2) | 2 T247G+c292A<br>1 c292A<br>2 c292T All:+G515A |
| | | | 26°C | 14% (2/2) | 0% (2/2) | 4 T245A+g293A+g344A<br>1 T245A+g293A+g344A+T573C<br>1 T245A+g293A+g344A+G343A<br>1 T245A+g293A+Tt399CA<br>2 del(293-301)+T396C All:+G515A |
| $M_D$; $M_E$;$M_{DE}$ | 0% (1/1) each | not sequenced | | | | |
| $M_{ABCDE}$ | 35% (1/1) | 11 none | 20°C | 71% (2/2) | 57-100% (2/2) | 5 t61A+A597(2A)<br>1 t61A+A597(2A)+C574T<br>5 t61A+A597(3A) |
| $M_{ALL+D'}$ | 35% (1/1) | 4 none | 20°C | 71% (2/2) | 57-71% (2/2) | 6 A60G<br>1 A60G+A422G<br>2 A60G+C76T<br>1 A60G+C76T+C583T+ ATGAA612AGTT($M_{ORFVII}$***) |
| $M_{AB}$ | 40% (2/2) | not sequenced | 20°C | 0% (2/2) | 0% (2/2) | 3 A60G |
| | | | 26°C | 29% (2/2) | 0% (2/2) | 3 t61A |
| $M_{AC}$ | 40% (2/2) | not sequenced | 20°C | 0-36% (2/2) | 7% (2/2) | 2 t61A |
| | | | 26°C | 29-71% (2/2) | 14% (2/2) | 3 A60G |
| $M_{A(ACG)}$ | 10% (1/1) | 11 none | 20°C | 14% (2/2) | 0% (2/2) | 4 c62T |
| | | | 26°C | 0% (2/2) | 0% (2/2) | 2 c62T<br>1 c62T+G-1C |
| $M_{A(TTT)}$ | 20-45% (2/2) | not sequenced | 20°C | 14-29% (2/2) | 0% (2/2) | 3 A57G<br>1 A57G+C444T<br>1 A57G+T591(4A)<br>1 T48G<br>1 T48G+G343A<br>1 T48G+C-50T |
| $M_{ALL(TTT)}$ | 35-100% (2/2) | not sequenced | 20°C | 50% (2/2) | 14-71% (2/2) | 7 A60G+T64C+g293A+g344A<br>1 A60G+T64C+g293A+g344A+ A597(3A)<br>1 T64C+g293A+g344A |
| $M_{A(gcT)}$ | 0-10% (2/2) | not sequenced | 20°C | 0% (2/2) | 0% (2/2) | 8 no reversion<br>1 C2G+A624C |
| $M_{ALL+A(gcT)}$ | 0-10% (2/2) | not sequenced | 20°C | 0% (2/2) | 0% (2/2) | 2 c292T+C298T+g344A<br>1 G249A+c292T<br>1 G249A+c292T+T396C<br>1 g293A+g344A+del377+C586T |
| $M_{A(TGg)}$ | 0-20% (2/2) | not sequenced | 20°C | 14-29% (2/2) | 0-57% (2/2) | 8 no reversion<br>1 G-42A<br>1 C330T |
| $M_{ALL+A(TGg)}$ | 55-100% (2/2) | not sequenced | 20°C | 71% (2/2) | 57-71% (2/2) | 5 g293A+g344A<br>1 g293A+g344A+T372A<br>1 A133G + g293A + g344A<br>1 c292A |
| $M_{A(ACA)}$***** | 200% (1/4) | 5* A60G****<br>1 A60G+ A140G+T528A<br>1 T64G<br>5 none | 20°C | 0% (2/2) | 3 A60G<br>1 A60G+del(61-63)<br>1 A60G+del(61-63)+T144G($M_C$***) | |
| $M_{ALL+A(ACA)}$ | 175% (1/2) | not sequenced | 20°C | 71% (2/2) | 7 A60G+T247G+g293A<br>2 A60G+T247G+g293A+g344A<br>1 A60G+T247G+g293A+ duplication of 448-459 +T460A | |
| $M_{A(stop)}$***** | 200% (1/4) | 14 - c72G | 20°C | 0% (2/2) | not sequenced | |
| $M_{ALL+A(stop)}$ | (0/2) | | | | | |

* number of clones with a given sequence
** infected per inoculated plants
*** mutation of a sORF start codon
**** nucleotide changes (see Fig. 2)
***** a higher concentration of DNA (5 µg/µl) was used for primary inoculation

Colours show first and second site reversions in: sORF A (see Fig. 5); stem section 3; bowl (see Fig. 4)

FIG. 3. Restoration of leader features on passage of other sORF mutants in turnip plants.

possible that a small proportion of reverted genomes (<7%) might have contributed protein products in *trans*.

To study further the eventual evolution of more-efficient viral genomes, new plants were inoculated with the sap from the primary-infected plants at either 20 or 26°C. After the first passage, there was still a significant delay in symptom appearance of 50% (20°C) or 71% (26°C). Upon further passaging, the delay time gradually decreased; after the fourth passage at 20°C, no difference from that of the wild-type virus was discernible. Already after the second passage, all recovered viral genomes contained sequence differences compared to the original inoculum (Fig. 2). In all recovered mutants, the originally altered base of the start codon of sORF D′ had been changed (from a C to an A or a T at position 292 [c292A or c292T]). In no case was a new start codon generated at this position, suggesting that either a primary sequence motif or a structural determinant was affected by the original mutation. This reversion was in almost all cases accompanied by one or two additional mutations which created one or two upstream sORFs. One of these mutations (G249A) created a sORF of nine codons that does not exist in any CaMV strain. It also slightly increases the stability of the predicted secondary structure by changing a GU base pair to AU within stem section 3 (Fig. 4) and thus may partially compensate for the reduced stability of this stem due to the original mutation in the start codon of sORF E (Fig. 4). The other mutation (t61A) restored the wild-type sORF A.

While the G249A mutation was present in only a subset of the viral genomes, a reverted sORF A was present in all viral genomes recovered after further propagation. Besides the true reversion to the wild-type sORF A sequence, a second-site substitution, generating a novel ATG upstream of the original sORF A start codon, was frequently observed (A60G in Fig. 2 and 5). After the third and fourth passages, additional mutations in stem section 3 (T247G), increasing the stability of the stem (Fig. 4), and the mutated start codon of sORF D (g344A, not restoring a start codon) became predominant. A tendency for the occurrence of reversions in the region of the start codon for sORF E″ existed (Fig. 2 and 4), but all other original mutations appeared to be stable. Very few mutations in other regions of the CaMV 35S RNA leader were observed in the sequences of 68 independent clones (Fig. 2), suggesting either that the replication mechanism is not so error prone after all or that selection for an active leader is tight and allows very few changes.

Our analysis indicated that besides sORF A, no other sORF is required for optimal leader function. Other revertants seemed to restore structural features but not sORFs. At 20°C, reversion of sORF A, strengthening of stem section 3, and some changes in the sORF D′ and D ATG regions were sufficient to produce a fully competent viral leader sequence which no longer caused a delay in symptom production. After the fourth passage at 20°C, the virus pool appeared to be quite uniform; all clones analyzed contained a wild-type sORF A, the T247G mutation in stem section 3, and the c292A and g344A mutations in the start-codon regions of sORFs D′ and D, respectively. A true reversion of the sORF E″ ATG codon was evident in some clones (tc467AT [Fig. 2 and 5]). Strikingly, the sORF-producing mutation G249A was missing in this population, although it was present in the inoculum (Table 2). The pool recovered at 26°C was more variable and contained individuals with the additional sORF; in this case, no E″ reversion was detected. It is unclear whether these differences represent different selection criteria at the two temperatures and whether they cause the slight delay (14%) in symptom appearance at 26°C.

