The EMBO Journal vol.14 no.14 pp.3552–3562, 1995

Splicing of cauliflower mosaic virus 35S RNA is essential for viral infectivity

Zsuzsanna Kiss-László, Stéphane Blanc[1] and Thomas Hohn[2]

Friedrich Miescher Institute, P.O. Box 2543, CH-4002 Basel, Switzerland and [1]Station de Recherches de Pathologie Comparée, INRA-CNRS, 30 380 Saint Christol-lez-Alès, France

[2]Corresponding author

A splicing event essential for the infectivity of a plant pararetrovirus has been characterized. Transient expression experiments using reporter constructs revealed a splice donor site in the leader sequence of the cauliflower mosaic virus (CaMV) 35S RNA and three additional splice donor sites within open reading frame (ORF) I. All four donors use the same splice acceptor within ORF II. Splicing between the leader and ORF II produces an mRNA from which ORF III and, in the presence of the CaMV translational transactivator, ORF IV can be translated efficiently. The other three splicing events produce RNAs encoding ORF I–II in-frame fusions. All four spliced CaMV RNAs were detected in CaMV-infected plants. Virus mutants in which the splice acceptor site in ORF II is inactivated are not infectious, indicating that splicing plays an essential role in the CaMV life cycle. The results presented here suggest a model for viral gene expression in which RNA splicing is required to provide appropriate substrate mRNAs for the specialized translation mechanisms of CaMV.

Key words: cauliflower mosaic virus/gene expression/pararetrovirus/splicing

Introduction

Cauliflower mosaic virus (CaMV) is a plant pararetrovirus that has been the focus of extensive study, both because of its widely used transcriptional control signals, and because it has evolved specialized translational control mechanisms (for reviews, see Hohn and Fütterer, 1992; Rothnie *et al.*, 1994). CaMV produces two primary transcripts (Covey and Hull, 1981), the 19S RNA and the 35S RNA (Figure 1A). The 19S RNA is a monocistronic mRNA covering ORF VI. The 35S RNA is terminally redundant (Guilley *et al.*, 1982) and covers the whole genome with its seven major open reading frames (ORFs). In addition to being the template for reverse transcription (Hull and Covey, 1983), it is also used as an mRNA for at least some viral proteins.

The use of the 35S RNA as mRNA is thought to be regulated by the 600 nt leader, which contains several small ORFs (sORFs) that would be predicted to inhibit the classical translational scanning process (Kozak, 1989). Translation in the presence of the leader is nevertheless possible and has been interpreted as being accomplished by a shunt of the scanning complex, allowing inhibitory regions to be bypassed (Fütterer *et al.*, 1993). The polycistronic nature of the 35S RNA is another feature requiring modification of standard translation models. Translation of ORFs on the 35S RNA has been proposed to occur by a 'relay race' mechanism (Sieg and Gronenborn, 1982; Dixon and Hohn, 1984) in which ribosomes re-initiate protein synthesis at the nearest AUG codon after passing a termination codon. This process requires the presence of a translational transactivator (TAV), encoded by ORF VI of CaMV (Bonneville *et al.*, 1989), which is itself translated from the monocistronic 19S RNA (Figure 1A). Translational transactivation functions on artificial dicistronic mRNAs (Fütterer and Hohn, 1991) and is independent of the length and sequence of the ORFs and the region between them, but is strongly enhanced if a sORF precedes the first large ORF (Fütterer and Hohn, 1992). A similar mechanism exists in the related figwort mosaic virus (FMV; Gowda *et al.*, 1989; Scholthof *et al.*, 1992). In the viral context, transactivated translation has been demonstrated in the case of ORF I translation (Bonneville *et al.*, 1989; Zijlstra and Hohn, 1992; DeTapia *et al.*, 1993), and relay race models of CaMV and FMV gene expression have been suggested in which TAV might transactivate translation of all viral ORFs on the pre-genomic RNA (Fütterer *et al.*, 1990b; Scholthof *et al.*, 1992), including ORF VI itself (Driesen *et al.*, 1993).

Although models invoking the mechanisms of non-linear ribosome scanning and transactivation of translation from polycistronic mRNAs can potentially explain translation of all CaMV ORFs, here we re-examine the premise that these translational strategies act on the full-length 35S RNA. It has been a long standing question whether other viral RNA species are produced, either by transcription from an internal promoter or by splicing of the 35S RNA, and whether such mechanisms would allow alternative expression strategies for some, or all, viral gene products. Such RNAs might have escaped attention so far in Northern blots of extracts from infected plants, because these are obscured by a smear of partially digested virus RNA molecules released during reverse transcription. Although there are indications that splicing can occur in caulimoviruses [some naturally occurring deletion mutants of CaMV and FMV can be interpreted as having arisen by reverse transcription of spliced viral RNAs (Hirochika *et al.*, 1985; Vaden and Melcher, 1990; Scholthof *et al.*, 1991)] and also in the human hepatitis B virus (Chen *et al.*, 1989), evidence for a functionally important role of splicing in viral gene expression was lacking.

Unlike in retroviruses, where splicing is essential in order to express certain viral proteins, it has previously been thought that there is no obligate role for splicing in either plant or animal pararetroviruses. The finding of

splicing in rice tungro bacilliform (badna) virus (RTBV; Fütterer *et al.*, 1994) was the first case to break this rule.

In this work, we describe the detection of spliced CaMV RNAs in infected plants and transfected protoplasts. Four splice donors were identified, one in the leader and three within ORF I, all of which can use the same splice acceptor within ORF II. We show that splicing to this acceptor is essential for virus viability. These results lead to the modification of current models of CaMV gene expression to include the essential contribution of RNA splicing to producing mRNAs from which viral proteins can be translated via the specialized translation mechanisms described above.

Results

Transient expression of reporter gene fusions to CaMV ORFs

Expression of individual ORFs of the CaMV 35S RNA was investigated in transient expression experiments. A series of plasmids was constructed in which specific CaMV ORFs, together with their preceding upstream sequences, were fused to either the β-glucuronidase (GUS; Figure 1B) or the chloramphenicol acetyltransferase (CAT; Figure 1C) reporter genes. Both reporter genes retain their original start codons. The resulting constructs were used to measure reporter gene activity relative to appropriate monocistronic control plasmids after transfection of cell culture protoplasts derived from the CaMV host plant *Orychophragmus violaceus*.

Confirming earlier results (Bonneville *et al.*, 1989; Fütterer *et al.*, 1990a), we found considerable levels of expression from a fusion to the first ORF (i.e. ORF VII; pVII–GUS) and very low levels from fusions to ORFs I and IV (pI–GUS; pIV–CAT). Surprisingly, GUS fusions to codon 110 of ORF II (pIIb–GUS) and to codon 34 of ORF III (pIII–GUS) gave rise to significant constitutive levels of expression. In all cases other than the monocistronic controls, expression was increased in the presence of TAV, but the effect of transactivation was markedly stronger with reporter gene fusions to ORFs I and IV, than to ORFs II or III. Expression of an ORF V–GUS fusion was barely detectable with or without transactivation (data not shown).

Comparison of GUS expression from the monocistronic control plasmids pmono–GUS and pmonoIII–GUS to that of pmonoIIb–GUS revealed that the latter GUS fusion is intrinsically very poorly expressed. The reduction of GUS activity is presumably the result of the fusion of 110 amino acids from ORF II at the GUS N-terminus (such N-terminal extensions of the GUS protein have been reported to inhibit GUS activity; Restrepo *et al.*, 1990; Varagona *et al.*, 1992). This suggested that the GUS expression observed with pIIb–GUS might result from translation beginning at the AUG of the GUS gene itself, which is retained in this construct. Consistent with this interpretation, the introduction of a frameshift mutation into ORF II upstream of the GUS gene did not affect GUS expression significantly (Figure 1B, pIIb–GUSfs). Two additional constructs were prepared in which CAT (which is unaffected by the N-terminal extension) was used as the reporter gene. In pIIa–CAT, the reporter gene is fused to codon 1 of ORF II; pIIb–CAT is analogous to

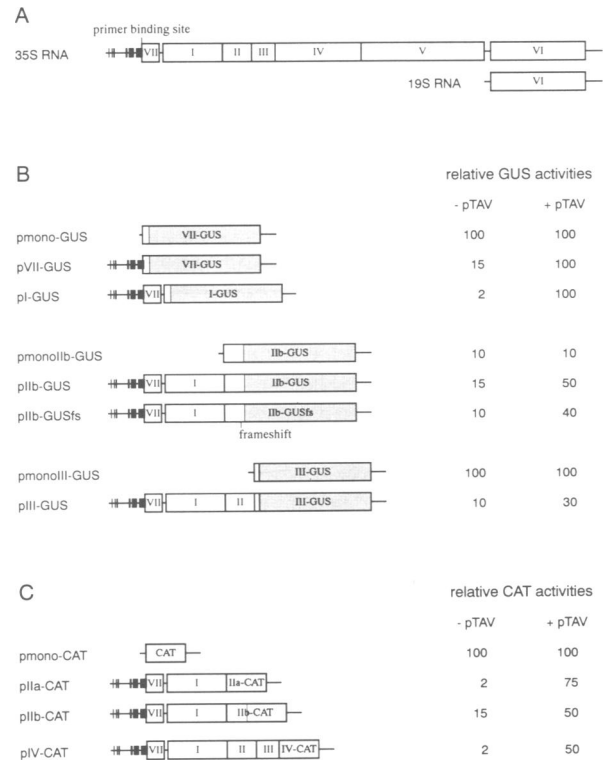


Fig. 1. Transient expression of downstream ORFs on the polycistronic 35S RNA. (**A**) Map of CaMV primary transcripts showing their large ORFs (VII–VI) [ORF VII, unknown function; ORF I, movement protein; ORF II, aphid transmission factor (ATF); ORF III, DNA binding minor capsid component; ORF IV, capsid protein; ORF V, enzymatic functions; ORF VI, cytoplasmic inclusion body protein and transactivator (TAV)]. The six sORFs in the leader are shown as black lines. The primer binding site for reverse transcription of the 35S RNA is indicated. (**B**, **C**) GUS or CAT reporter genes are fused to ORF VII, I, II, III or IV and expression of the chimaeric fusions after transfection of protoplasts is compared with the appropriate leaderless monocistronic control with (+) or without (–) supplying the transactivator plasmid (pTAV). The position of the frameshift mutation introduced into pIIb–GUSfs is indicated. The reporter coding regions are shaded.

