
The leading sequence of caulimovirus large RNA can be folded into a large stem-loop structure

J.Fütterer, K.Gordon, J.M.Bonneville, H.Sanfaçon, B.Pisan, J.Penswick and Th.Hohn*

Friedrich Miescher-Institut, PO Box 2543, CH-4002 Basel, Switzerland

Received June 21, 1988; Revised and Accepted August 2, 1988

ABSTRACT

The 600 nt long sequences preceding the first large ORFs (ORF VII) of three caulimoviruses, although varying in primary sequence, can be folded into a large stem/loop structure centered around a conserved stretch of 36 nucleotides. Deletions of the conserved sequence delay symptom appearance considerably, but do not affect expression of a reporter gene in plant protoplasts. Another striking similarity between the leaders concerns the number and distribution of small open reading frames (sORF) they carry. Expression of two of these sORFs was tested by fusion of a reporter gene: both were expressed in plant protoplasts.

INTRODUCTION

It is unusual for eukaryotic mRNAs to have long leader sequences, and especially for these to contain small open reading frames (sORFs). Such leader sequences have been implicated in the control of gene expression at various levels (1-11). The leader sequences of the caulimoviruses are especially long and have hitherto hardly been studied.

Caulimoviruses can be considered to be the retroviruses of the plant kingdom, since their genomes are replicated by reverse transcriptase and one of their genomic forms is terminally redundant RNA like retrovirus genomic RNA. In contrast to mammalian retroviruses, however, CaMV genomes form circular minichromosomes in the nuclei of plant cells and are stored as open circular DNA molecules in virions. Six strains of the type member cauliflower mosaic virus (CaMV; strain S (12); strain CM1841 (13); strain D/H (14); strain CM4-184 (15); strain XJ (16); strain JI, J.Stanley, pers. comm.) and one strain each of carnation etched ring virus (CERV; 17) and figwort mosaic virus (FMV; 18) have been sequenced. The cap site and 3'-terminus of the genomic (35S) RNA of CaMV have also been determined (19). The cap site is located 31 nt downstream of the TATA box of the CaMV 35S-promoter. In the case of the other two caulimoviruses, promoter consensus sequences have been found in a region corresponding to the one of CaMV and

the cap-sites can be assumed to lie an equivalent distance downstream. The regions between the cap-sites and the larger coding sequences (i.e. the first ORF larger than 200 nt) are 600 nt long and defined here as caulimovirus leaders.

The corresponding sequences (R + U5 + L) of retroviruses are in general shorter, although in some cases they also reach considerable lengths (mouse mammary sarcoma virus [MMSV], 225 nt; mouse mammary tumor virus [MMTV], 300 nt; Rous sarcoma virus [RSV], 380 nt; spleen necrosis virus [SNV] 470 nt) (20). Leader regions in retroviruses have been suggested to contain signals for splicing, RNA dimer linkage and packaging (20); the structure of the RSV leader has been implicated in translational control (6). The long caulimovirus leaders may further reveal some peculiarities of plant virus/host interaction or of plant gene control.

MATERIALS AND METHODS

1) Software.

A VAX 8600 computer and the "Sequence Analysis Software Package" of the Genetics Computer Group, University of Wisconsin (21) were used. Minimal free energy folds of RNA were performed using their adaptation of the "FOLD" program of Zuker and Stiegler (22) with the free energy values as defined by (23). Graphical presentations were obtained with "SQUIGGLES" and "CIRCLES". The two types of graphical presentations are each optimal for certain purposes. SQUIGGLES presents a simple stem-loop structure more directly, but the features of more complex structures are more easily discerned on CIRCLES. In addition to these presentations the sequence and proposed structure of the CaMV goblet was drawn by hand (Fig.4). Sequence comparisons were performed with the adaptation of "COMPARE" and "DOTPLOT" (24).

2) Numbering

Two types of numberings are used in this manuscript. Those referring to published DNA sequence data call the first nucleotide of the primer binding site "1", whereas those referring to RNA structure (e.g. in Figs. 1, 3 and 4) begin numbering at the first nucleotide of the 35S RNA. Both types of numberings are in the direction of transcription.

3) CaMV strains used in experiments and their growth.

Ca379, a deletion in the CaMV leader region is as described by Penswick et al. (25). In this mutant, basepairs 7234 - 7381 (sequence and numbering: ref. 15) are replaced by TCCCCGCTCGAGC. This strain and the parent virus CM4-184 (26) were grown as clones in the E. coli plasmid pUC8 using the

isogenic and unique SalI site. The CaMV sequences were excised from the bacterial vector at the cloning site and inoculated as described (27).

4) Construction of plasmids for electroporation

Plasmid pLC20 is a derivative of plasmid pDH51 (28). The CaMV 35S promoter sequences in pDH51 are extended to the ATG of CaMV ORF VII, which is fused directly to the ORF for a bacterial chloramphenicol acetyl transferase. To allow easy preparation of single stranded DNA for mutagenesis procedures, the pUC plasmid background of pDH51 was changed to pTZ19u (29) in pLC20. In plasmids pLC15 and pLC31 the CAT ORF is fused to the second codon of the CaMV ORF D and E, respectively. All fusions were obtained by oligonucleotide directed deletion mutagenesis in the final step. Exactness of the fusions was verified by dideoxy sequencing. Plasmid pLC421 is a derivative of pLC20 in which the wild type CaMV leader sequence was replaced with the leader sequence of Ca379.

These plasmids are members of a larger series of constructs, which will be described later. Exact sequences and a detailed pathway of construction is available upon request.

5) Transient expression experiments.