**Mutations at the 3′ end of the leader.** As a first step in investigating which mutations contribute to the 50% delay in symptom appearance associated with $M_{ALL}$, the start codons of sORFs E″ and F were mutated individually or in combination (yielding $M_{E''}$, $M_F$, and $M_{E''F}$). These sORFs are particularly interesting because sORF E″ is absolutely conserved in all CaMV strains, the start codon of E″ is in a perfect sequence context for translation initiation (Table 1), and sORF F overlaps the first longer ORF downstream of the leader (Fig. 1). Both features might therefore influence downstream translation. However, all mutant viruses infected plants with the same characteristics as the wild-type virus (with the exception of $M_{E''}$, with which symptom appearance was slightly delayed on passage at 20°C [Fig. 3]), indicating that these sORFs do not play an important role in virus fitness. In later passages, a first-site mutation in sORF F (a519C) was observed in one clone of $M_{E''F}$; however, this mutation did not restore the ATG codon and thus might instead produce a more favorable primary- or secondary-structure element.

**Mutations in the center of the leader.** The start codons of sORFs D′, D, and E were mutated together in mutant $M_{D'DE}$. Symptom appearance with this mutant virus was delayed by 25%, but already after two passages, reversion to wild-type appearance of infection occurred (Fig. 3). At 20°C, only those substitutions at positions 292 and 247 described above were detected. At 26°C, most clones contained the g344A reversion and two new mutations, g293A (instead of the c292A/T mutation) and T245A. The last mutation is equivalent to the T247G mutation described above in increasing the stability of stem section 3 (Fig. 4). A double mutation in the sORF E ATG region was observed in one clone (Tt399CA). This mutation would also lead to increased stability of stem section 3 if the original leader sequence were present at positions 245 and 246. In this particular case, however, the increased stability was somewhat compromised by the T245A mutation present in the same clone. In general, substitutions that increase the stability of stem section 3 are more frequently observed at 26°C than at 20°C with this mutant (Fig. 3).

The G249A mutation, which was so predominant in early revertants obtained after infection with the $M_{ALL}$ mutant, was not found with the $M_{D'DE}$ mutant.

These data indicate that stem section 3 has considerable importance as a virulence determinant. However, mutation of the sORF E start codon, which destabilizes this stem, either alone or in combination with mutation of sORF D did not cause any delay in symptom appearance ($M_E$ and $M_{DE}$, Fig. 3). Also, in mutants in which all ATGs besides that of D′ have been mutated ($M_{ALL+D'}$) or in which the ATGs of sORFs A, B, C, D, and E have been mutated ($M_{ABCDE}$), the only reversion after the third passage was in the start codon of sORF A (Fig. 3). However, without restoration of stem section 3 stability, these revertants are still delayed in symptom appearance by about 70% and therefore clearly have not yet evolved to an optimal configuration.

We noticed that whenever sORF D′ was present, no reversions of the mutation in the D start codon occurred.

From these data, it appears that certain structures other than sORFs have to be conserved in the center of the leader. For mutations within stem section 3, the importance of structure is obvious because of the restoration of base pairing. For mutations in sORF D′ and D, this is less apparent, since these mutations reside in single-stranded loops. However, computer predictions for the structure of this region suggest that an optimal conformation is indeed altered in the mutant RNAs, and such a mutant conformation is weakened when the observed reversions or second-site substitutions are incorporated

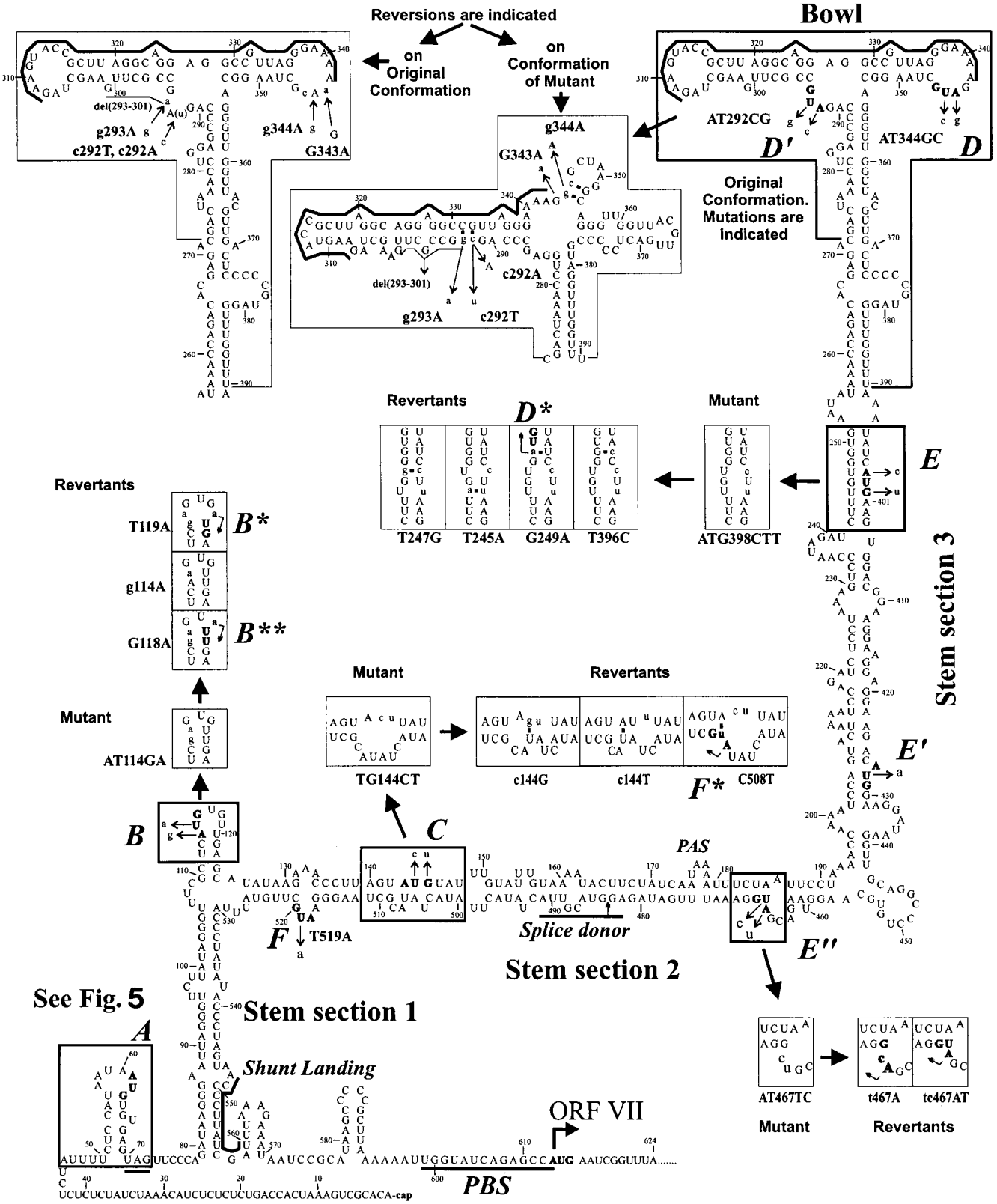


FIG. 4. Reversions of the 35S RNA leader secondary structure and sORFs generated in planta. The MFold-predicted (20°C) hairpinlike secondary structure of the strain CM4-184 leader is shown. Here we refer to three parts of the main stem, which are divided by the side branches, as stem sections 1, 2, and 3 in ascending order and to the top of the structure as the bowl. Elements of this structure, which are influenced by the sORF mutations and the ensuing reversions, are boxed. The start codons of the sORFs are indicated in boldface and also by uppercase italic letters (A to F) along the sequence. Mutations are indicated by thin arrows. Mutant and revertant conformations of the structural elements are drawn in boxes nearby and indicated by thick arrows. Bent thin arrows within boxes show novel start codons created by the reversions in some cases (B*, B**, D*, F*). Nucleotides that do not occur in the wild-type sequence are in lowercase letters. Thick lines define the conserved purine-rich sequence of the bowl, the shunt landing sequence, the sORF A stop codon, the PBS, and the splice donor. PAS, polyadenylation signal.