pIIb–GUS, with the AUG of CAT being fused to codon 110 of ORF II (Figure 1C). Reporter gene expression from pIIb–CAT was similar to that from pIIb–GUS, with a significant basic level of expression which was only moderately increased upon transactivation. In contrast, CAT was only very poorly expressed from pIIa–CAT but, upon addition of TAV, the level of expression was stimulated by a factor similar to that observed for the I–GUS fusion.

Taken together, these results suggest that translation of the reporter genes in the pIIb fusions is initiated at the original GUS or CAT AUG codon and that the effect of TAV on translation depends on whether the II AUG or the reporter gene AUG 110 codons further downstream is used. This in turn suggests that translation initiating at the reporter gene AUG may occur on a separate RNA, lacking the II AUG, originating either from transcription from an internal promoter or by splicing. To search for such an RNA, we performed RNase A/T1 protection analysis of pIIb–GUS transcripts isolated from transfected *Arabidopsis thaliana* protoplasts (*Arabidopsis* protoplasts were used for RNA analysis because of the difficulty in isolating high amounts of undegraded RNA from transfected

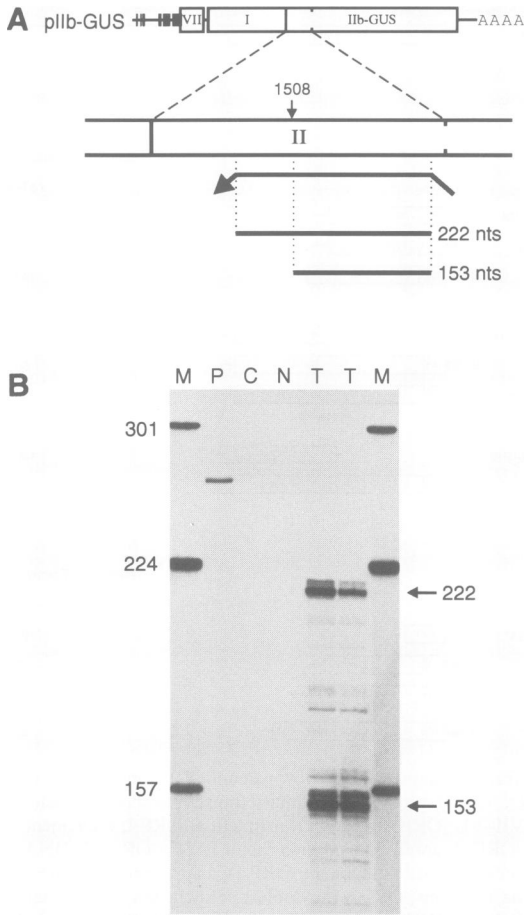


Fig. 2. RNase A/T1 analysis of pIIb–GUS transcripts. (**A**) Schematic representation of the primary transcript produced from pIIb–GUS. The antisense RNA probe complementary to 222 nt of ORF II is represented by an arrow. Thick lines show protected fragments corresponding to unspliced (222 nt) or spliced (153 nt) transcripts. The putative splice acceptor site at position 1508 is indicated. (**B**) RNAs were mapped from either non-transfected (N) or transfected *A.thaliana* protoplasts (T; duplicates). Lanes (P) and (C) represent non-digested and RNase A/T1-treated probe, respectively. (M): ^{32}P-labelled DNA fragments (pBR322/*Hin*fI).

O.violaceus protoplasts; there was no significant difference in the results of reporter gene expression in the two protoplast systems). An antisense probe covering 222 nt of ORF II (still present in pIIb–GUS) protected a fragment of 153 nt in addition to the expected 222 nt fragment derived from protection of the full-length transcript (Figure 2). The shorter fragment suggests that either a splice acceptor site or a transcription start site is located between the AUG of ORF II and the GUS gene in pIIb–GUS. We could find no evidence for any promoter activity in the sequences upstream of this region (Z.Kiss-László, unpublished data), suggesting that splicing of the 35S RNA was the more likely explanation. Analysis of naturally occurring deletion mutants of CaMV suggested that a splice acceptor might be present at position 1508 within ORF II (Hirochika *et al.*, 1985). In the experiment shown in Figure 2, spliced RNA using an acceptor at this position would result in a protected fragment of 153 nt, which is exactly what we observe. Thus, in Figure 2, we align the 153 nt fragment to the putative splice acceptor at position 1508.

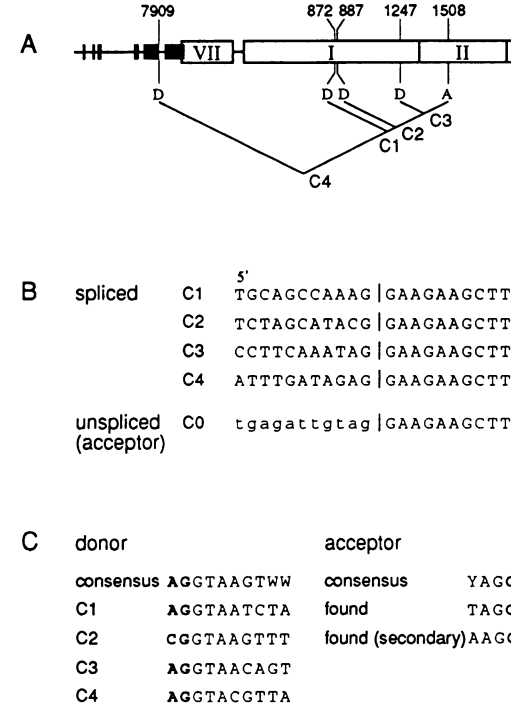


Fig. 3. Localization of the splice sites in the pIIb–GUS transcript. (**A**) Schematic representation of splicing events between the four splice donor sites found at positions 7909 at the 3′ end of the leader and 872, 887, 1247 within ORF I and the acceptor site in ORF II (1508). (**B**) The sequences around the splicing junctions found in the cDNAs are shown in all four cases (C1–C4). C0 indicates the sequence of unspliced transcripts. (**C**) Sequence comparison of the splice sites found in CaMV and plant consensus splicing signals.

Characterization of spliced RNAs from transfected protoplasts

To determine whether the 153 nt protected fragment was derived from RNA splicing, we analysed transcripts produced from the pIIb–GUS construct in transfected *A.thaliana* protoplasts by RT–PCR. First strand cDNA molecules were synthesized using random primers, and a phosphorylated 20mer oligonucleotide was ligated to their 3′ ends (see Materials and methods). The resulting ligation products were PCR amplified using a primer located 35 nt downstream of the putative splice acceptor site and a second primer complementary to the oligonucleotide extension. Sequence analysis of 50 individual clones indicated five classes of cDNAs (C0–C4; Figure 3). Each class was represented by at least five independent clones, with the exception of class C4 (two clones). Class C0 represented the RT–PCR-amplified unspliced 35S RNA. Analysis of the sequences of classes C1–C4 confirmed that they had undergone RNA splicing, using a splice acceptor site at position 1508 of the CaMV sequence, i.e. at codon 53 of ORF II, which is located between the II AUG and the fusion site in pIIb–GUS. The splice donor sites were located at positions 872, 887 and 1247 within ORF I (D1, D2 and D3) and at position 7909 close to the 3′ end of the leader (D4). All sites conform to plant splice donor and acceptor consensus sequences (Figure 3; Filipowicz *et al.*, 1995) and these sites are conserved in all available sequences of CaMV isolates.