Cell suspension cultures of Orychopragmus violaceus, a crucifer and host plant of CaMV were obtained from C.Matsui and maintained and protoplasts prepared according to him (30). Electroporation of 2.10^6 protoplasts in 0.7 ml electroporation buffer was carried out according to (31) by discharging an 820 mF capacitor at 200V. CAT assays were performed according to (32).

RESULTS AND DISCUSSION

Secondary structure of the Caulimovirus leaders

Fig. 1 shows the fold of the first 1200 nts of CaMV RNA in the "CIRCLES" graphical presentation. "CIRCLES" arranges the sequence counterclockwise starting at "3h" and connects pairing bases by arcs. A large series of such arcs can be discerned connecting the stretch of nucleotides 52 to 280 with 346 to 572. This series corresponds to a large stem interrupted by a number of loops and short hairpins (SQUIGGLES printout; Fig.2). The arcs flank a "pair of eyes", corresponding to two short branches and loops. This pair we name here "bowl" and the total structure "goblet". The two side stems (Sb_1 and Sb_2) branching from the main stem (Fig.2) are seen in Fig.1 as small arcs around 2h and 11h.

The leaders of the other two caulimoviruses with published sequence (CERV and FMV) can also be folded into "goblet" like structures carrying "bowls" on top (Figs. 2 and 3). In all three cases the first 50 to 80 nt of

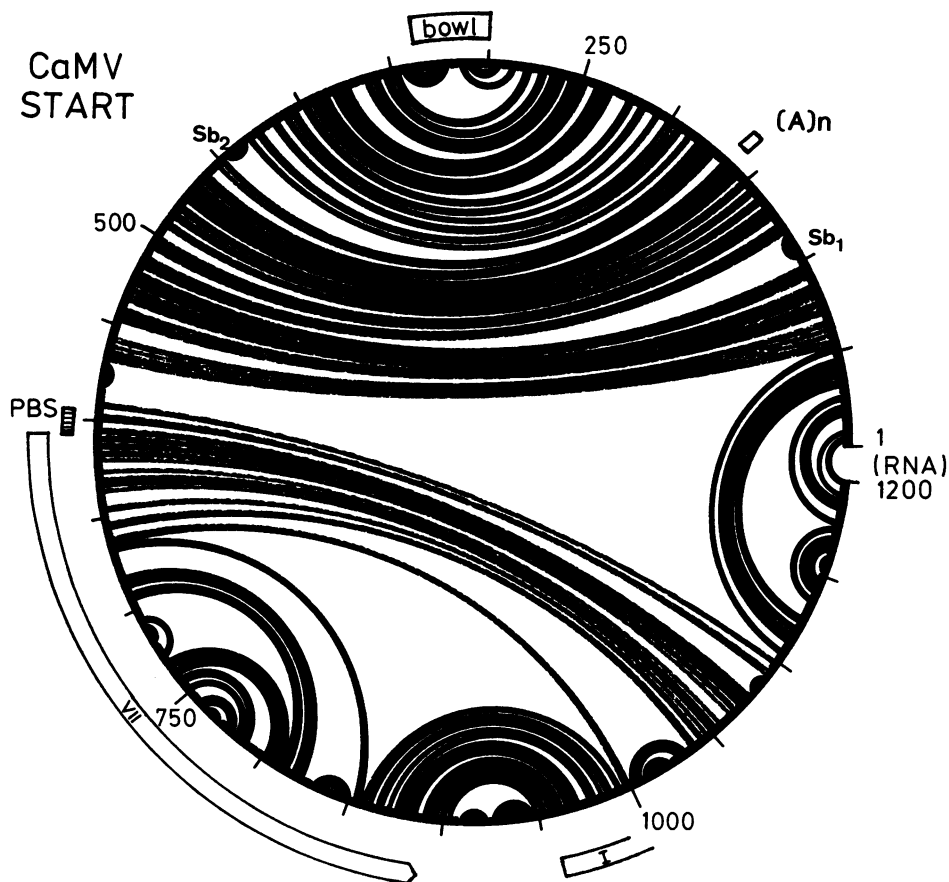


Fig.1. Secondary structure (CIRCLE presentation) proposed by the FOLD program for the 1200 nt of CaMV (strain S) genomic RNA starting with the CAP site. The polyadenylation site [(A)n], "bowl" (-sequence), primer binding site (PBS), ORF VII and the beginning of ORF I are indicated.

the leader have little secondary structure. The stem has a varying number of short side branches, zero for CERV, one for FMV and, as mentioned, two for CaMV. Free minimal energies for the goblets are -116 (FMV), -120 (CERV) and -144 (CaMV) kCal/mol. The sequences of the stems were randomly shuffled (21) 5 times, reconnected to the bowls and folded again; free minimal energies of the shuffled sequences were in the mean 20% higher than for the original structures (e.g. -111, -110, -107, -118 and -112 for the first 5 shuffles for CaMV). In all cases the ascending arm of the stem contains the polyadenylation signal (Fig.1 and 3) and the descending arm terminates ahead