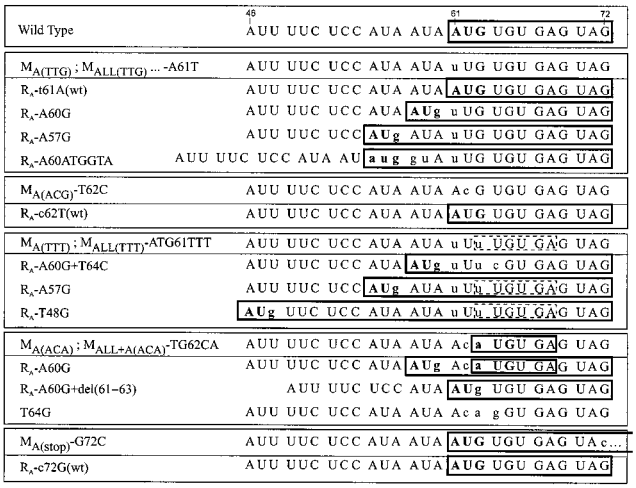


FIG. 5. True and second-site reversions of sORF A generated in planta. Different types of sORF A mutations (designated on the left) and the corresponding reversions are grouped in panels. The coding sequences of the wild-type sORF A (in the uppermost panel) and the reverted sORF As are boxed. Lowercase letters indicate nucleotides that do not occur in the wild-type sequence. The mini-sORF created in place of sORF A by some mutations is also boxed (a dashed box shows the same mini-sORF which might be opened by the TTG codon). In mutant $M_{A(stop)}$ (in the bottom panel), sORF A extends for an additional 23 codons.

(Fig. 4). Our computer predictions favor the structure previously termed a bowl (17). Chemical and enzymatic structural analyses have led to the suggestion of an alternative crosslike structure (30) that shares with the bowl the two stem-loops but has a different organization of the intervening sequence. One revertant, represented by two of nine clones obtained from the $M_{D'DE}$ mutant (Fig. 3), exhibited a deletion of 9 nt that removed a sequence that in the cross structure is largely unpaired but in the bowl structure is paired. However, structure prediction for this revertant suggests a third conformation different from all of the others (data not shown). This deletion is always accompanied by another mutation increasing the stability of stem section 3 (T396C [Fig. 4]).

Structural comparisons revealed that 4 of 11 CaMV isolates (NY8153, CM1841, JI, and CMV1) carry the base pair inversion (C397-G248 to G397-C248) maintaining the structure of stem section 3 but at the same time introducing an in-frame ATG just upstream of the start codon of sORF E (Fig. 1; note that CM1841 lacks the latter ATG because of an additional substitution [A400G]). Since sORF E itself appears to be dispensable for infectivity, the latter finding again supports the importance of the secondary structure in this region.

**Mutations at the beginning of the leader.** A recurrent feature of the mutants $M_{ALL}$, $M_{ABCDE}$, $M_{ALL+D'}$ is the reappearance of sORF A after two passages. For these mutants, we never observed reversions in the sORF C region and only occasionally found a reversion in sORF B (g114A, in only two clones of the $M_{ALL}$ offspring [Fig. 2]). Also, when sORF A was mutated in combination with either sORF B or C, only substitutions restoring sORF A were found (mutants $M_{AB}$ and $M_{AC}$ [Fig. 3]). In contrast, in one particular case of mutant $M_{ABC}$, besides restoration of sORF A, changes in the sORF B and C regions or in the complementary regions of the stem-loop structure were frequently observed (Fig. 2 and 4). The most frequent substitution in sORF B did not change the GAG codon to which the original start codon had been altered but introduced a new start codon by a T119A transversion. The new sORF was not in frame with the original sORF B. This mutation might also restore structural features at the top of the sORF B side stem (Fig. 4). The second-most-frequent change was a G118A transition, which did not alter the predicted structure (Fig. 4). This transition might create a new sORF with a non-ATG start codon (ATT in a good sequence context for translation initiation), but since such non-ATG codons are only weak initiators unless they are supported by specific structural features, an unknown structural effect is as likely as a translational effect. The third substitution restores one of the mutated bases (g114A) and might again improve the structure (Fig. 4). These changes were usually accompanied by a C508T transition, which improves the base pairing in the sORF C region and, in addition, creates a new sORF which overlaps with sORF F (Fig. 4, F*; also occurs naturally in strain D/H [Fig. 1]). In addition, we found subpopulations in the $M_{ABC}$ offspring containing other first-site substitutions in the sORF C region (c144T or c144G, not restoring an ATG) which might improve either the base pairing (Fig. 4) or some primary sequence motif, while the sORF B region remained unchanged (Fig. 2).

**A translatable sORF A is an important determinant of infectivity.** In the mutants $M_{AB}$, $M_{AC}$, $M_{ABC}$, $M_{ABCDE}$, $M_{ALL+D'}$, and $M_{ALL}$, the sORF A start codon has been altered from ATG to TTG. In planta, in all these mutants a sORF A was recreated by point mutations that introduced an ATG at either the original position (t61A) or one of two immediately upstream triplets (A60G and A57G). When this mutation was tested individually in mutant $M_{A(TTG)}$, the same substitutions occurred as well as an additional one creating an ATG by an insertion of 5 nt (Fig. 2 and 5).

Reversion to the wild-type sequence also occurred when the sORF A start codon was changed to ACG instead of TTG [mutant $M_{A(ACG)}$]. After the second passage, the c62T transition was predominant in the entire progeny at both temperatures (Fig. 3). The presence of all of these revertants suggests that the restoration of a translatable sORF is important. However, symptom development with mutants $M_{A(TTG)}$ and $M_{A(ACG)}$ was delayed by only 10%, and no reversions were found in the progeny of the primary infection cycle. Thus, even though the presence of a sORF A apparently increases the competitiveness of a viral genome so much that all of the progeny from a number of different sORF A mutants have acquired it, the mutant virus is almost as infectious as a virus with a wild-type sORF A. We had altered the ATG codon in only one position and created codons that might function as non-ATG start codons (see Discussion). We therefore produced an additional mutant with a TTT triplet instead of the ATG. Symptom development with this mutant [$M_{A(TTT)}$] was delayed by 20 to 45% in the primary-infected plants. This delay was only slightly reduced in the first passage. In the second passage, no delay was observed and all the sequenced progeny clones had regained a sORF A via a second-site substitution upstream of the original site (A57G or T48G [Fig. 3 and 5]). The TTT mutation in the context of a sORF free leader [mutant $M_{ALL(TTT)}$] showed a delay of up to 100%. Second-passage progeny showed (besides the substitutions in the center of the leader described above) a sORF A created by an A60G transition. This reversion was always accompanied by a T64C transition (Fig. 5). It is noteworthy that in the TTT mutants, second-site substitutions restoring an in-frame ATG by a single base change at each of three possible positions upstream of the original start codon were generated, whereas the three other positions at which an ATG would be out of frame were never used (Fig. 5). Such an out-of-frame start codon was artificially generated by construction of the mutant $M_{A(ACA)}$. In this