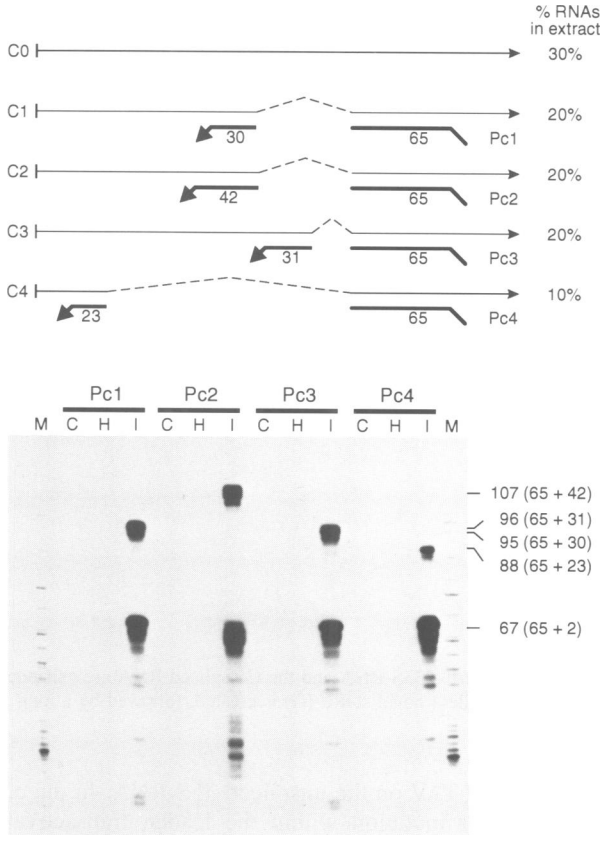


Fig. 4. Characterization of spliced transcripts from turnip plants infected with CaMV-S strain. RNase A/T1 mapping of total RNA isolated from CaMV-S strain-infected plants (I) and healthy plants (H). Transcripts were characterized using antisense cDNA probes derived from RT–PCR covering all four splicing junctions (Pc1–Pc4). Digested probes are represented in lanes C. For each probe, protected fragments larger than 67 nt indicate transcripts processed using the corresponding splice donor (Figure 3). The remaining unspliced RNAs appear as protected bands of 67 nt (2 bp longer than expected from the cleavage at the 3′ splice site because of 2 nt additional homology between the cDNAs and unspliced transcripts (AG). The fraction of the RNAs represented in cell extract is depicted in (%)s.

Detection of novel viral mRNAs from infected plants

To verify that splicing also occurs in intact plants during virus infection, RNA isolated from CaMV-infected turnip plants was subjected to RNase A/T1 mapping, using derivatives of the cDNAs C1–C4 covering the respective splice junctions as probes (Figure 4A). With each of these probes a spliced RNA species was detected, represented by a protected fragment of larger than 67 nt (Figure 4B). The 67 nt fragment corresponds in each case to the sum of unspliced transcripts and spliced transcripts using the other three donor sites. The relative abundance of each of the RNAs was determined by phosphor-imaging analysis. In infected plant cells, the splice donor sites within ORF I were used with an efficiency of 20% for each case, while the splice donor site in the leader was used with an efficiency of 10% (Figure 4); the remaining RNA population (30%) had not used the splice acceptor at position 1508 within ORF II. These values are consistent with those obtained from the mapping analysis in transfected protoplasts using an ORF II probe (Figure 2), suggesting that quantification is accurate and that there are no additional RNAs spliced in this region which had escaped our attention.

We also analysed CaMV-specific DNA in infected plants. PCR analysis did not reveal DNA sequences that might have originated from deletion events that could have served as templates for shortened transcripts or that might have been derived from reverse transcription of the spliced species (data not shown). This shows that most, possibly all, the spliced RNA is excluded from the reverse transcription process.

Notably, all of the splicing events between ORF I and ORF II produced in-frame ORF I–II fusions. Splicing from the leader to ORF II generates an mRNA consisting of the whole leader sequence up to CaMV position 7909 and a new sORF (sORF G), which is followed by a longer start codon-free region and ORF III (Figure 5). The sORF G is present out-of-frame within ORF II of CaMV-S strain.

Translation from the RNA spliced between the leader and ORF II

To test whether splicing between the leader and ORF II would result in a functionally active mRNA, we constructed pIII–GUS-C4, a plasmid derived from pIII–GUS lacking the intron between the leader and codon 53 of ORF II (Figure 6A). Upon transfection into *O.violaceus* protoplasts, pIII–GUS-C4 gave rise to a significant level of GUS expression, which did not respond significantly to the presence of TAV (Figure 6A). This result confirmed that ORF III can be translated efficiently and constitutively from the RNA produced by this splicing event. We also tested expression from an ORF IV fusion in the context of the C4-type leader. The level of CAT activity from this construct, pIV–CAT-C4 (Figure 6B), is very similar to that obtained with the original pIV–CAT, supporting the hypothesis that the ORF IV fusion in each case is expressed from the same RNA, i.e. the C4-spliced RNA. CAT expression from both these constructs is significantly increased upon transactivation (Figure 6B). These results suggest that ORF III is expressed constitutively and that ORF IV can be expressed from the C4-spliced RNA in the presence of TAV.

Splice site mutations reduce ORF III expression in transiently transfected protoplasts

Further evidence that removal of the intron from the 35S RNA is involved in ORF III expression came from a series of mutations of the splice signals. The splice donor and acceptor sites of pIII–GUS were mutated using oligonucleotide-directed site-specific mutagenesis. Expression of the mutants was tested in transfected *O.violaceus* protoplasts. Mutation of the splice donor site within the leader (GGT→CCA; pIII–GUSmd; see Figure 3 for the context), or the splice acceptor site within ORF II (AGG→CCG; pIII–GUSma1; see Figure 8 for sequence context), nearly abolished constitutive expression of the III–GUS fusion (Figure 7A). We also tested another mutant, which carries an AGG→AGA mutation in the 3′ splice site. Introduction of this sequence into pIII–GUS (pIII–GUSma3) also resulted in a dramatic decrease in GUS expression (Figure 7A). RNase A/T1 protection analysis confirmed the lack of the relevant spliced products for all three mutants (Figure 7B). Phosphor-imaging analysis showed that although the steady-state level of RNA

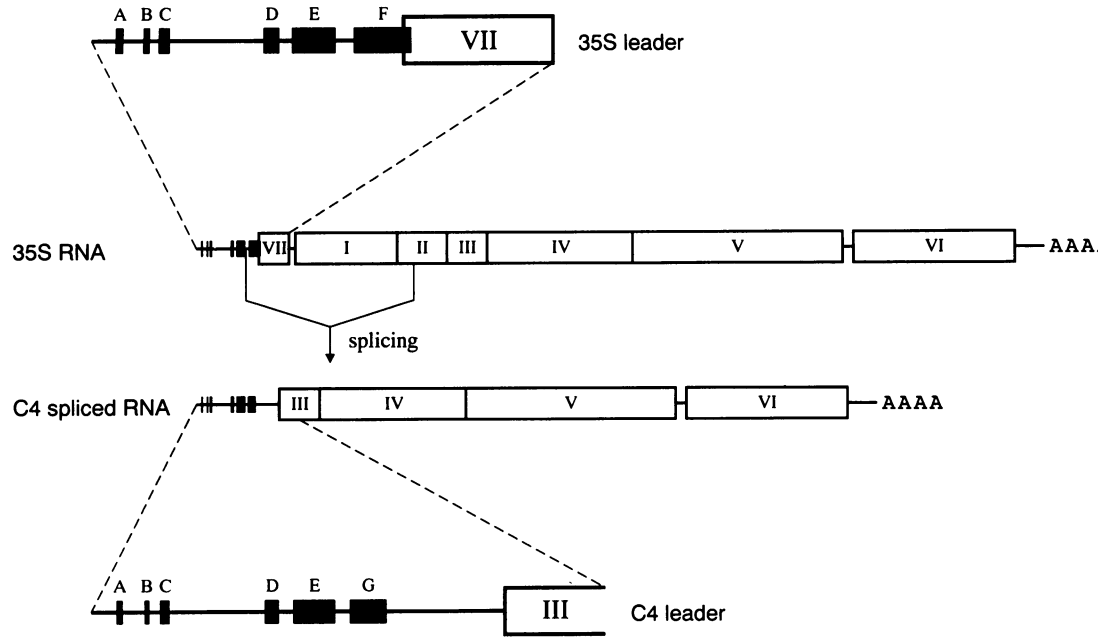


Fig. 5. Differences within the leader of the 35S RNA before and after splicing. The leaders of the 35S RNA and the C4-spliced RNA are expanded to show the arrangement of the sORFs. After removal of the intron, sORF F is eliminated while a novel sORF (G) is created, followed by a long AUG-free region upstream of ORF III.

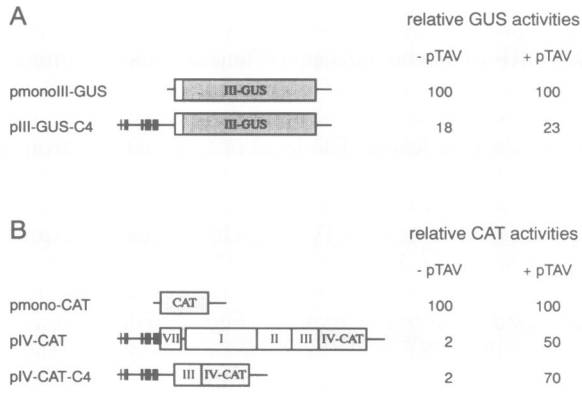


Fig. 6. Translation from C4-type pIII–GUS and pIV–CAT constructs. (**A**) and (**B**) represent the activities of GUS and CAT reporter genes respectively in the context of pIII–GUS and pIV–CAT fusions lacking the intron between the leader and ORF II with (+) and without (–) transactivation (pTAV).

produced by the mutants was reduced, spliced species could still potentially have been detected, but none were observed even after long exposure.

Because we also wanted to analyse splice acceptor site mutations in the context of the virus without altering the amino acid sequence of ORF II, an AGG→CGG mutation was introduced into pIII–GUS (pIII–GUSma2) which did not change the coding sequence. Unexpectedly, this mutant still expressed GUS, albeit with lower efficiency. RNase A/T1 protection analysis revealed that this mutation had not inactivated the acceptor (Figure 7B); the C4 splicing event still occurs with an efficiency of 7% (wild-type value = 10%).