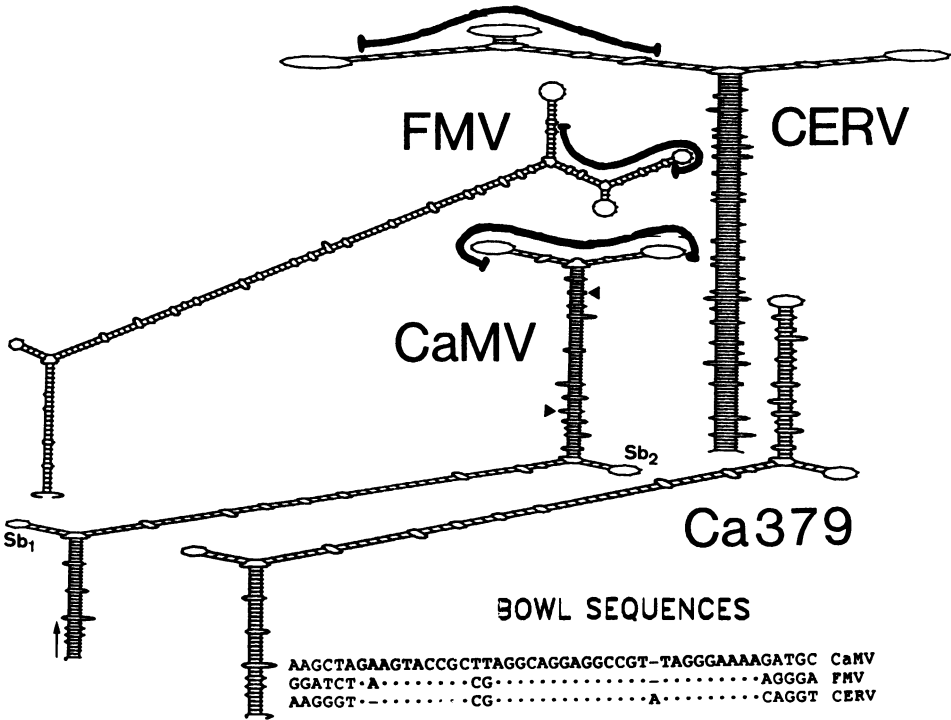


Fig.2 Comparison of secondary structures proposed by the FOLD program for CaMV (strain S), FMV, and CERV (SQUIGGLES presentation). Only the "goblets" are shown; for details of the surrounding sequences please consult Fig.3. The start and end points of the CaMV "bowl" deletion positions are indicated on the CaMV stem/loop by arrowheads; the resulting structure of this deletion (Ca379) is also shown. The "bowl" sequences are printed at lower right and their positions on the individual goblets are marked.

of the primer-binding site and before the first major ORF begins. In CERV a second (smaller) stem/loop structure is located between the large stem and the first ORF (Fig. 3a). The FOLD program used shows only one possible state, namely the one of lowest free energy, while other states of slightly higher free energy remain undetected. Those states might become more important in presence of relevant proteins. Moreover, features of tertiary structure, such as pseudoknots (33, 34) are not considered and the structures determined by the computer have not been confirmed experimentally (e.g. according to 35). Nevertheless the similarity of the structures obtained for the three caulimoviruses is remarkable and therefore they probably reflect an important biological function (which remains, of course,