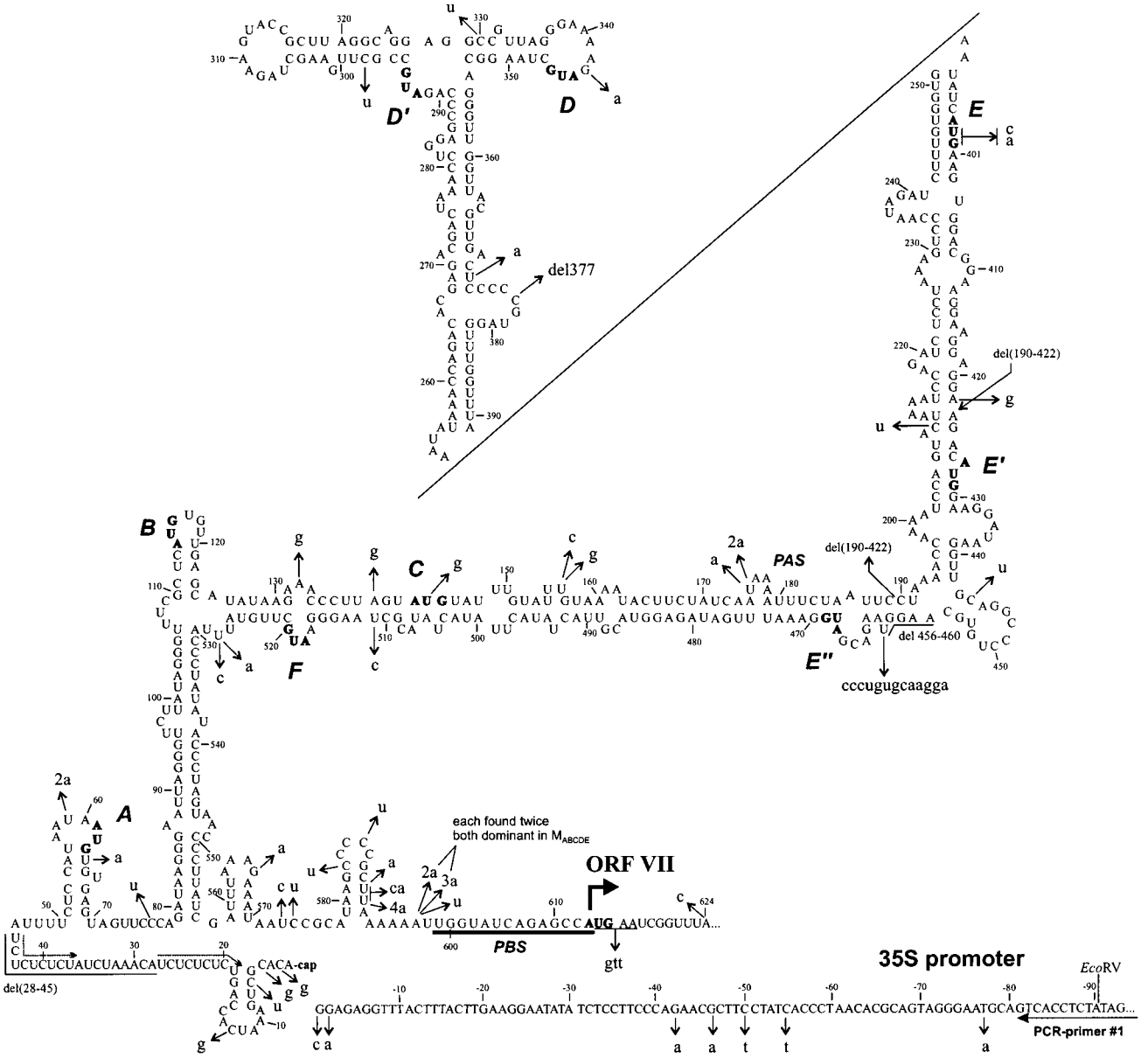


FIG. 6. Microvariants of the 35S RNA leader found in the offspring of different sORF mutants. Nucleotide substitutions, deletions, and insertions are in lowercase letters and are indicated by thin arrows on the hairpinlike structure (disrupted in stem section 3 for the sake of compactness). A bent thick arrow shows the position of the ORF VII start codon; the CaMV PBS is underlined. The contiguous sequence of the CaMV 35S RNA promoter is shown at the bottom; the annealing site of PCR primer 1 is indicated. The start codons of sORFs are indicated in boldface and also by uppercase italic letters (A to F). Dotted arrows show a 10-nt CT repeat. PAS, polyadenylation signal.

mutant, the sORF A sequence ATG.TGT.GAG has been converted to ACA.TGT.GAG (an ATG start codon is underlined), creating the shortest possible sORF, with a start codon followed immediately by a stop codon. Initial attempts to infect plants with this mutant failed. After a fivefold increase in the DNA concentration of the inoculum and after a delay of 200%, one of four inoculated plants developed symptoms. In half of the analyzed clones, a normal sORF A was recreated by an A60G transition; in one clone, the mini sORF was destroyed by a T64G transversion (Fig. 5). The remaining five clones showed no alterations, suggesting that complementation between the reverted and the nonreverted clones can occur. However, these sequences were lost upon further passaging, as was the delay in symptom development (Fig. 3). The progeny consisted only of clones with the A60G transition, either alone or together with a further modification that removed the ACA codon, a deletion of nt 61 to 63 [del(61–63)]. Consequently, the latter revertants had a sORF A with the wild-type sequence (Fig. 5). The same ACA mutation in the context of the ATG-free leader [$M_{ALL+A(ACA)}$] also resulted in a long delay (175%) before symptom development. The progeny invariably contained the sORF A reversion (A60G) together with some of the typical substitutions in stem section 3 and the sORF D′-D region (Fig. 3).

These results may indicate that the nature of the sORF is of some importance. We therefore changed the second codon of

sORF A from TGT to either GCT or TGG. In the GCT mutant [$M_{A(gcT)}$], the codon change also improved the context of the start codon (Table 1). Even though these mutants showed a slight delay in some infection experiments, the majority of the offspring still contained the original mutations and no characteristic revertants accumulated (Fig. 3). The delay was more pronounced with the TGG mutant [$M_{A(TGg)}$]. This was particularly evident when the mutants were tested in the context of the ATG-free leader, wherein the TGG mutant [$M_{ALL+A(TGg)}$] is still delayed for symptom development by about 50% in the second passage. In the offspring of the GCT mutant [$M_{ALL+A(gcT)}$], the typical reversions in stem section 3 and the sORF D′-D region are represented (Fig. 3). With the $M_{ALL+A(TGg)}$ mutant, only the reversions in the sORF D′-D region are found.