In the presence of TAV, GUS expression could be observed with all splice site mutants. A low level of transactivated expression from pIII–GUSma1 (Figure 7A) indicates that translation of ORF III might also occur by the action of TAV on the unspliced 35S RNA. In the case of the donor mutation within the leader, transactivated expression was lower than in wild-type but higher than in the acceptor mutations, probably reflecting expression from RNAs with ORF I–ORF II fusions.

Viral mutants carrying mutations inactivating the 3′ splice site are not infectious

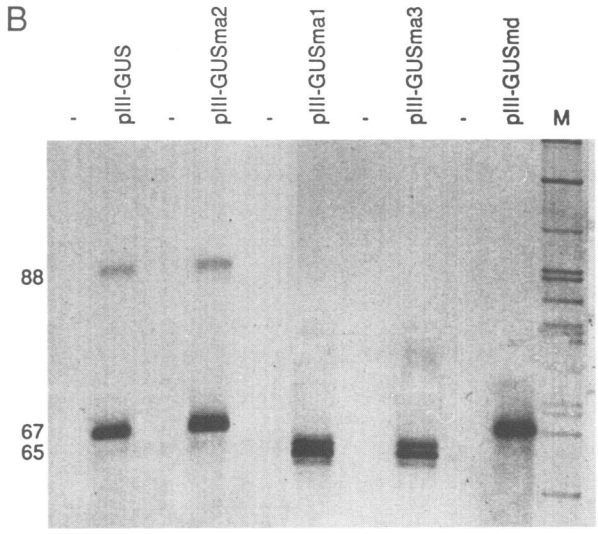
To examine the significance of this splicing event for viral replication, we studied the effect of mutations in the splice acceptor in ORF II on viability of CaMV. We tested plasmid-derived virus DNA from virus mutants Ca37ma1, ma2, and ma3, harbouring the same 3′ splice site mutations as the respective pIII–GUS series (Figure 7A). Turnip plants were inoculated mechanically with plasmid DNA carrying either wild-type Ca37 (Lebeurier *et al.*, 1980), or one of the mutant derivatives. After 3 weeks, plants inoculated with Ca37ma1 and Ca37ma3 were still symptom free, while plants inoculated with Ca37ma2 or Ca37 showed distinct systemic symptoms (Figure 7C). No symptoms ever appeared on the 60 plants inoculated with Ca37ma3.

The results obtained with the viral mutants correlate with the expression data, i.e. splice acceptor mutants which are inactive (ma1 and ma3; see Figure 7B) do not express GUS in the pIII–GUS context and are not infectious in the viral context. This strongly suggests that splicing using the acceptor at position 1508 is essential for virus viability.

After an additional 3 weeks, two of the 60 plants inoculated with Ca37ma1 finally showed symptoms. Rare and delayed appearance of symptoms usually suggests that the original inoculum was defective and that first or second site reversions restored virus viability leading finally to symptom appearance. To determine the nature of the reversion, DNA fragments containing the region

A

	5' splice site 7909	3' splice site 1508	relative GUS activities - pTAV	+ pTAV
pmonoIII-GUS	–	–	100	100
pIII-GUS	wt	wt	10	30
pIII-GUSmd	GGT to CCA	wt	2	20
pIII-GUSma1	wt	AGG to CCG	2	6
pIII-GUSma2	wt	AGG to CGG	6	25
pIII-GUSma3	wt	AGG to AGA	2	8



C

virus	infectious
Ca37	+
Ca37ma1	-
Ca37ma2	+
Ca37ma3	-
Ca37ma1 revertants	+

Fig. 7. Effect of different splice site mutations on reporter gene expression in transfected protoplasts and on virus viability in infected plants. (**A**) Mutations in either the 5′ splice site in the leader or the 3′ splice site in ORF II are indicated. Reporter gene activity was measured after transfection of protoplasts with (+) and without (–) transactivation (pTAV). The values are the rounded means of 10 independent experiments and are calculated relative to the expression value of pIII–GUS. (**B**) RNase A/T1 mapping of pIII–GUS derivatives. RNA fragments are mapped from non-transfected (–) and transfected protoplasts using the Pc4 antisense RNA probe (see Figure 4). C4-specific protected fragments (88 bp) are seen with pIII–GUS and pIII–GUSma2, while this spliced transcript is not seen in the case of pIII–GUSma1, ma3 and md. Protected fragments corresponding to the non-spliced transcripts indicated as 65+2 (pIII–GUS, pIII–GUS ma2, pIII–GUSmd) contain 2 nt additional homology between the Pc4 probe and non-spliced transcripts. This 2 nt homology is not present in ma1 and ma3; the unspliced transcripts from these constructs generate protected fragments of 65 bp. M: ^{35}S-labelled G-ladder of pBluescript II. (**C**) Virus mutants carrying the same mutations as the corresponding constructs in A. The results of the infectivity test are depicted by (+) and (–).

between 1037 in ORF I and 1931 in ORF III were isolated by PCR from plants infected with Ca37ma1 and 2 and ORF II was sequenced. No change with respect to the inoculated DNA was seen for Ca37ma2. In Ca37ma1, the original mutation was retained, but a second site mutation had occurred downstream of the first mutation: a deletion of 58 bp truncating ORF II by bringing a stop codon in-frame with the coding sequence (Figures 7C and 8A). Inspection of the deleted sequences revealed G-rich stretches which could inhibit a secondary splice acceptor site located further downstream on the RNA (Figure 8A).

To study the mechanism of ORF III expression in second site revertant viruses, we introduced the rearranged sequences of ORF II, harbouring the ma1 splice site mutation (AGG→CCG) and the 58 bp deletion, into the pIII–GUS context. This pIII–GUSma1rev construct gave rise to levels of constitutive and transactivated GUS expression similar to those observed for pIII–GUS (Figure 8B). This suggested that a cryptic splice site(s) may have been used. Inspection of sequences downstream of the deletion in the revertants located potential splice acceptors. In addition, sequence comparison between our second site revertants and a naturally occurring deletion mutant, which lacks 421 bp of ORF II (CM4-184; Howarth *et al.*, 1981), revealed a putative 3′ splice site which is conserved in CM4-184 (Figure 8A). To check whether this site is used in CM4-184, the 421 bp deletion in ORF II was introduced into the pIII–GUS context (pIII–GUSCM4-184). Constitutive and transactivated GUS expression from pIII–GUSCM4-184 was comparable with the pIII–GUSma1rev and pIII–GUS constructs (Figure 8B), again indicating use of this 3′ splice site. RNase A/T1 mapping confirmed that this site at position 1385 is indeed used (Figure 9). An antisense RNA probe covering the region of ORF II (still present in strain CM4-184) revealed a 67 nt protected fragment corresponding to the expected spliced fragment. A 103 nt protected fragment represented the non-spliced RNAs. These data are consistent with the interpretation that splicing is required for ORF III expression not only in CaMV-S (Strasbourg) but also in strain CM4-184, and suggests that activation of the cryptic splice site in pCa37ma1rev is responsible for restored viability.