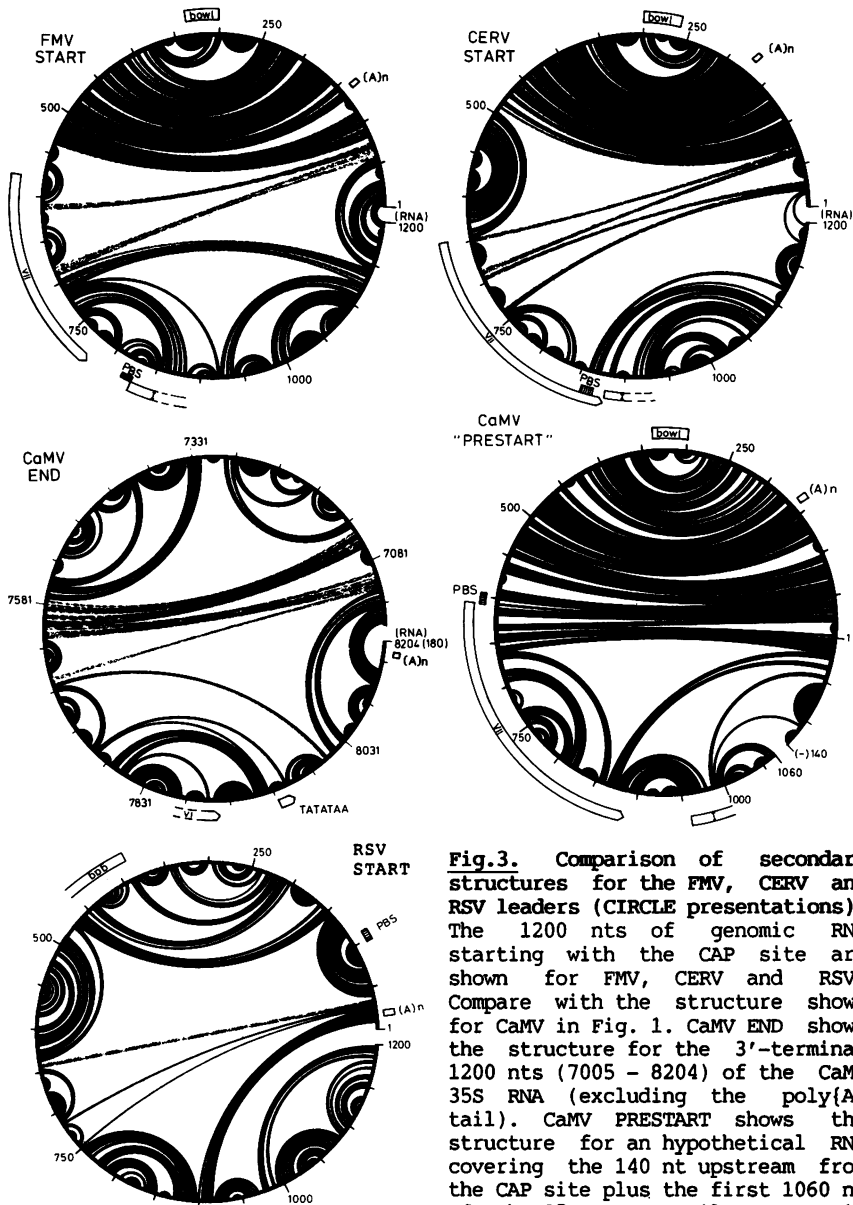


Fig. 3. Comparison of secondary structures for the FMV, CERV and RSV leaders (CIRCLE presentations). The 1200 nts of genomic RNA starting with the CAP site are shown for FMV, CERV and RSV. Compare with the structure shown for CaMV in Fig. 1. CaMV END shows the structure for the 3'-terminal 1200 nts (7005 - 8204) of the CaMV 35S RNA (excluding the poly{A} tail). CaMV PRESTART shows the structure for an hypothetical RNA covering the 140 nt upstream from the CAP site plus the first 1060 nt of the 35S RNA. Details are as in Fig. 1.

to be established). Analysis of the remainder of the caulimovirus genomes in portions of 1200 nt (the limit of the program) and starting at every 400th nt showed no structures of similar extent (see Fig. 3 "CaMV END" for a

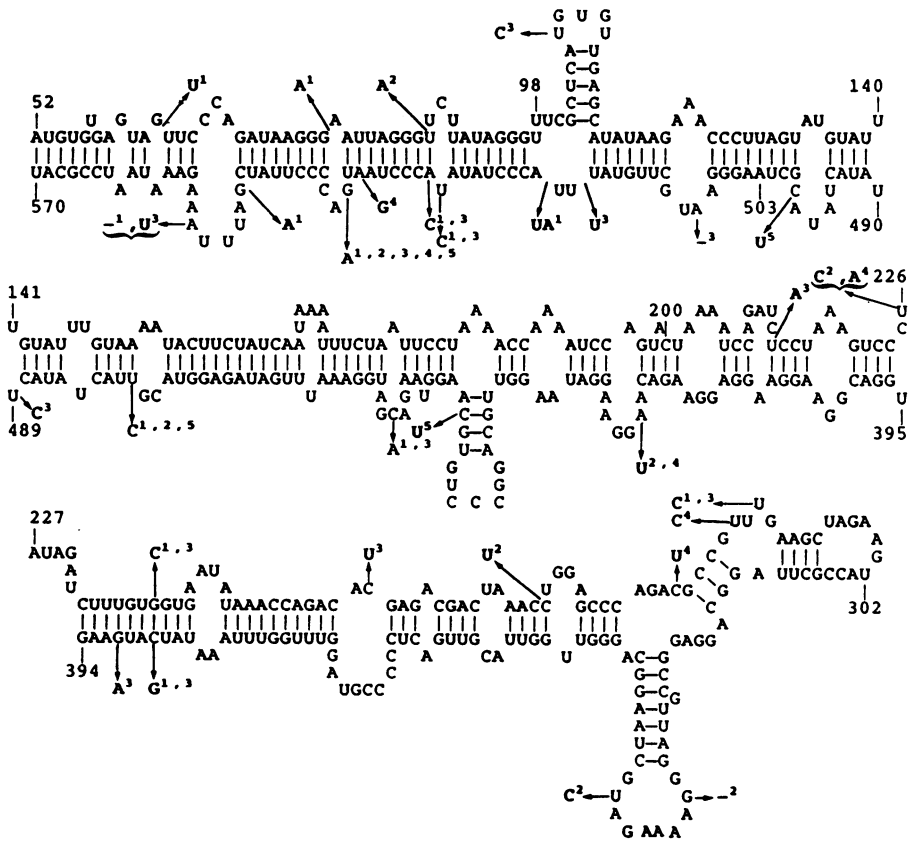


Fig.4. Sequence of CaMV strain S in the proposed folded structure. The sequence is numbered from the first nucleotide of the leader. The base exchanges observed in 5 other strains of CaMV are shown: ¹J1; ²XJ; ³CM-1841; ⁴CM4-184; ⁵D/H.

typical example). However, the length of the CaMV stem was increased when sequences starting artificially at position -140 were analyzed (Fig. 3 "CaMV PRESTART"). This phenomenon was not observed with the other caulimoviruses. We found only much shorter stems in leaders of the retroviruses RSV (Fig. 3; see also 6), MMSV, MMTV and SNV (not shown).

The sequences of the six CaMV strains (references above) were compared for the potential stem/loop structure in the leader (Fig.4). Twenty-one sequence variant sites (nucleotide exchanges, insertions, deletions) were found in the 189 nt single strand regions, of which one variation could lead to an increase in base-pairing. Ten were located in the base-paired stems