In mutants $M_{A(stop)}$ and $M_{ALL+A(stop)}$, the stop codon of sORF A was changed from TAG to TAC, which extended the sORF by an additional 23 codons. When a mutant with this elongated sORF in an otherwise ATG-free leader [$M_{ALL+A(stop)}$] was tested, it proved to be noninfectious. After infection with a fivefold-concentrated inoculum of a mutant with an otherwise normal leader [$M_{A(stop)}$], one plant out of four developed symptoms after a 200% delay. In all of the offspring clones, the original mutant had reverted to the wild-type sequence (Fig. 3 and 5).

**Microvariations in mutant progeny.** In addition to the reversions described above, we found in the offspring of mutant viruses a few occasional substitutions, deletions, and insertions, most of which occurred only once in one clone (Fig. 2, 3, and 6). Such microvariations were found in 21% of the 349 sequenced clones. Most of these were simple base changes probably caused by misincorporation during transcription or reverse transcription. The overall ratio of transitions to transversions was about 1.5. It was almost the same for both occasional and dominant substitutions. Among occasional substitutions the most frequent were the C-to-T transition and the T-to-A transversion. Overall, the pattern of nucleotide sequence variation reported here is strikingly similar to that found among CaMV isolates by Chenault and Melcher (6). The only exception is an increased frequency of A-to-G transitions in our case. However, most of them represent the dominant reversions of sORF A, for which a strong selective pressure apparently exists. Among the occasional substitutions at other sites, the A-to-G transition occurred sixfold less frequently.

These additional variations were rather scattered along the leader, although two hot spots of mutation appeared to exist. The first is immediately upstream of the PBS (nt 599 to 613) and the second is around the transcription start site (Fig. 6), respectively suggesting that the frequency of errors made by the reverse transcriptase is higher near the critical regions of binding and template switching (see reference 3 for a review). In particular, insertions of additional adenosines in the oligo(A) stretch at nt 592 to 597 were frequently observed (Fig. 6).

Most of the occasional substitutions are not disruptive for the hairpinlike secondary structure because they either fall into single-stranded regions or, in many cases, produce G-U base pairs (Fig. 6). Again, this suggests an importance of the secondary structure in the CaMV leader.

The most extreme mutation was found in one clone of the $M_{ALL}$ progeny. A deletion spanning nt 190 to 422 removed half of the leader sequence (Fig. 2 and 6). It most probably occurred via premature template switching of the reverse transcriptase from position 422 to the 3′ end of the 35S RNA 3′ untranslated region, which has been mapped to position 190 or 192 (51). This deletion is probably lethal in the absence of a helper virus. Another deletion which can be explained by premature template switching spans positions 28 to 45 in one clone of the $M_{ABC}$ progeny. It could have occurred via template switching between two consecutive 10-nt CT repeats (Fig. 6).

In two clones of the $M_{ABC}$ progeny, after the third passage the T at position 175 within the polyadenylation signal (PAS), AATAAA, was replaced with either A or AA (Fig. 2 and 6). It has been shown that in plant protoplasts the CaMV PAS can tolerate various mutations, including the substitution of A for T (51).

It has been recently suggested that many of the sequence variations observed in virus populations are artifacts of the analysis, which generally involves many PCR amplification steps (56). By using a PCR polymerase with a proofreading activity and optimized PCR conditions (see Materials and Methods), we could minimize this problem. In fact, in our mutagenesis experiments we never found any accidental misincorporations, deletions, or insertions within about 800 nt of the amplified product, even after as many as 140 amplification cycles in total. We therefore conclude that the observed sequence variations have been generated in planta.

## DISCUSSION

Many viruses use the 5′ ends of their genomic or pregenomic RNAs in processes coordinating gene expression and replication. This requires the presence of regulatory sequences or structures which may influence each other and may also vary in response to viral or cellular factors such that identical viral RNAs can perform different functions in the course of the infection cycle. Paradigms for such complex leaders can be found, e.g., in the animal picornaviruses and in some complex retroviruses. All plant pararetroviruses possess complicated leader sequences containing a number of sORFs and having extensive secondary structure. An elegant way to study sequence requirements in these elaborate and condensed systems is to allow nature to choose the optimal virus genome by forcing evolution from suboptimal genomes. This method is particularly applicable for RNA viruses and retroviruses which employ in their replication enzymes lacking proofreading capability and therefore are supposed to evolve quickly. For the pararetroviruses, these enzymes are the cellular RNA polymerase II and the virus-encoded reverse transcriptase, which is known to produce errors not only by misincorporation but also by template switching within one or between different RNA molecules (for a review, see reference 32).

We have used this approach of studying mutations in the leader sequence of CaMV, and the ensuing reversions, to address the question of whether any of the sORFs are important for viral infectivity. In the first cycle, plants were infected with in vitro-generated, uniform mutant DNA. For subsequent passages, sap from infected plants, which contained a pool of viral genomes packaged into particles, was used. Different coinfecting CaMV genotypes can be propagated together, and since recombinants between such genotypes can be generated, they must exist together in one cell. Cross-protection has been reported for CaMV; 2 days' advantage is sufficient for an otherwise inferior virus to preclude infection by an otherwise dominant variant (60). It is likely that competition between genomes will occur mainly during the very early phases of infection, when only very low copy numbers of the given genomes exist in the nucleus. In newly developing leaves that show a characteristic mosaic pattern, all virus genomes are probably derived from the pool of virus genomes that occupies the infected area near the growing tip (1, 40). If selection occurs before this pool is built up, there will be only a small

number of competitive events. Alternatively, such events may be numerous if this pool contains a large variety of clones. We very rarely observed reversions in the sequence pool analyzed after the primary inoculation, even though the considerable delays in infection rates showed that the mutant viruses were less efficient than the wild-type virus. However, this pool certainly contained a low percentage such revertants, since they invariably dominated in subsequent passages. This suggests that only a few selective infections occur during one plant infection cycle. More-effective genomes would probably be present only in some of the many cells infected in parallel and could exert their advantage only in later passages.

We have altered start codons of sORFs in the leader and, thus, certainly changed translational features. These changes and the direct effect of the sequence alterations also alter the structure of the leader, and this in turn may directly or indirectly influence reverse transcription, e.g., by affecting the packaging of RNA. CaMV genomes with an ATG-free leader proved to be infectious, although symptom appearance was delayed. Most reversions involved stem section 3 and the sequence at the top of the leader hairpin. Computer-aided predictions of structure suggest that the previously described (17) bowl conformation of this region is the most stable one (Fig. 4); however, a number of suboptimal structures with only slightly higher energy are conceivable. One of those is a cross-like structure that has recently been suggested on the basis of in vitro probing (30). The formation of either structure is disturbed by the mutations in the start codons of sORFs D′ and D (Fig. 4) and, according to the predictions obtained with the MFold program, also by mutations in stem section 3 (e.g., sORF E's start codon). All of the ensuing reversions and second-site mutations generated in planta lead to at least partial restoration of the structure (Fig. 4). The D′ region seems to be of special importance, since it invariably reverted very quickly while reversions in the D region were detected only occasionally, later in the passaging, and never when the D′ region was intact. Also, mutations in stem section 3 were tolerated when the D′ region was left intact ($M_{ALL+D'}$ and $M_{ABCDE}$). Strikingly, reversions in response to mutations in D′ and D occurred in almost all cases directly at the mutated nucleotides and not—as for almost all other mutants—at secondary sites, which is consistent with a requirement for not only the secondary structure but also a conserved primary sequence in the region. In fact, the top of the hairpin harbors the most conserved sequence in the caulimoviral leaders (Fig. 4) (17). A similar purine-rich sequence is also found in the badnavirus RTBV (29).