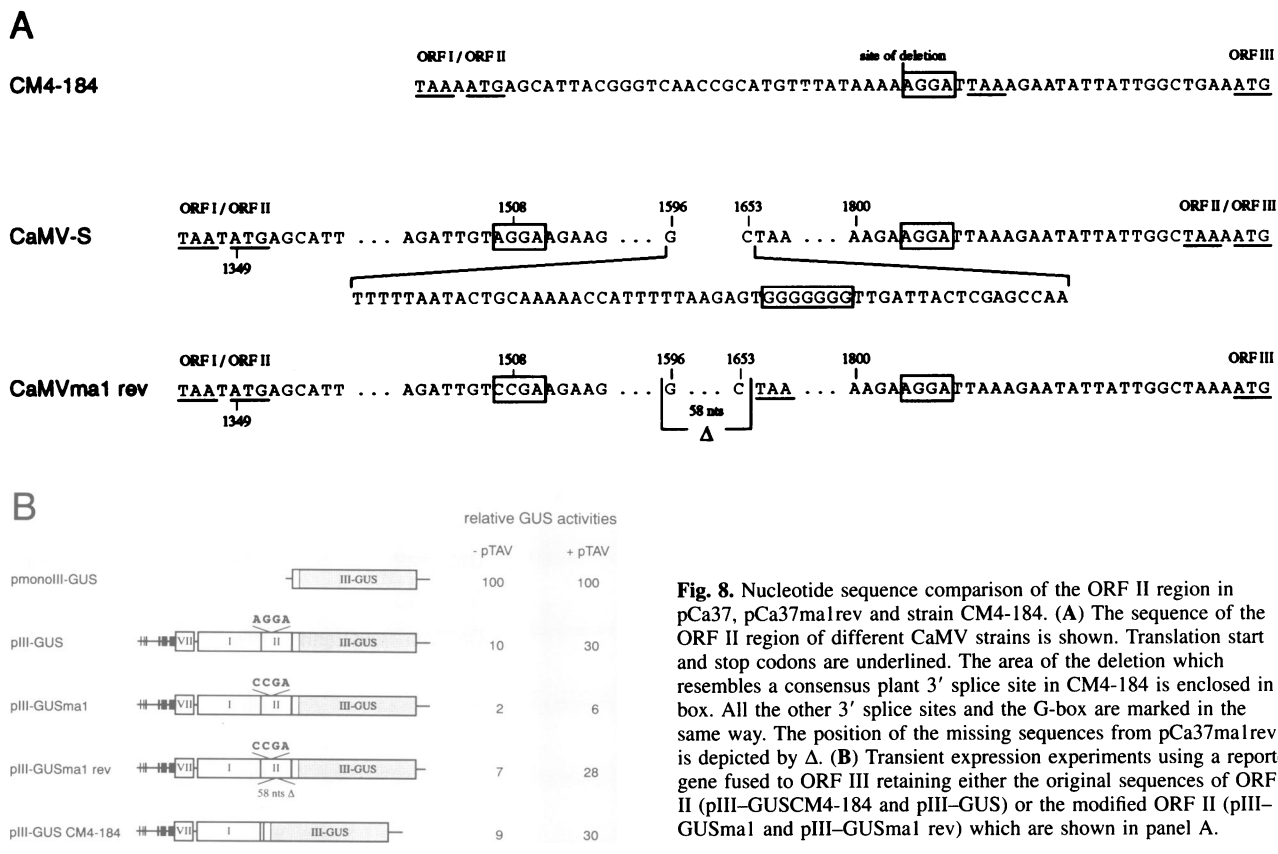


Fig. 8. Nucleotide sequence comparison of the ORF II region in pCa37, pCa37ma1rev and strain CM4-184. (**A**) The sequence of the ORF II region of different CaMV strains is shown. Translation start and stop codons are underlined. The area of the deletion which resembles a consensus plant 3′ splice site in CM4-184 is enclosed in a box. All the other 3′ splice sites and the G-box are marked in the same way. The position of the missing sequences from pCa37ma1rev is depicted by Δ. (**B**) Transient expression experiments using a reporter gene fused to ORF III retaining either the original sequences of ORF II (pIII–GUSCM4-184 and pIII–GUS) or the modified ORF II (pIII–GUSma1 and pIII–GUSma1 rev) which are shown in panel A.

Discussion

The mechanism of expression of downstream ORFs on the polycistronic CaMV 35S RNA was investigated in transient expression experiments. Four novel RNA species were detected as transcripts of a CaMV-derived reporter RNA in transfected protoplasts. Analysis of cDNA copies of these RNAs revealed deletions either from a site close to the 3′ end of the CaMV leader or from three alternative positions within ORF I to one site within ORF II. The corresponding RNAs were also detected in CaMV-infected plants.

Three lines of evidence support the conclusion that these deletions are derived from the primary transcript by splicing. First, in all four cases, consensus splice donor and acceptor sequences were used. Second, mutations of the proposed splice donor and acceptor sites in the chimaeric pIII–GUS plasmid resulted in considerable reduction of GUS activity in transfected protoplasts. Third, no deleted version of CaMV DNA, from which the novel RNA species could have been transcribed, could be detected by PCR analysis of DNA from infected plants. In addition, the splice donor and acceptor sites are conserved in all available CaMV strains.

Splicing from the leader to ORF II creates an mRNA from which ORF III can be translated efficiently in the absence of transactivator. The other three splicing events produce fused ORFs comprising parts of ORF I and ORF II.

The most compelling evidence that splicing is essential for virus viability was obtained from analysis of the infectivity of virus mutants. Mutations abolishing splicing (pCa37ma1 and pCa37ma3) render CaMV non-infectious unless gene rearrangements occur. The rare appearance of second site revertants (pCa37ma1rev, Figure 8A) could be explained by the activation of cryptic splice site(s). G-rich sequences close to splice acceptors have been shown to inhibit splicing in plants (Goodall and Filipowicz, 1989; Luehrsen and Walbot, 1994) and therefore the deletion of G-rich sequences in the second site revertants suggests activation of a cryptic splice acceptor site which is not used in the wild-type CaMV-S strain. Comparison with a naturally occurring deletion mutant of CaMV lacking most of ORF II (strain CM4-184; Howarth *et al.*, 1981) revealed that this cryptic splice acceptor is also conserved in this strain (Figure 8A). Transient expression constructs containing the rearranged ORF II sequences from Ca37ma1rev and CM4-184 (pIII–GUSma1rev and pIII–GUSCM4-184; Figure 8B) gave rise to ORF III expression levels comparable with that of pIII–GUS. RNase A/T1 mapping analysis confirmed that the secondary 3′ splice site is indeed used in pIII–GUSCM4-184. Activation of a cryptic splice site in the revertant and in the naturally occurring mutants supports the conclusion that this splicing event is a prerequisite for viral replication.

The splice donors and acceptors characterized here conform to consensus plant splice site sequences (Figure 3) and the introns have features typical of plant introns, although they are longer than average. Dicot introns are usually UA-rich and shorter than in vertebrates; two-thirds of known dicot introns are shorter than 150 nt and relatively few are longer than 2–3 kb. The C1, C2 and C3 introns are 266, 621 and 636 nt long, with an A+U content of 68, 71 and 71%. The C4 intron is 1623 nt in length with 65% A+U. Very few splice site mutants in plant introns have been analysed to date. Mutations of

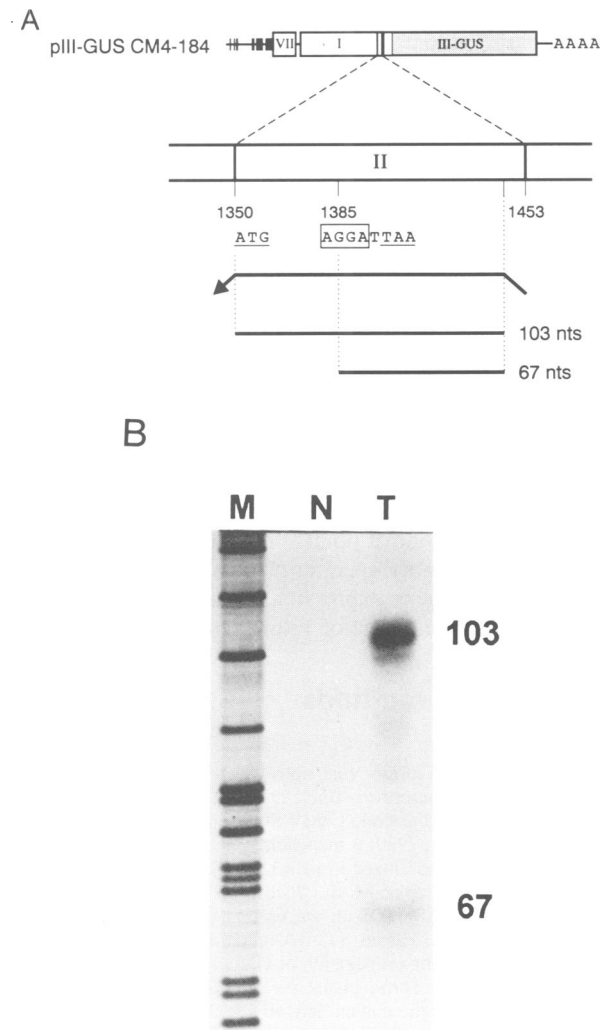


Fig. 9. (**A**) Schematic representation of the primary transcript produced from pIII–GUSCM4-184. The antisense RNA probe is represented by an arrow covering the region of sequence of strain CM4-184 between 1453 and 1350. The putative splice acceptor site is enclosed in a box. Thick lines show protected fragments corresponding to unspliced (103) or spliced (67) transcripts.(**B**) RNAs from *A.thaliana* protoplasts transfected with pIII–GUSCM4-184 (T) or non-transfected (N) were mapped. (M): ^{35}S-labelled G-ladder of pBluescript II.

consensus positions result in the activation of adjacent cryptic sites or decrease splicing efficiency (reviewed by Filipowicz *et al.*, 1995). One of our 3′ splice site mutations remained partially active (AGG→CGG in pIII–GUSma2, Figure7B), still allowing considerable levels of splicing both in transfected protoplasts and in pCa37ma2-infected plants. This was unexpected, because the central AG motif of the acceptor site is considered absolutely essential for splicing function (Moore *et al.*, 1993). The same AGG→CGG mutation has been observed to decrease splicing efficiency to 4% of wild-type in *Saccharomyces cerevisiae* (Parker and Siliciano, 1993). It could be that an optimal wider consensus compensates for the defect of the mutation in our case.

One of the splicing events creates a suitable mRNA for essential viral proteins; ORF III can be translated efficiently from the C4-type spliced RNA, and the ORF arrangement of this RNA suggests that ORF IV could also be produced from it under the influence of the translational transactivator. To date, expression of ORFs III and IV has been explained by the relay race model (Sieg and Gronenborn, 1982; Dixon and Hohn, 1984), according to which all viral ORFs are translated from the polycistronic 35S RNA. This model states that, after one of the ORFs has been translated, the translation process is re-initiated either by the same or another ribosome at the ORF that follows. This relay race translation is now known to depend on the action of TAV and has been shown to occur both with dicistronic viral transcripts (CaMV and FMV; Bonneville *et al.*, 1989; Gowda *et al.*, 1989) and with model dicistronic RNAs (Fütterer and Hohn, 1991; Zijlstra and Hohn, 1992). However, on polycistronic RNAs, it can be expected that re-initiation efficiency is reduced with each new ORF.