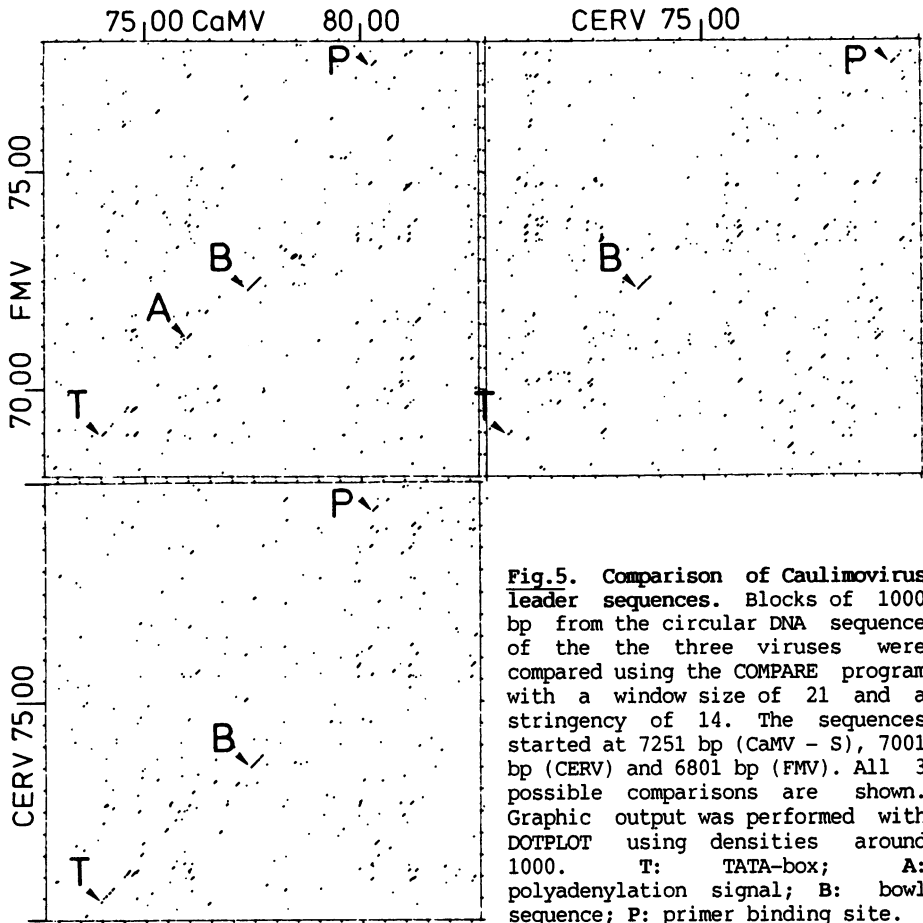


Fig.5. Comparison of Caulimovirus leader sequences. Blocks of 1000 bp from the circular DNA sequence of the three viruses were compared using the COMPARE program with a window size of 21 and a stringency of 14. The sequences started at 7251 bp (CaMV - S), 7001 bp (CERV) and 6801 bp (FMV). All 3 possible comparisons are shown. Graphic output was performed with DOTPLOT using densities around 1000. T: TATA-box; A: polyadenylation signal; B: bowl sequence; P: primer binding site.

(330 nt). Of these, changes at six sites preserved base-pairing (including the complementary changes at nt 239 and 388 in two strains), whereas only 4 would abolish base-pairing (at nt 86, 214, 480 and 530) (Fig.4). Thus the overall bias is towards changes which would not affect the proposed structure, providing indirect evidence for its significance.

A short conserved sequence in the Caulimovirus leaders

The three caulimoviruses have considerable sequence homology in the region encoding ORFs I - VI (17, 18), but little outside this region. Dot matrices (19; Fig.5) of the non coding regions including the leaders show that the only conserved sequences are the TATA box, a CT rich motif between the cap site and the first SORF, the polyadenylation signal, the primer

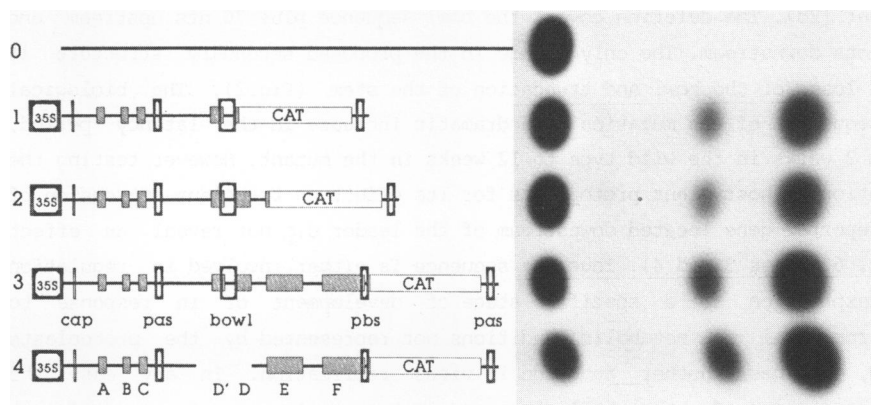


Fig.6. Transient expression experiments. All constructs contain the CaMV 35S enhancer/promoter region and the polyadenylation signals and sites as in pDW2 (28). Between these regions was placed the CaMV 35S leader either in its complete form (plasmid pLC20, line 3) or without the bowl-sequence (pLC421, line 4). The bacterial CAT coding sequence (without its own ATG) is either fused to the first codon of ORF VII (lines 3,4) or to the second codon of sORFs D (pLC15, line 1) or E (pLC31, line 2). The inserts are shown schematically, with sORFs in the leader being indicated as hatched boxes and signal sequences as cartridges: 35S: 35S-promoter; cap: RNA start; pas: facultative polyadenylation signal; bowl: the conserved 36 nt sequence; pbs: the primer binding site for reverse transcription. Transient expression was studied after electroporation of *Orychophragmus violaceus* protoplasts and the CAT assay was performed as described in MATERIALS & METHODS. Thin layer chromatography (right) separates unreacted radioactively labelled chloramphenicol (left spots) from 1-acetyl chloramphenicol (middle spots) and 3-acetyl chloramphenicol (right spots). All experiments were made at least 4 times with similar results. Line 0 shows CAT activity in protoplasts electroporated with calf thymus DNA.

binding site and, most remarkably, a stretch of 36 nts of unknown function (Fig. 2; also observed by [18]). Interestingly in all 3 cases this sequence is located in the bowl (Fig. 2 and 3) at the top of the structure. Thus the goblet consists of a central portion with conserved primary sequence ("bowl-sequence") imbedded in flanking portions with conserved secondary structure (the stem). This conservation of secondary structure, despite the lack of primary sequence conservation is probably indicative of some important function(s). It may be compared to the conservation of secondary structure among tRNAs and viroids. Moreover, conserved sequences in the single-strand loops of stems lacking sequence conservation have been shown to be involved in the post-transcriptional regulation of iron responsive mRNAs (7).