The bowl region has been shown to be dispensable for the unusual translation mechanisms of CaMV (17, 19, 22) and is not known to contain sequences affecting any other step of gene expression. The infectivity of the respective mutants and the ease with which revertants are generated argue for its importance late in the infection cycle, e.g., in the production of stable virions or in (long-distance) transport.

For mutations affecting secondary structure, temperature-dependent effects are conceivable. In our experiments, we observed a slight increase in stem section 3 restoration at 26°C compared to that at 20°C, but the effect was not drastic (Fig. 2 and 3).

One of the frequently occurring mutations (G249A) found in early-passage progeny of the $M_{ALL}$ mutant only slightly improved the predicted stability of stem section 3 by altering a G-U base pair to A-U. This mutation, however, also created a new sORF of nine codons. This particular mutation was dominant only in a pool of viral sequences that contained only a few members with an intact sORF A or with a reversion in the D region. It was outcompeted later by other revertants containing a sORF A, reversions in the D region, and another stem section 3-restoring mutation (T247G [Fig. 2]). It remains to be investigated whether the novel sORF had a function in altering the ribosome flow on this otherwise ATG-free leader or whether the effect of the mutation was only structural.

In other cases, very few sORFs were created by reversions. Since most start codons had been mutated at two positions, direct reversions would have been difficult, but usually nearby codons could have been used to create an ATG codon by single-site mutations. It therefore appears that none of the sORFs is absolutely required for virus infectivity. However, all mutants tested acquired a sORF A after at most two additional passages. ATG-creating mutations occurred at different sites at or upstream of the original sORF A ATG (Fig. 5), suggesting that the presence of a sORF per se, rather than a structural feature, provides for the selective advantage while the length and the N-terminal coding potential of the sORF is not important. Sequence changes within the wild-type sORF A were also tolerated [$M_{A(TGg)}$, $M_{ALL+A(TGg)}$, $M_{A(gcT)}$, and $M_{ALL+A(gcT)}$]. The TGG mutant, however, showed a slightly increased delay period length and therefore may be impaired in infectivity. In the different mutants that lack sORF A, ATG reversions occurred at every one of the three nearby codons at which an ATG in frame with the original one could be generated by a single-site mutation. Significantly, none of the three nearby out-of-frame possibilities was used (Fig. 5). The introduction of an out-of-frame ATG immediately followed by a stop codon [$M_{A(ACA)}$ and $M_{ALL+A(ACA)}$] resulted in a dramatic decrease in fitness. Mutant $M_{A(ACA)}$ evolved either by a knockout of the out-of-frame ATG by point mutation to AGG or, in a dominant subpopulation, by creation of an in-frame ATG at an upstream position. Furthermore, the artificial ACA codon was deleted from some genomes of the latter progeny. This is an indication that either the additional start codon was detrimental even when located within the sORF A or the mutated sORF A coding sequence was suboptimal. It is noteworthy that about half of the progeny of the primary infection with $M_{A(ACA)}$ retained the mutant sequence, suggesting that these genomes could be propagated together with the revertant ones although they were severely compromised when present alone. This, together with the severely reduced infectivity and the very early occurrence (and selection) of revertants, led us to conclude that the mutation affected an aspect of gene expression and that these mutants contain all of the *cis*-acting sequences necessary for replication and transport.

The infection characteristics of a mutant in which the stop codon of sORF A was altered [$M_{A(stop)}$] were similar to those of $M_{A(ACA)}$. This mutation causes an elongation of sORF A from 4 to 27 codons and resulted in only occasional infection, and only after a 200% delay period. In all sequenced clones, a true reversion with recreation of the stop codon had occurred. Interestingly, a reversion was detected only at the original position and not at various other nearby codons at which a stop codon could have been generated by a single-site mutation. This is in contrast to the different second-site mutations restoring a sORF A start codon and may indicate that termination of sORF A translation at the original site is critical, either positionally or mechanistically. All functional sORFs A terminate in GAGTAG (Fig. 5). It remains to be investigated whether the type of termination signal, the last codon, or the position of the signal with respect to other leader regions (e.g., stem section 1; see below) is important. In other systems, unknown features of sORF coding potential or termination have been shown to influence the control of translation by such

sORFs. In particular, reinitiation capacity could be influenced (see reference 31 for a review). We have shown previously that the sORF A region is important for CaMV ORF VII translation by the ribosome shunt mechanism in artificial reporter gene constructs (22) and also that sORFs can be important for the reinitiation mechanism activated by the CaMV transactivator protein on polycistronic mRNAs (20, 21). The effects of the different mutations and reversions on these translation processes are currently being investigated.

The apparent advantage of sORF A revertants raises the questions of why most of the sORF A mutants are infectious with only a slight delay and why revertants are selected only in later passages. We assume that the important feature in sORF A may still function partially even in most of the mutants. The strongly structured stem section 1, 15 nt downstream of the sORF A start codon (Fig. 4), is suitably located to enhance translation initiation at codons in an unfavorable sequence context or at non-ATG codons (39). Therefore, a certain level of translation initiation should occur at the ATG codons that have been changed at only one position (TTG or ACG [28]) or at the naturally occurring ATA codon immediately upstream of sORF A (the latter codon might be used when the ATG is replaced with noninitiating triplet ACA or TTT). This low level of initiation could be sufficient for the sORF to exert its function if a mechanism like ribosome stalling is involved. Stalling seems to be the mechanism by which a sORF in the leader of the human cytomegalovirus gpUL4 RNA controls downstream translation (5) and the *Neurospora crassa arg-2* gene sORF attenuates translation in the presence of arginine (58). Such a mechanism might be also responsible for the effects of other sORFs (see, e.g., references 8, 42 and 44). Stalling depends on sequence features but is not significantly affected by the initiation frequency. In the case of CaMV, stalling might be caused by the sequence around the termination codon, which was unchanged in all functional sORFs A. It can be further proposed that stalling of the translating or terminating ribosome near the sORF A stop codon is an initial step of the ribosome shunt mechanism. Strikingly, the predicted secondary structure brings sORF A into the close spatial vicinity of the shunt landing sequence (22) that is involved in the formation of the base of stem section 1 (Fig. 4). Stalling of the ribosome before dissociation, which might be further prompted by the close proximity of stem section 1, might allow a physical interaction of the 40S ribosomal subunit and the landing sequence. In this model, precise positioning of the sORF A stop codon 6 nt upstream of the base of the stem should be critical, which is consistent with our in planta findings. Structural comparison reveals strikingly similar positioning of a 5′-proximal sORF near the base of large hairpin structures in the pregenomic RNA leaders of different plant pararetroviruses (48a). Significantly, in RTBV RNA (which has been shown to utilize a ribosome shunt mechanism [3]), a precisely defined shunt landing sequence is partially involved in the formation of the hairpin base (on the descending strand) and a 5′-proximal sORF 1 terminates in GAGTAG 7 nt upstream of the base of the hairpin. This similarity to CaMV strongly supports the proposed model of ribosome shunting.

Besides sORF A, the only other sORF absolutely conserved in all CaMV strains is sORF E″. Its removal had no apparent detrimental effect, but interestingly, true revertants were obtained in one experiment ($M_{ALL}$ after four passages) even though two mutations are required for the sORF E″ restoration. This sORF is present in a region to which a cellular protein of unknown function binds (11), and its importance remains to be investigated.