The product of ORF III is a minor capsid component (Giband *et al.*, 1986). Its function is still not clearly defined but it has been proposed to play a role in compaction of the viral genome after reverse transcription (Mougeot *et al.*, 1993), since it binds to DNA non-specifically (Giband *et al.*, 1986). ORF IV encodes the CaMV capsid protein. The amounts of these structural proteins, if exclusively translated from the 35S RNA, might not be sufficient for optimal virus production. We therefore suggest that the relay race mechanism acts on the spliced C4-type mRNA, allowing translation of ORF III and, in the presence of TAV, ORF IV.

The leader created by the C4 splicing event is 800 nt long and contains all the sORFs present in the original CaMV leader with the exception of sORF F (which is known to be inhibitory for translation; Fütterer *et al.*, 1990b). Instead of sORF F, a new sORF (sORF G) is created. Although longer, this leader is even less of a hindrance for translation than the original CaMV leader (compare pVII–GUS in Figure 1 and pIII–GUS-C4 in Figure 6), perhaps because, unlike sORF F, sORF G does not overlap the following large ORF. Translation from the C4 RNA could occur by internal initiation as in animal picornaviruses (e.g. Pelletier and Sonenberg, 1988; Jang and Wimmer, 1990), by leaky scanning as for the yeast GCN4 system (Hinnebusch, 1988) or via ribosome shunt as in the case of CaMV 35S RNA (Fütterer *et al.*, 1993) and possibly also of RTBV (Chen *et al.*, 1994). Experiments following the strategy described by Fütterer *et al.* (1993) on the 35S RNA leader will clarify this question. If the shunt mechanism is also used for the C4 RNA, as preliminary experiments indicate, then the comparison of features of the 35S and C4 leaders (as well as those of the other caulimo- and badnaviruses) will hopefully lead to a detailed understanding of the shunt mechanism.

With the CaMV strain used here [CaMV-S (Strasbourg)], the spliced RNA species are evidently excluded from the packaging/reverse transcription process; the corresponding cDNAs are absent from infected plants. This is easily explainable for the C4 RNA: the intron that is removed includes the primer binding site required for reverse transcription (see Figure 1). Interestingly, a corresponding CaMV deletion is not included in the set of CaMV deletion mutants found by Hirochika *et al.* (1985). For the C1–C3

introns one would have to postulate that they contain additional signals required for packaging and/or reverse transcription. The exclusion of spliced RNA from the reverse transcription process might not be as efficient for the latter cases and for some other CaMV strains, e.g. the isolate CaMV-S (Japan), where a DNA subpopulation obviously derived from spliced RNA has been observed (Hirochika *et al.*, 1985). Two of the five different deletions described by these authors correspond exactly to introns found by us (C1 and C2).

Splicing between ORF I and ORF II gave rise to fused ORFs comprising parts of ORF I and ORF II. A possible function for the ORF I–II fusion might simply be to maintain an arrangement of ORFs which is a suitable substrate for the action of TAV. Polypeptides corresponding in size to the fused ORFs were detected in infected plants (Z.Kiss-László, unpublished observation). So far we do not know their role, if any. They would be not essential for virus replication and systemic spread, since mechanically inoculated virus genomes, in which ORF II is missing or replaced by a foreign coding region, lead to systemic infections. Also their obligate roles in insect transmission is questionable, since Blanc *et al.* (1993) showed that the ORF II protein alone is sufficient for aphid transmission: when produced from baculovirus vectors in insect cells and first acquired by aphids through parafilm membranes, it mediates the transmission of aphid non-transmissible CaMV isolates CM1841 and CM4-184.

Another possible function for the fusion proteins could be a role in plants growing under specific conditions. In this respect, it is interesting to note the experiments of Qiu and Schoelz (1993 and personal communication), who analysed the requirements of CaMV strains to infect *Solanaceae*. They found that CaMV strain W260 and chimaeric viruses, with ORF II, IV, VI and part of ORF I derived from W260, can grow in *Nicotiana bigelovii*. The requirement for ORF II in this host was conditional. At low light intensity the source of ORF II did not influence whether a chimaeric virus infected *N.bigelovii*. At high light intensity, however, ORF II had to be derived from W260. They mapped this property to a single codon within ORF II, which would be present in the fusion proteins that we detected.

The splicing events characterized here would be predicted to reduce the amounts of movement protein (pI) and aphid transmission factor (pII). Thus, regulation of splicing could be a mechanism to adjust the level of production of these proteins during viral infection. Splicing could also be regulated to control either the temporal expression or quantitative production of the products of ORFs III and IV. The contribution of splice sites detected in ORF I suggests the possibility that these events might be controlled by regulating competition between alternative splice donors.

The discovery of a spliced mRNA for ORF III of CaMV has led to the modification of the relay race translation model. Although splicing has previously been thought to play no role in CaMV gene expression, the results presented here show that it is required to generate a greater variety of mRNA substrates for the specialized CaMV translational control mechanisms, thus ensuring appropriate production of viral gene products. Because our search for splicing signals was directed specifically towards a limited region of the genome, it remains an open question whether other spliced CaMV RNAs exist which might be required for constitutive or transactivated translation of other viral ORFs. For example, expression of the enzymatic functions of ORF V could be explained by a relay race-type model acting on the C4-type mRNA described here, but it remains possible that another, as yet uncharacterized, mRNA is produced for translation of this ORF (Plant *et al.*, 1985; Schultze *et al.*, 1990).

A spliced RNA has already been suggested to play a role in virus gene expression in RTBV (Fütterer *et al.*, 1994), and there is evidence that splicing is important also in duck hepatitis B virus gene expression (H.Schaller and S.Obert, personal communication). These findings break the dogma that pararetroviruses use only pregenomic RNAs or RNAs directed from internal promoters for translation of viral products. Gene expression can be regulated at various levels, and all these processes provide further potential control points. The need to elucidate the mechanisms of highly specific quantitative, temporal and spatial control of gene expression in plant pararetroviruses will be a fascinating area of future research.

Materials and methods

Plants and viruses

Brassica rapa plants (Turnip 'Just right') were grown at 20°C with a 16 h photo period as described (Blanc *et al.*, 1993). The CaMV isolate used was derived from the strain CaMV Strasbourg (Franck *et al.*, 1980). pCa37 (Lebeurier *et al.*, 1980) is the whole genome of CaMV Strasbourg linearized with *Sal*I and cloned in pBR322. pCa37ma1 and pCa37ma2 were constructed by replacing the *Bst*EII–*Bam*HI fragment of pIII–GUSma1 and pIII–GUSma2 with the original *Bam*HI–*Kpn*I of pCa37 (see below). pCa37ma3 carries a G→A mutation in ORF II at position 1509, which arose during propagation of viral clones in *Escherichia coli* (S.Blanc, unpublished). Turnip plants were inoculated mechanically with either pCa37 or one of the mutant derivatives. Generally 20 plants were used for each case, however 60 plants were used in the case of Ca37ma1 and Ca37ma3.

Plasmids for transient expression in plant protoplasts

Plasmids pVII–GUS, pI–GUS, pII–GUS (called pIIb–GUS here), pIII–GUS and pIV–CAT have been described previously (Fütterer *et al.*, 1990b). In pIIa–CAT, the CAT gene is fused to the first codon of ORF II; pIIb–CAT is analogous to pIIb–GUS with the CAT ORF fused to codon 110 of ORF II. The frameshift mutant pIIb–GUSfs was generated from pIIb–GUS by filling in the *Bam*HI site at the border between ORF II and the GUS gene.

To generate mutations in the 3′ splice site, the *Bst*EII–*Sph*I fragment from pIII–GUS was first subcloned into a modified pBluescript II (KS+) containing *Bst*EII–*Sph*I sites introduced on an oligonucleotide insertion into the polylinker. The oligonucleotides 5′-GTAAGCTTCTTCGGACAATCTCAT-3′ and 5′-GTAAGCTTCTTCCGACAATCTCAT-3′ (mutated bases underlined) were used to generate ma1 and ma2 respectively by *in vitro* oligonucleotide-directed site-specific mutagenesis (Amersham oligonucleotide-directed *in vitro* mutagenesis system version II). The mutated region was confirmed by sequencing and the corresponding fragments were replaced in pIII–GUS.

pIII–GUSmd containing a mutation in the 5′ splice site in the leader was constructed by replacing an *Eco*NI–*Cla*I fragment of pIII–GUS with oligonucleotides carrying the desired nucleotide exchanges. Sense oligonucleotide: 5′-GCAACCAAAGACGATGGAAATTTGA TAGAC CAACGTTACTATACTTATACTATACGCTAAGGGAATGCTTGTATTTACCCTATATACCCTAATGACCCCTTAT-3′; antisense oligonucleotide: 5′-CGATAAGGGGTCATTAGGGTATATAGGGTAAATACAAGCATTCCCTTAGCGTATAGTATAAGTATAGTAACGTTGGTCT ATCAAATTTCCATCGTCTTTGGTTG-3′.

The construct lacking the intron between the leader and gene II, pIII–GUS-C4, was constructed by replacement of an *Eco*NI–*Hin*dIII fragment of pIII–GUS with oligonucleotides to recreate the splicing junction cDNA sequence. Sense oligonucleotide: 5′-GCAAGGTAAGACG-

ATGGAAATTTGATAGAGGAAGA-3'; antisense oligonucleotide: 5'-AGCTTCTTCCTCTATCAAATTTCCATCGTCTT ACCTT G -3'.