Ca379 is a CaMV mutant in which 147 nts have been deleted from the

leader (25). The deletion covers the bowl sequence plus 70 nts upstream and 10 nts downstream. The only change in the proposed secondary structure is the loss of the bowl and truncation of the stem (Fig.2). The biological consequence of the mutation is a dramatic increase in the latency period, from 2 weeks in the wild type to 12 weeks in the mutant. However testing the deletion in host plant protoplasts for its effect on transient expression of a reporter gene located downstream of the leader did not reveal an effect (Fig. 6, lines 3 and 4). Thus the sequence is either involved in regulation of expression at a specific stage of development or in response to environmental or metabolic conditions not represented by the protoplasts used, or has another function in virus replication. In any case the conservation of the bowl sequence between viruses from different caulimovirus groups is striking. Experiments to elucidate the function are in progress.

Presence of small open reading frames in the caulimovirus leaders.

Another striking similarity between the leaders of caulimoviruses concerns the number and distribution of the short ORFs (sORFs) they carry. The first 150 nucleotides of the leaders (corresponding to the R region) from all the caulimovirus sequences published contain either 2 or 3 very short (1 - 6 codons) ORFs. These sORFs (A, B and C) are followed by a stretch of 90 (FMV) to 190 (CERV) nucleotides lacking AUG codons and then by several longer sORFs (Fig. 7). The amino acid sequences of sORFs A - E are well conserved among the six CaMV strains (Fig. 7). The first three sORFs (A to C) of CaMV and CERV are homologous and encode peptides containing cysteine (sORFs A and B) or aromatic amino acids (sORF C) (Fig. 7). A feature of the subsequent sORFs (D', D and E) is that for each virus at least one encodes a distinctly basic peptide. CaMV and CERV sORFs E show some sequence homology. Although the sORFs of FMV differ in sequence from the other viruses, there are some homologies between sORFs C and D of FMV and sORFs D' and D of CERV. The sORFs at the extreme 3' end of the leader, if present at all, vary widely in sequence even among the CaMV strains. This is mainly due to changes in phase resulting from insertions or deletions. The loss of sORF B in strain CM1841 and the gain of sORF D' in strain CM4-184 are due to exchanges of ATG and ACG. ACG codons may, however, still allow translation initiation, albeit at a lower efficiency (36; M. Schultze, personal comm.). If non-AUG initiation codons are considered, the actual number of sORFs in the leader may be even higher.

In order to determine whether the sORFs in the leader are expressed in vivo, we fused the CAT reporter gene to the second codons of CaMV sORFs D