Mutations in the leader sequence, which could not be connected to reversion events due to the incorporated mutations, were observed at a surprisingly low frequency even after four passages. This suggests either that the selection for functional genomes is tight and few alterations are tolerated or that the polymerase is more precise than has been assumed. The notion that reverse transcriptases are error prone is mainly derived from studies of retroviral enzymes. However, even for retroviruses, the error frequency of the reverse transcriptase is different for different viruses and seems to depend at least in part on the nature of amino acid X in the conserved YXDD center (47). In mammalian viruses, a certain error frequency may be an evolutionary advantage because it allows the generation of new variants which may escape immune surveillance. In plants, virus defense mechanisms are less well studied. They may also include the recognition of a specific protein (46), but this seems not to be a general mechanism, and therefore rapid alteration of sequence information may not be as advantageous.

## ACKNOWLEDGMENTS

We are grateful to Matthias Müller, Sandra Corsten, and Diana Lewetag for excellent technical assistance. Special thanks go to Herbert Angliker for expert sequencing work. We thank Witold Filipowicz and Helen Rothnie for critical reading of the manuscript and Diana Dominguez, Lyubov Ryabova, and Etienne Herzog for helpful discussions. M.M.P. thanks his wife, Natalia Kolman, for understanding and moral support and Konstantin Skryabin for his steady help and interest.

This work was partially supported by an EMBO East-European postdoctoral fellowship to M.M.P.

## REFERENCES

1. **Al-Kaff, N. S., and S. N. Covey.** 1996. Unusual accumulations of cauliflower mosaic virus in local lesions, dark green leaf tissue, and roots of infected plants. Mol. Plant-Microbe Interact. **9:**357–363.
2. **Baughman, G., and S. H. Howell.** 1988. Cauliflower mosaic virus 35S RNA leader region inhibits translation of downstream genes. Virology **167:**125–135.
3. **Bonneville, J.-M., and T. Hohn.** 1993. A reverse transcriptase for cauliflower mosaic virus: state of the art, 1992, p. 357–390. *In* N. Skalka and S. Goff (ed.), Reverse transcriptase. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
4. **Bonneville, J.-M., H. Sanfaçon, J. Fütterer, and T. Hohn.** 1989. Posttranscriptional *trans*-activation in cauliflower mosaic virus. Cell **59:**1135–1143.
5. **Cao, J., and A. P. Geballe.** 1996. Inhibition of nascent-peptide release at translation termination. Mol. Cell. Biol. **16:**7109–7114.
6. **Chenault, K. D., and U. Melcher.** 1994. Patterns of nucleotide sequence variation among cauliflower mosaic virus isolates. Biochimie **76:**3–8.
7. **Curran, J., and D. Kolakofsky.** 1988. Scanning independent ribosomal initiation of the Sendai virus X protein. EMBO J. **7:**2869–2874.
8. **Damiani, R. D., Jr., and S. R. Wessler.** 1993. An upstream open reading frame represses expression of *Lc*, a member of the *R/B* family of maize transcriptional activators. Proc. Natl. Acad. Sci. USA **90:**8244–8248.
9. **Dixon, L. K., and T. Hohn.** 1984. Initiation of translation of the cauliflower mosaic virus genome from a polycistronic mRNA: evidence from deletion mutagenesis. EMBO J. **3:**2731–2736.
10. **Dixon, L. K., J. Jiricny, and T. Hohn.** 1986. Oligonucleotide directed mutagenesis of the cauliflower mosaic virus genome DNA using a repair-resistant nucleoside analogue: identification of an agnogene initiation codon. Gene **41:**225–231.
11. **Dominguez, D. I., T. Hohn, and W. Schmidt-Puchta.** 1996. Cellular proteins bind to multiple sites of the leader region of cauliflower mosaic virus 35S RNA. Virology **226:**374–383.
12. **Donzé, O., P. Damay, and P.-F. Spahr.** 1995. The first and third uORFs in RSV leader RNA are efficiently translated: implications for translational regulation and viral RNA packaging. Nucleic Acids Res. **23:**861–868.
13. **Donzé, O., and P.-F. Spahr.** 1992. Role of the open reading frames of Rous sarcoma virus leader RNA in translation and genome packaging. EMBO J. **11:**3747–3757.
14. **Dowson-Day, M. J., J. L. Ashurst, S. F. Mathias, and J. W. Watts.** 1993. Plant viral leaders influence expression of a reporter gene in tobacco. Plant Mol. Biol. **23:**97–109.
15. **Franck, A., H. Guilley, G. Jonard, K. Richards, and L. Hirth.** 1980. Nucleotide sequence of cauliflower mosaic virus DNA. Cell **21:**285–294.
16. **Fütterer, J., J.-M. Bonneville, K. Gordon, M. DeTapia, S. Karlsson, and T.**