In pIV–CAT-C4, the *Cla*I–*Sph*I fragment covering ORF III and the CAT gene from pIV–CAT was introduced into the corresponding restriction sites of pIII–GUS-C4.

To reconstruct pIII–GUSma1rev, Ca37ma1rev-specific virus DNA was amplified by oligonucleotides covering the first 18 nt of ORF I (sense 5'-ATGGATTTGTATCCAGAAGA-3') and a *Kpn*I site in ORF III (antisense 5'-GTTGGGTACCTAAGGCTTCTAATATCTC-3') and sequenced. We used this amplified fragment digested with *Nsi*I–*Bam*HI and exchanged the corresponding region in pIII–GUS. pIII–GUSCM4-184 was constructed with the same *Nsi*I–*Bam*HI fragment exchange using the original CM4-184 strain.

Plant cell cultures and protoplast transformation

Culture conditions for *O.violaceus* suspension cultures, protoplast preparation and transfection have been described previously (Fütterer *et al.*, 1989, 1990a). Routinely, 2×10^6 protoplasts were transfected by electroporation (Fromm *et al.*, 1985) with 10–15 μg of circular plasmid DNA. Alternatively, 6×10^5 protoplasts made from *A.thaliana* cell suspension culture (Trezzini *et al.*, 1993) were transfected using polyethylene glycol (PEG)-mediated transformation (Goodall *et al.*, 1990). After overnight incubation (12–18 h), a protein extract was prepared and reporter gene activities were determined (Bonneville *et al.*, 1989).

Isolation of RNA and analysis of transcripts by RNase A/T1 mapping

The guanidium thiocyanate/phenol/chloroform extraction method (Goodall *et al.*, 1990) was used for isolation of RNA from either infected turnip leaves or transfected *A.thaliana* protoplasts.

RNase A/T1 mapping was carried out as described by Goodall *et al.* (1990). Protected fragments were separated on 6% denaturing polyacrylamide gels and quantified by means of a Phosphor Imager (Molecular Dynamics).

Specific antisense RNA probes were synthesized *in vitro* using T3 or T7 RNA polymerase, [α-^{32}P]UTP (sp. act. 30–40 Ci/mmol), and linearized recombinant plasmid as template. The probes were purified on polyacrylamide gels.

pII was derived from ORF II. An *Eco*RI–*Sma*I fragment encompassing 222 nt of ORF II was generated by PCR using primers homologous to regions 1440–1460 and 1646–1670 of pIIb–GUS and cloned into pBstII(KS+). The identity of the clone was confirmed by sequencing. The plasmid was linearized with *Eco*RI and transcribed with T7 polymerase to produce antisense RNA probe. Sense primer: 5'-GCAGAATTCGTTCCTCAAAAGGGA-3'; antisense primer: 5'-GACCTACCCGGGGATCCTCTAGAGT-3'.

A probe for mapping RNAs produced after transfection of pIII–GUS CM4-184 was generated by the same approach as described above. Primers complementary to regions 1350–1371 and 1425–1453 of pIII–GUS CM4-184 were designed containing *Not*I and *Pst*I sites respectively. Sense primer: 5'-TGATGCGGCCGCATGAGCATTACGGGTCAACCG-3'; antisense primer: 5'-GACTCTGCAGGATTTCAGAGACTTCTTTCTGGA-3'. The plasmid was linearized with *Not*I and transcribed with T3 RNA polymerase to produce antisense RNA probe.

Production and characterization of the cDNAs

Total RNA isolated from *A.thaliana* protoplasts transfected with pIIb–GUS was used for first strand cDNA synthesis. The 20 μl reaction contained 50 mM Tris–HCl pH 8.1, 60 mM $MgCl_2$, 0.4 M KCl, 10 mM DTT, 1 mM dNTPs, 100 pM [N_6] random oligonucleotides, 10 U of AMV reverse transcriptase (Boehringer), 10 U of RNasin (Promega) and 20 μg of total RNA. The reaction was incubated at 42°C for 1 h. The 5' phosphorylated 20mer oligonucleotide 5'-pCATCTCGAGCGGCCGCATCA-3' was ligated to the 3' end of the first strand cDNA using T4 RNA ligase. (Edwards *et al.*, 1991). After phenol–chloroform treatment, the resulting ligation product was used as a template in a PCR carried out as described (Frohman *et al.*, 1988). The primers were a 34mer oligonucleotide (5'-GACTCTGCAGGGTCTTTGCTTAATCCGAAG-3') comprising the CaMV sequence at positions 1546–1567 and a 20mer oligonucleotide (5'-TGATGCGGCCGCTCGAGATG-3') complementary to the one which was ligated to the 3' end of the cDNA. Amplified fragments were cloned using the PCR-introduced restriction sites into the *Not*I–*Pst*I sites of pBst(KS+). The recombinant plasmids were selected by colony hybridization (Maniatis *et al.*, 1989) using an oligonucleotide downstream of the putative 3' splice site at positions 1524–1545 (5'-GAATATGGAAGATCAACTCAT-3'). Selected clones were further screened with a second oligonucleotide corresponding to a region 1481–1504 (5'-CATCTTAACAACCTCAATGAGATT-3') upstream of the 3' splice site (1508). Colonies which were positive in the first round of screening and negative in the second round were selected for further analysis.

Hybridization was carried out overnight at 37°C in a solution containing 6× SSC, 1× Denhardts, 0.05% Na pyrophosphate and 100 μg/ml yeast tRNA. Filters were washed twice in 6× SSC for 30 min, at 55 and 60°C respectively. The sequence of 50 selected clones was analysed by the dideoxy sequencing method (Sanger *et al.*, 1977) and a Sequenase 2.0™ Kit (USB).

One individual clone from each group of cDNAs covering the four different splicing junctions was used for RNase A/T1 mapping. The plasmids were linearized with *Not*I and transcribed with T3 RNA polymerase to produce antisense RNA probes.

Analysis of CaMV DNA from plants infected with pCa37, pCa37ma1 and pCa37ma2

Virus DNA was extracted from young systemically infected plant leaves (Gardner and Shepherd, 1980) 3 weeks after inoculation and amplified by PCR as described by Berthomieu and Meyer (1991). To characterize C1–C3 type-specific DNAs, a sense primer was designed at position 7854–7866 of the leader while the antisense oligonucleotide was the same as described above at position 1546–1567 within ORFII. pCa37ma1 and pCa37ma2 specific virus DNA was amplified and sequenced between ORFI and ORFIII.

Acknowledgements

We thank Stefan Karlsson for some of the reporter constructs used, Christine Perbal, Simon Covey and Roger Hull for antibodies, Imre Somssics for kindly providing an *Arabidopsis* cell culture, Heinz Schaller and James Schoelz for personal communication. We very much appreciate the technical assistance of Mathias Müller, Sandra Corsten and Diana Lewetag. We thank all our colleagues for their interest and various types of valuable advice and/or critical reading of the manuscript, especially Helen Rothnie, Waltraud Schmidt-Puchta, Johannes Fütterer and Gordon Simpson. Special thanks go to Mike Rothnie for preparation of the figures.

References

Berthomieu,P. and Meyer,C. (1991) Direct amplification of plant genomic DNA from leaf and root pieces using PCR. *Plant. Mol. Biol.*, **17**, 555–557.

Blanc,S., Cerutti,M., Usmany,M., Vlak,J.M. and Hull,R. (1993) Biological activity of cauliflower mosaic virus aphid transmission factor expressed in a heterologous system. *Virology*, **192**, 643–650.

Bonneville,J.M., Sanfaçon,H., Fütterer,J. and Hohn,T. (1989) Post-transcriptional *trans*-activation in cauliflower mosaic virus. *Cell*, **59**, 1135–1143.

Chen,G., Müller,M., Potrykus,I., Hohn,T. and Fütterer,J. (1994) Rice tungro bacilliform virus: transcription and translation in protoplasts. *Virology*, **204**, 91–100.

Chen,P., Chen,C., Sung,J. and Chen,D. (1989) Identification of a doubly spliced viral transcript joining the separated domains for putative protease and reverse transcriptase of hepatitis B virus. *J. Virol.*, **63**, 4165–4171.

Covey,S.N. and Hull,R. (1981) Transcription of cauliflower mosaic virus DNA. Detection of transcripts, properties, and location of the gene encoding the virus inclusion body protein. *Virology*, **111**, 463–474.

DeTapia,M., Himmelbach,A. and Hohn,T. (1993) Molecular dissection of the cauliflower mosaic virus translational transactivator. *EMBO J.*, **12**, 3305–3314.

Dixon,L.K. and Hohn,T. (1984) Initiation of translation of the cauliflower mosaic virus genome from a polycistronic mRNA: evidence from deletion mutagenesis. *EMBO J.*, **3**, 2731–2736.

Driesen,M., Benito-Moreno,R., Hohn,T. and Fütterer,J. (1993) Transcription from the CaMV 19S promoter and autocatalysis of translation from CaMV RNA. *Virology*, **195**, 203–210.

Edwards,J.B.D.M., Delort,J. and Mallet,J. (1991) Oligodeoxyribonucleotide ligation to single-stranded cDNAs: a new tool for cloning 5' ends of mRNAs and for constructing cDNA libraries by *in vitro* amplification. *Nucleic Acids Res.*, **19**, 5227–5232.

Filipowicz,W., Gniadkowski,M., Klahre,U. and Liu,H.-X. (1995) Pre-mRNA splicing in plants. In Lamond,A. (ed.), *Pre-mRNA Processing*. RG. Landes Publishers, Georgetown, TX, pp 65–77.