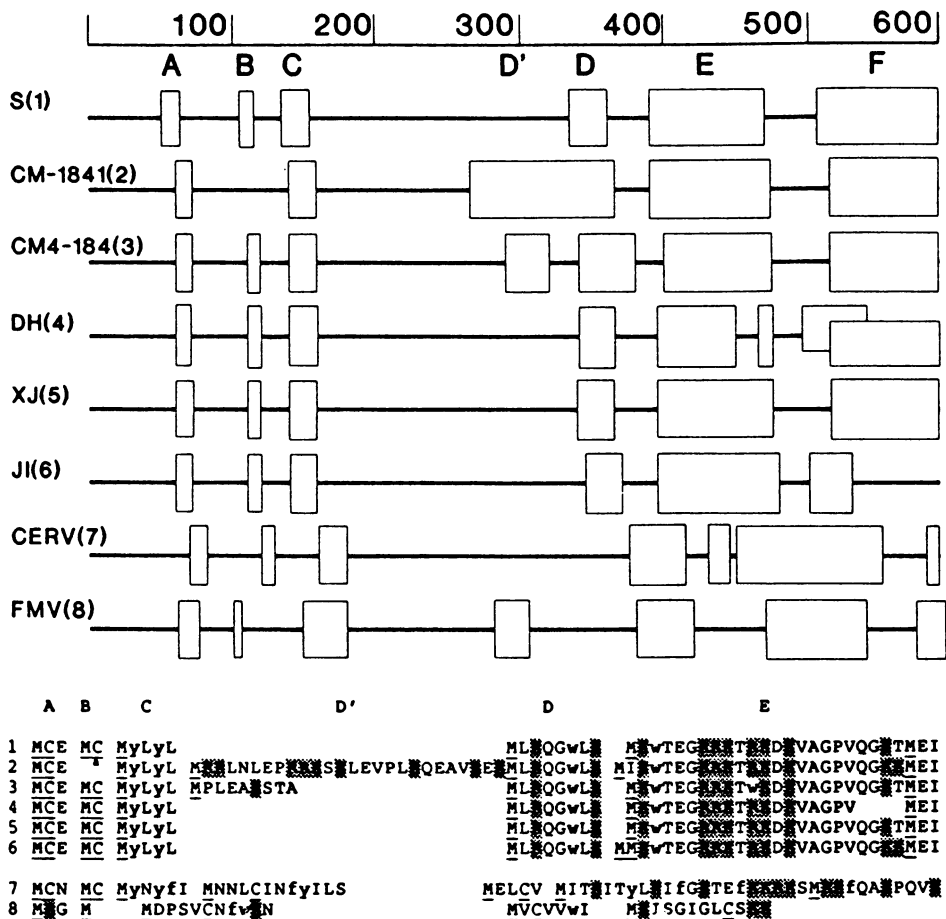


Fig.7. Location of short ORFs in the leader regions of caulimoviruses. The open boxes represent sORFs in the 600 nt leaders of six strains of CaMV (numbered 1-6), CERV (nr.7) and FMV (nr.8). The sORFs are labelled A - F. D' refers to the fourth sORF in (nr.3) and in (nr.7). Amino acid sequences of the peptides encoded by sORFs A - E are given in the lower part of the figure. Sulphur-containing residues are underlined, aromatic amino acids printed in lower case and basic residues stippled. The sequence of sORF F is omitted (see text).

and E. Both fusions induced comparable amounts of CAT activity in *Orychopragmus violaceus* protoplasts in a transient expression experiment (Fig.6, lines 1 and 2). This shows that the start codons of these sORFs are in a sequence context that allows efficient translation initiation and that these AUG codons are reached by ribosomes inspite of the upstream sORFs. A

possibility that must, however, be considered is that the AUGs in the leader are occluded within the secondary structure when the complete leader sequence is present.

A role for the leader

A complex mRNA leader sequence, with extensive secondary structure and sORFs, is likely to be involved in the regulation of translation. Both secondary structure of sufficient extent and the presence of small upstream ORFs generally reduce expression of a downstream gene (8-10, 37) but both features can be involved in regulation mechanisms (38-39). Small ORFs, e.g., serve as regulatory elements in response to environmental changes for the yeast GCN4 (2) and CPA1 (11) mRNAs. Interestingly, sORFs A, B and C of CaMV and CERV resemble those in the leader of the GCN4 mRNA. Translational regulation of gene expression by an interplay of RNA secondary structure and the expression of a small upstream ORF was suggested for SV40 coat protein synthesis (40), and for some bacterial antibiotic resistance genes (41), and in classical attenuation of the bacterial tryptophan operon (42).

The presence of a highly conserved sequence motif in the leader of all caulimoviruses and the exposed localization of this motif in the proposed secondary structure may be indicative for an interaction of this sequence with conserved cellular factors. Such factors could again act in translation control, possibly in response to conditions not covered by the transient expression assay used by us.

The CaMV 35S RNA probably has at least two functions in the life cycle of CaMV: as the mRNA for most of the viral genes (43,44) and as the template for reverse transcription. The distribution of the 35S RNA between these functions may be subject to regulation. It is interesting that in all caulimoviruses the primer binding site for reverse transcription is located downstream of the leader sequence in close vicinity to the initiation codon of a longer ORF known to be expressed in vivo and that this site occurs always within a small ORF so that translation could directly interfere with priming of reverse transcription. In addition, sequences at the 3' end of the leader, adjacent to the primer binding site, may play some role in reverse transcriptase binding. On the other hand, reverse transcription has so far always been found to be coupled to packaging of the template RNA (45). The Ψ (RNA packaging) signal in some retroviruses has been found to be located either fully (e.g. 46) or partially (47) within the leader of these viruses and a corresponding function is also to be expected in CaMV (45).

Finally, sequences regulating transcription could be located downstream of the RNA cap-site (48) and sequence and structure around the

polyadenylation signal might contribute to its facultative use, typical for retroviruses and other retroid elements.

Although it is possible to discern interesting features, including secondary structure, conserved sequence elements and SORFs in the caulimovirus leaders, we are a long way from understanding these complex RNAs. A number of approaches, including site directed mutagenesis, assays for infectivity in vivo, in vitro transcription and translation, RNase protection, transient expression experiments in transformed protoplasts, and studies of plants transgenic for virus constructs will help to sort out their functions.

ACKNOWLEDGEMENTS

We highly acknowledge the stimulating discussions and critical readings of this publication by Karin Wiebauer and Joe Jiricny, the generous gift of Orychopragmus suspension cells by C. Matsui, the privileged communication with J. Davies, and J. Stanley and the expert technical assistance of Hanny Schmid and Mathias Müller. The VAX computer and the relevant programs were selected and installed by Christoph Nager and we thank him and Thomas Nyffenegger for some valuable teaching sessions. A high degree of CaMV RNA secondary structure had been proposed at a conference (Jerusalem, 1983).