**Hohn.** 1990. Expression from polycistronic cauliflower mosaic virus pregenomic RNA, p. 347–357. *In* J. E. G. McCarthy and M. F. Tuite (ed.), Post-transcriptional control of gene expression. NATO ASI series, vol. H49. Springer-Verlag, Berlin, Germany.
17. **Fütterer, J., K. Gordon, J.-M. Bonneville, H. Sanfaçon, B. Pisan, J. Penswick, and T. Hohn.** 1988. The leading sequence of caulimovirus large RNA can be folded into a large stem-loop structure. Nucleic Acids Res. **16:**8377–8390.
18. **Fütterer, J., K. Gordon, P. Pfeifer, H. Sanfaçon, B. Pisan, J.-M. Bonneville, and T. Hohn.** 1989. Differential inhibition of downstream gene expression by the cauliflower mosaic virus 35S RNA leader. Virus Genes **3:**45–55.
19. **Fütterer, J., K. Gordon, H. Sanfaçon, J.-M. Bonneville, and T. Hohn.** 1990. Positive and negative control of translation by the leader sequence of cauliflower mosaic virus pregenomic 35S RNA. EMBO J. **9:**1697–1707.
20. **Fütterer, J., and T. Hohn.** 1991. Translation of a polycistronic mRNA in the presence of the cauliflower mosaic virus transactivator protein. EMBO J. **10:**3887–3896.
21. **Fütterer, J., and T. Hohn.** 1992. Role of an upstream open reading frame in the translation of polycistronic mRNAs in plant cells. Nucleic Acids Res. **20:**3851–3857.
22. **Fütterer, J., Z. Kiss-László, and T. Hohn.** 1993. Non-linear ribosome migration on cauliflower mosaic virus 35S RNA. Cell **73:**789–802.
23. **Fütterer, J., I. Potrykus, Y. Bao, L. Li, T. M. Burns, R. Hull, and T. Hohn.** 1996. Position-dependent ATT initiation during plant pararetrovirus rice tungro bacilliform virus translation. J. Virol. **70:**2999–3010.
24. **Galvez, A. F., and B. O. DeLumen.** 1995. Generation of an Nco I restriction site for translation fusions using PCR-mediated, site-directed mutagenesis. Plant Mol. Biol. Rep. **13:**232–242.
25. **Gardner, R. C., A. J. Howarth, P. Hahn, M. Brown-Luedi, R. J. Shepherd, and J. Messing.** 1981. The complete nucleotide sequence of an infectious clone of cauliflower mosaic virus by M13mp7 shotgun sequencing. Nucleic Acids Res. **9:**2871–2888.
26. **Geballe, A. P.** 1996. Translational control mediated by upstream AUG codons, p. 173–197. *In* J. W. B. Hershey, M. B. Mathews, and N. Sonenberg (ed.), Translational control. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
27. **Geballe, A. P., and D. R. Morris.** 1994. Initiation codons within 5′-leaders of mRNA as regulators of translation. Trends Biochem. Sci. **19:**159–164.
28. **Gordon, K., J. Fütterer, and T. Hohn.** 1992. Efficient initiation of translation at non-AUG triplets in plant cells. Plant J. **2:**809–813.
29. **Hay, J. M., M. C. Jones, M. L. Blakebrough, I. Dasgupta, J. W. Davies, and R. Hull.** 1991. An analysis of the sequence of an infectious clone of rice tungro bacilliform virus, a plant pararetrovirus. Nucleic Acids Res. **19:**2615–2621.
30. **Hemmings-Mieszczak, M., G. Steger, and T. Hohn.** 1997. Alternative structures of the cauliflower mosaic virus 35S RNA leader: implications for viral expression and replication. J. Mol. Biol. **267:**1075–1088.
31. **Hinnebusch, A. G.** 1996. Translational control of *GCN4*: gene-specific regulation by phosphorylation of eIF2, p. 199–244. *In* J. W. B. Hershey, M. B. Mathews, and N. Sonenberg (ed.), Translational control. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
32. **Hohn, T.** 1994. Recombination of plant pararetroviruses, p. 25–38. *In* J. Paszkowski (ed.), Homologous recombination and gene silencing in plants. Kluwer Academic Press, Dordrecht, The Netherlands.
33. **Hohn, T., and J. Fütterer.** 1997. The proteins and functions of plant pararetroviruses: knowns and unknowns. Crit. Rev. Plant Sci. **16:**133–161.
34. **Howarth, A. J., R. C. Gardner, J. Messing, and R. J. Shepherd.** 1981. Nucleotide sequence of naturally occurring deletion mutants of cauliflower mosaic virus. Virology **112:**678–685.
35. **Hull, R., and S. N. Covey.** 1995. Retroelements: propagation and adaptation. Virus Genes **11:**105–118.
36. **Kiss-László, Z., S. Blanc, and T. Hohn.** 1995. Splicing of cauliflower mosaic virus 35S RNA is essential for viral infectivity. EMBO J. **14:**3552–3562.
37. **Kiss-László, Z., and T. Hohn.** 1996. Pararetro- and retrovirus RNA: splicing and the control of nuclear export. Trends Microbiol. **4:**480–485.
38. **Kozak, M.** 1989. The scanning model for translation: an update. J. Cell Biol. **108:**229–241.
39. **Kozak, M.** 1990. Downstream secondary structure facilitates recognition of initiator codons by eukaryotic ribosomes. Proc. Natl. Acad. Sci. USA **87:**8301–8305.
40. **Leisner, S. M., R. Turgeon, and S. H. Howell.** 1992. Long distance movement of cauliflower mosaic virus in infected turnip plants. Mol. Plant-Microbe Interact. **5:**41–47.
41. **Li, J.** 1996. Molecular analysis of late gene expression in budgerigar fledgling disease virus. Ph.D. thesis. University of Giessen, Giessen, Germany.
42. **Lohmer, S., M. Maddaloni, M. Motto, F. Salamini, and R. D. Thompson.** 1993. Translation of the mRNA of the maize transcriptional activator *Opaque*-2 is inhibited by upstream open reading frames present in the leader sequence. Plant Cell **5:**65–73.
43. **Mesnard, J. M., and C. Carriere.** 1995. Comparison of packaging strategy of retroviruses and pararetroviruses. Virology **213:**1–6.
44. **Michelet, B., M. Lukaszewicz, V. Dupriez, and M. Boutry.** 1994. A plant plasma membrane proton-ATPase gene is regulated by development and environment and shows signs of a translational regulation. Plant Cell **6:**1375–1389.
45. **Padegimas, L., O. A. Shulga, and K. G. Skryabin.** 1993. Screening of transgenic plants with polymerase chain reaction. Engl. Transl. Mol. Biol. (Moscow) **27:**583–585.
46. **Padgett, H. S., and R. N. Beachy.** 1993. Analysis of a tobacco mosaic strain capable of overcoming N gene-mediated resistance. Plant Cell **5:**577–586.
47. **Pandey, V. N., N. Kaushik, N. Rege, S. G. Sarafianos, P. N. S. Yadav, and M. J. Modak.** 1996. Role of methionine 184 of human immunodeficiency virus type 1 reverse transcriptase in the polymerase function and fidelity of DNA synthesis. Biochemistry **35:**2168–2179.
48. **Pierce, D. A., I. J. Mettler, A. R. Lachmansingh, L. M. Pomeroy, E. A. Weck, and D. Mascarenhas.** 1987. Effect of 35S leader modifications on promoter activity, p. 301–310. *In* J. L. Key and L. McIntosh (ed.), Plant gene systems and their biology. Alan R. Liss, Inc., New York, N.Y.
48a. **Pooggin, M. M.** Unpublished data.
49. **Riederer, M. A., N. H. Grimsley, B. Hohn, and J. Jiricny.** 1992. The mode of cauliflower mosaic virus propagation in the plant allows rapid amplification of viable mutant strains. J. Gen. Virol. **73:**1449–1456.
50. **Rothnie, H. M., Y. Chapdelaine, and T. Hohn.** 1994. Pararetroviruses and retroviruses: a comparative review of viral structure and gene expression strategies. Adv. Virus Res. **44:**1–67.
51. **Rothnie, H. M., J. Reid, and T. Hohn.** 1994. The contribution of AAUAAA and the upstream element UUUGUA to the efficiency of mRNA 3′-end formation in plants. EMBO J. **13:**2200–2210.
52. **Sanfaçon, H., P. Brodmann, and T. Hohn.** 1991. A dissection of the cauliflower mosaic virus RNA polyadenylation signal. Genes Dev. **5:**141–149.
53. **Sanfaçon, H., and T. Hohn.** 1990. Proximity to the promoter inhibits recognition of cauliflower mosaic virus polyadenylation signal. Nature **346:**81–84.
54. **Scholthof, H. B., S. Gowda, F. C. Wu, and R. J. Shepherd.** 1992. The full-length transcript of a caulimovirus is a polycistronic mRNA whose genes are *trans* activated by the product of gene VI. J. Virol. **66:**3131–3139.
55. **Sieg, K., and B. Gronenborn.** 1982. Introduction and propagation of foreign DNA in plants using cauliflower mosaic virus as a vector, p. 154. *In* Abstracts of the NATO Advanced Studies Institute/FEBS advanced course on structure and function of plant genomes, Porto Portese, Italy.
56. **Smith, D. B., J. McAllister, C. Casino, and P. Simmonds.** 1997. Virus 'quasispecies': making a mountain out of a molehill? J. Gen. Virol. **78:**1511–1519.
57. **Sonstegard, T. S., and P. B. Hackett.** 1996. Autogenous regulation of RNA translation and packaging by Rous sarcoma virus $Pr76^{gag}$. J. Virol. **70:**6642–6652.
58. **Wang, Z., and M. S. Sachs.** 1997. Ribosome stalling is responsible for arginine-specific translation attenuation in *Neurospora crassa*. Mol. Cell. Biol. **17:**4904–4913.
59. **Yueh, A., and R. J. Schneider.** 1996. Selective translation initiation by ribosome jumping in adenovirus-infected and heat-shocked cells. Genes Dev. **10:**1557–1567.
60. **Zhang, X. S., and U. Melcher.** 1989. Competition between isolates and variants of cauliflower mosaic virus in infected turnip plants. J. Gen. Virol. **70:**3427–3437.
61. **Zuker, M., and P. Stiegler.** 1981. Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. Nucleic Acids Res. **9:**133–148.