Franck,A., Guilley,H., Jonard,G., Richards,K.E. and Hirth,L. (1980)

Nucleotide sequence of cauliflower mosaic virus DNA. *Cell,* **21**, 285–294.

Frohman,M.A., Dush,M.K. and Martin,G.R. (1988) Rapid production of full length cDNAs from rare transcripts: amplification using a single gene specific oligonucleotide primer. *Proc. Natl Acad. Sci. USA,* **85**, 8998–9002.

Fromm,M., Taylor,L.P. and Walbot,V. (1985) Expression of genes transferred into monocot and dicot plant cells by electroporation. *Proc. Natl Acad. Sci. USA,* **82**, 5824–5828.

Fütterer,J. and Hohn,T. (1991) Translation of polycistronic mRNA in the presence of the cauliflower mosaic virus transactivator protein. *EMBO J.,* **10**, 3887–3896.

Fütterer,J. and Hohn,T. (1992) Role of an upstream open reading frame in the translation of polycistronic mRNAs in plant cells. *Nucleic Acids Res.,* **20**, 3851–3857.

Fütterer,J., Gordon,K., Pfeiffer,P., Sanfaçon,H., Pisan,B., Bonneville,J. and Hohn,T. (1989) Differential inhibition of downstream gene expression by the cauliflower mosaic virus 35S RNA leader. *Virus Genes,* **3**, 45–55.

Fütterer,J., Bonneville,J., Gordon,K., DeTapia,M., Karlsson,S. and Hohn,T. (1990a) Expression from polycistronic cauliflower mosaic virus pregenomic RNA. In Mc Carthy,J.E.G. and Tuite,M.F. (eds), *Posttranscriptional Control of Gene Expression.* Nato Asi Series, Springer, Berlin, pp. 349–357.

Fütterer,J., Gordon,K., Bonneville,J., Sanfacon,H. and Hohn,T. (1990b) Positive and negative control of translation by the leader sequence of cauliflower mosaic virus pregenomic 35S RNA. *EMBO J.,* **9**, 1697–1707.

Fütterer,J., Kiss-László,Z. and Hohn,T. (1993) Nonlinear ribosome migration on cauliflower mosaic virus 35S RNA. *Cell,* **73**, 789–802.

Fütterer,J., Potrykus,I., Valles-Brau,M.P., Dasgupta,I., Hull,R. and Hohn,T. (1994) Splicing in a plant pararetrovirus. *Virology,* **198**, 663–670.

Gardner,R.C. and Shepherd,R.J. (1980) A procedure for rapid isolation and analysis of cauliflower mosaic virus DNA. *Virology,* **106**, 159–161.

Giband,M., Mesnard,J.M. and Lebeurier,G. (1986) The gene III product (p15) of cauliflower mosaic virus is a DNA-binding protein while an immunologically related p11 polypeptide is associated with virions. *EMBO J.,* **5**, 2433–2438.

Goodall,G.J. and Filipowicz,W. (1989) The AU-rich sequences present in the introns of plant nuclear pre-mRNAs are required for splicing. *Cell,* **58**, 473–483.

Goodall,G.J., Wiebauer,K. and Filipowicz,W. (1990) Analysis of pre-mRNA processing in transfected plant protoplasts. *Methods Enzymol.,* **181**, 148–161.

Gowda,S., Wu,F.C., Scholthof,H.B. and Shepherd,R.J. (1989) Gene VI of figwort mosaic virus (caulimovirus group) functions in posttranscriptional expression of genes on the full-length RNA transcript. *Proc. Natl Acad. Sci. USA,* **86**, 9203–9207.

Guilley,H., Dudley,R.K., Jonard,G., Balázs,E. and Richards,K.E. (1982) Transcription of cauliflower mosaic virus DNA: detection of promoter sequences and characterization of transcripts. *Cell,* **30**, 763–773.

Hinnebusch,A.G. (1988) Novel mechanisms of translational control in *Saccharomyces cerevisiae. Trends Genet.,* **4**, 169–174.

Hirochika,H., Takatsuji,H., Ubasawa,A. and Ikeda,J. (1985) Site-specific deletion in cauliflower mosaic virus DNA: possible involvement of RNA splicing and reverse transcription. *EMBO J.,* **4**, 1673–1680.

Hohn,T. and Fütterer,J. (1992) Transcriptional and translational control of gene expression in cauliflower mosaic virus. *Curr. Opin. Genet. Dev.,* **2**, 90–96.

Howarth,A.J., Gardner,R.C., Messing,J. and Shepherd,R.J. (1981) Nucleotide sequence of naturally occurring deletion mutants of cauliflower mosaic virus. *Virology,* **112**, 678–685.

Hull,R. and Covey,S.N. (1983) Does cauliflower mosaic virus replicate by reverse transcription?. *Trends Biochem. Sci.,* **8**, 119–121.

Jang,S.K. and Wimmer,E. (1990) Cap-independent translation of encephalomyocarditis virus RNA: structural elements of the internal ribosome entry site and essential binding of a cellular 57 kDa protein. *Genes Dev.,* **4**, 1560–1572.

Kozak,M. (1989) The scanning model for translation—an update. *J. Cell Biol.,* **108**, 229–241.

Lebeurier,G., Hirth,L., Hohn,T. and Hohn,B. (1980) Infectivities of native and cloned cauliflower mosaic virus DNA. *Gene,* **12**, 139–146.

Luehrsen,K.R. and Walbot,V. (1994) Addition of A- and U-rich sequence increases the splicing efficiency of a deleted form of maize intron. *Plant Mol. Biol.,* **24**, 449–463.

Maniatis,T., Fritsch,E.F. and Sambrook,J. (1989) *Molecular Cloning. A Laboratory Manual.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Moore,M.J., Query,C.C. and Sharp,P.A. (1993) Splicing of precursors to mRNA by the spliceosome. In Gesteland,R.F. and Atkins,J.F., (eds), *The RNA World.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 303–357.

Mougeot,J., Guidasci,T., Wurch,T., Lebeurier,G. and Mesnard,J. (1993) Identification of C-terminal amino acid residues of cauliflower mosaic virus open reading frame III protein responsible for its DNA binding activity. *Proc. Natl Acad. Sci. USA,* **90**, 1470–1473.

Parker,R. and Siliciano,P.G. (1993) Evidence for an essential non-Watson–Crick interaction between the first and last nucleotides of a nuclear pre-mRNA intron. *Nature,* **361**, 660–662.

Pelletier,J. and Sonenberg,N. (1988) Internal initiation of translation of eukaryotic mRNA directed by a sequence derived from poliovirus RNA. *Nature,* **334**, 320–325.

Plant,A.L., Covey,S.N. and Grierson,D. (1985) Detection of a subgenomic mRNA for gene V, the putative reverse transcriptase gene of cauliflower mosaic virus. *Nucleic Acids Res.,* **13**, 8305–8321.

Qiu,S.G. and Schoelz,J.E. (1993) A point mutation within CaMV gene II influences host specificity in a light intensity-dependent manner. Abstr. No. P65–29 of The IX International Congress on Virology, 347 (Abstract).

Restrepo,M.A., Freed,D.D. and Carrington,J.C. (1990) Nuclear transport of plant potyviral proteins. *Plant Cell,* **2**, 987–998.

Rothnie,H.M., Chapdelaine,Y. and Hohn,T. (1994) Pararetroviruses and retroviruses: a comparative review of viral structure and gene expression strategies. *Adv. Virus Res.,* **44**, 1–67.

Sanger,F., Nicklen,S. and Coulson,A.R. (1977) DNA sequencing with chain-terminating inhibitors. *Proc. Natl Acad. Sci. USA,* **74**, 5463–5467.

Scholthof,H.B., Wu,F.C., Richins,R.D. and Shepherd,R.J. (1991) A naturally occuring deletion mutant of figwort mosaic virus (caulimovirus) is generated by RNA splicing. *Virology,* **184**, 290–298.

Scholthof,H.B., Gowda,S., Wu,F.C. and Shepherd,R.J. (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.

Schultze,M., Hohn,T. and Jiricny,J. (1990) The reverse transcriptase gene of cauliflower mosaic virus is translated separately from the capsid gene. *EMBO J.,* **9**, 1177–1185.

Sieg,K. and Gronenborn,B. (1982) Evidence for polycistronic messenger RNA encoded by cauliflower mosaic virus. NATO Advanced Studies Institute Advanced Course 154 (Abstract).

Trezzini,G.F., Horrichs,A. and Somssich,I.E. (1993) Isolation of putative defense-related genes from *Arabidopsis thaliana* and expression in fungial elicitor treated cells. *Plant Mol. Biol.,* **21**, 385–389.

Vaden,V.R. and Melcher,U. (1990) Recombination sites in cauliflower mosaic virus DNAs: implications for mechanisms of recombination. *Virology,* **177**, 717–726.

Varagona,M.J., Schmidt,R.J. and Raikhel,N.V. (1992) Nuclear localization signals required for nuclear targeting of the maize regulatory protein opaque-2. *Plant Cell,* **4**, 1213–1227.

Zijlstra,C. and Hohn,T. (1992) Cauliflower mosaic virus gene VI controls translation from dicistronic expression units in transgenic *Arabidopsis* plants. *Plant Cell,* **4**, 1471–1484.

Received on March 13, 1995; revised on April 25, 1995