*To whom correspondence should be addressed

REFERENCES

1. Kozak, M. (1987) *Nucleic Acids Res.* **15**, 8125-8146.
2. Müller P.P. & Hinnebusch A.G. (1986) *Cell* **45** 201-207.
3. Muesing M.A., Smith D.H., & Capon D.J. (1987) *Cell* **48** 691-701.
4. Hann, S.R., King, M.W., Bentley, D.L., Anderson, C.W. & Eisenman, R. N. (1988) *Cell* **52**, 185-195.
5. Werner, M., Feller, A., Messenguy, F. & Piérard, A. (1987) *Cell* **49**, 805-813.
6. Darlix, J.L., Zuker, M. & Spahr, P.F. (1982) *Nucl. Acids Res.* **10**, 5183-5196.
7. Hentze, M.W., Caughman, S.W., Rouault, T.A., Barriocanal, J.G., Dancis, A., Harford, J.B. & Klausner, R.D. (1987) *Science* **238**, 1570-1573.
8. Marth, J.D., Overell, R.W., Meier, K.E., Krebs, E.G. & Perlmutter, R.M. (1988) *Nature* **332**, 171-173.
9. Chevrier, D., Vézina, C., Bastille, J., Linaud, Ch., Sonenberg, N. & Boileau, G. (1988) *J. Biol. Chem.* **263**, 902-910.
10. Baim, S.B. & Sherman, F. (1988) *Mol. Cell. Biol.* **8**, 1591-1601.
11. Werner, M., Feller, A., Messenguy, F. & Piérard, A. (1987) *Cell* **49**, 805-813.
12. Franck, A., Guilley, H., Jonard, G., Richards, K. & Hirth, L. (1980) *Cell* **21**, 285-294.
13. Gardner, R.C., Howarth, A.J., Hahn, P., Brown-Luedi, M., Shepherd, R. J., & Messing, J. (1981) *Nucl. Acids Res.* **9**, 2871-2888.
14. Balazs, E., Guilley, H., Jonard, G., Richards, K. (1982) *Gene* **19**, 239-249.

15. Dixon, L.K., Nyffenegger, T., Delley, G., Martinez-Izquierdo, J. & Hohn, T. (1986) *Virology* **150**, 463-468.
16. Rongxiang, F., Xiaojun, W., Ming, B., Yingchuan, T., Faxing, C. & Kekiang, M. (1985) *Chin. J. Virol.* **1**, 247-256.
17. Hull, R., Sadler, J. & Longstaff, M. (1986) *EMBO J.* **12**, 3083-3090.
18. Richins, R.D., Scholthof, H.B. & Shepherd, R.J. (1987) *Nucl. Acids Res.* **15**, 8451-8466.
19. Guilley, H., Dudley, R.K., Jonard, G., Balazs, E. & Richards, K. (1982) *Cell* **30**, 763-773.
20. Weiss, R., Teich, N., Varmus, H. & Coffin, J. (1984, 1985) *RNA tumor viruses*. 2 Vols. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
21. Caballero, L., Devereux, J., Karp, Y., Marquess, P., Smith, M. & Winsborough, W. (1987) *Sequence Analysis Software Package* Vers. 5. University of Wisconsin Genetics Computer Group.
22. Zucker, M. & Stiegler, P. (1981) *Nucl. Acids Res.* **9**, 133-148.
23. Freier, S.M., Kierzek, R., Jaeger, J.A., Sugimoto, N., Caruthers, M.H., Neilson, T. & Turner, D.H. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 9373-9377.
24. Maizel, J.V. & Lenk, R.R. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 7665-7669.
25. Penswick, J., Hübler R., & Hohn, T. (1988) *J. Virol.* **62** 1460-1463.
26. Howarth, A.J., Gardner, R.C., Messing, J. & Shepherd, R.J. (1981) *Virology* **112**, 678-685.
27. Lebeurier, G., Hirth, L., Hohn, T. & Hohn, B. (1980) *Gene* **12**, 139-146.
28. Pietrzak, M., Shillito, M., Hohn, T. & Potrykus, I. (1986) *Nucl. Acids Res.* **14**, 5857-5868.
29. Mead, D.A., Szczesna-Skorupa, E. & Kemper, B. (1986) *Protein Engineering* **1**, 67-74.
30. Matsui, C., (1985), doctoral thesis, Nagoya University, Japan
31. Fromm, M. E., Taylor, L. P. & Walbot, V. (1986) *Nature* **319**, 791-793.
32. Gorman, C.M., Moffat, L.F. & Howard, B.H. (1982) *Mol. Cell. Biol.* **2**, 1044-1051.
33. Pleij, C.W.A., Rietveld, K. & Bosch, L. (1985) *Nucl. Acids Res.* **13**, 1717-1731.
34. Puglisi, J.D., Wyatt, J.R. & Tinoco, I., Jr. (1988) *Nature* **311**, 283-286.
35. Ehresmann, C., Baudin, F., Mougel, M., Romby, P., Ebel, J.P. & Ehresmann, B. (1987) *Nucl. Acids Res.*, **15** 9109-9128.
36. Curran, J. & Kolakofsky, D. (1988) *EMBO J.* **7**, 245-251.
37. Kozak, M. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 2850-2854.
38. Hackett, P.B., Petersen, R.B., Hensel, C.H., Albericio, F., Gunderson, S.I., Palmenberg, A.C. & Barany, G. (1986) *J. Mol. Biol.* **190**, 45-57.
39. Hull, R. (1984) *Plant Mol. Biol.* **3**, 121-125.
40. Hay, N. & Aloni, Y. (1985) *Mol. Cell. Biol.* **5**, 1327-1334.
41. Alexieva, Z., Duvall, E.J., Ambulos, N.P., Jr., Kim, U.J. & Lovett, P.S. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 3057-3061.
42. Yanofsky, C. (1987) *Trends in Genetics* **3**, 356-360.
43. Dixon, L. & Hohn, T. (1984) *EMBO J.* **3**, 2731-2736.
44. Sieg, K. & Gronenborn, B. (1982) *Abstract NATO Adv. Study Inst.* **154**.
45. Fütterer, J. & Hohn, T. (1987) *TIBS* **12**, 92-95.
46. Watanabe, S. & Temin, H.M. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 5986-5990.
47. Bender, M.A., Palmer, T.D., Gelinas, R.E. & Miller, D. (1987) *J. Virol.* **61**, 1639-1646.
48. Serfling, E., Jasin, M. & Schaffner, W. (1985) *TIG* **1**, 224-230.