Symbiosis, 2 (1986)
Balaban Publishers, Philadelphia/Rehovot

Cauliflower Mosaic Virus Potential Vector Gene: Genome Organization and Replication

GENEVIÉVE LEBEURIER

Laboratoire de Virologie, Institut de Biologie Moléculaire et Cellulaire, CNRS, 67000 Strasbourg, France
Tel. 88.61.02.02 Telex 880492F

Abstract

Cauliflower mosaic virus (CaMV) is one of the rare DNA plant viruses. Its DNA is in a double-stranded form with three interruptions. The complete sequence of the viral DNA has been established. The transcript of the α-strand contains six or eight open reading frames (ORFs) coding for viral proteins. The presence of five of these proteins has been immunodetected in infected plants and their functions determined or predicted. Only ORF VI encodes for a 19 S RNA which synthesizes *in vivo* and *in vitro* the inclusion body protein. ORF V encodes for a reverse-transcriptase which synthesizes single-stranded DNA from a 35 S RNA according to a process very close to that of Hepatitis B virus and Retroviruses. The usefulness of CaMV as a gene vector in higher plants is discussed.

The molecular biology of Caulimoviruses has already inspired quite a few review articles: I shall therefore limit myself here to a general description of the type-member of this group, namely cauliflower mosaic virus (CaMV), and detail the major results obtained both on the organization and the function of the molecular level of Caulimoviruses and especially cauliflower mosaic virus.

Caulimoviruses are, together with Geminiviruses, the only plant viruses with a DNA genome. DNA viruses represent at most 1 to 2% of the viruses infecting plants, all others being RNA viruses.

The caulimovirus group comprises 12 related members whose structure is very similar to that of CaMV. Consequently, most of the properties of CaMV can be extrapolated to the other members of this group.

CaMV is the type-member of Caulimoviruses. This family is represented by 12 groups. Cauliflower mosaic virus or CaMV is the best known of this family.

Keywords: CaMV — vector gene — genome organization — replication — *in vivo* recombination

0334-5114/86/$03.00 © 1986 Balaban Publishers

1. Structure of the Viral Genome

Double-stranded DNA extracted from purified virus preparation is chiefly found as relaxed circular molecules together with some linear molecules of variable length.

The circular molecules are relaxed owing to the presence of single stranded interruptions (Volovitch et al., 1978): one strand, called the negative strand or α strand, features invariably a single interruption: $\Delta 1$, while the other strand, "positive strand" or β strand, has two interruptions: $\Delta 2$ and $\Delta 3$ (Fig. 1).

The complete sequence of the DNA of the Strasbourg strain (S-CaMV) has been established in 1980 by a group of our laboratory (Franck et al., 1980). So a very peculiar triple-stranded structure of these interruptions has been described (Richards et al., 1981) thus for the α strand the 3'OH end overlaps by a few nucleotides an identical sequence at the 5' end; the β strand is continued in this region corresponding at the structure of $\Delta 1$. $\Delta 2$ and $\Delta 3$ have the same organization with the 3'OH end of the β strand overlapping an identical sequence of the 5' end but slightly longer than the $\Delta 1$ sequence. It is the same situation for $\Delta 3$ interruption (see Fig. 1). Now the generation of these interruptions is understood, it is the result of the mode of replication of virus; we will explain this process later. Their existence is not required for virus infectivity as we and another group have demonstrated (Howell et al., 1980; Hull and Howell, 1978; Hull and Covey, 1983; Kruse et al., 1986; Lebeurier et al., 1980).

The S-CaMV DNA cloned in pBR322 has lost its interruptions after amplification of plasmid in bacteria. Such a cloned S-CaMV DNA, linearized by enzyme which cuts at a restriction site used to integrate viral genome in pBR322, is always infection.

Upon infection the relaxed circular viral DNA or the linear viral DNA migrate to the nucleus of the infected cells. Both viral DNAs are converted into a covalently closed circular form found associated with nucleosomes to form a viral minichromosome (Ménissier et al., 1983; 1984; 1986; Olszewski et al., 1982). This minichromosome has the same structure as that of the chromatine of sensitive cells; its length corresponds to 42 nucleosomes (Ménissier et al., 1983). The minichromosome does not get encapsidated into virions.

2. Transcription Process

The transcription process is localized in the nucleus of the infected cells. The viral minichromosome is the template from which the host RNA polymerase II generates the virus-specific transcripts.

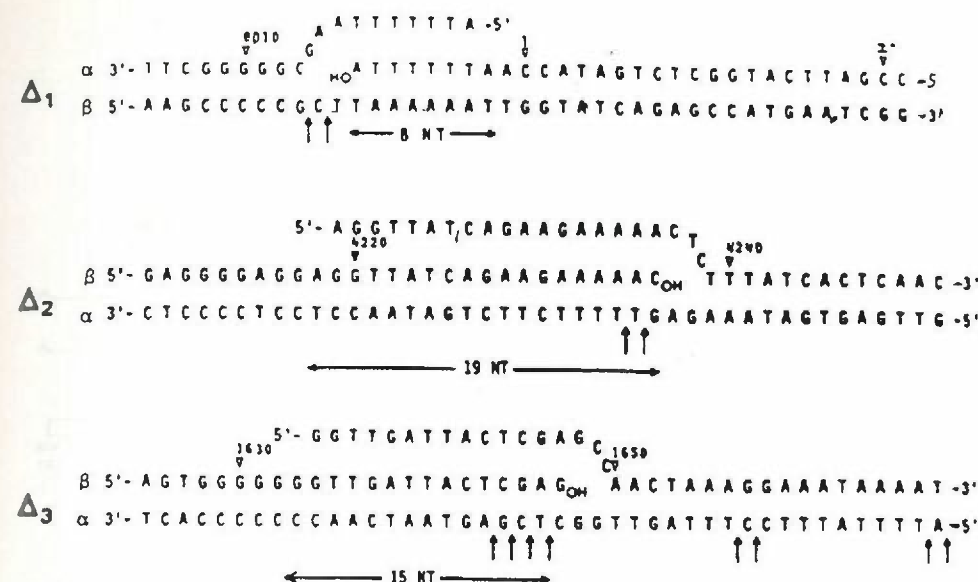

Figure 1. Sequence around the discontinuities of CaMV Cabb-S DNA. The location of the 5'-extremity of the discontinuities and the sequence of the complementary strand is from Franck et al. (1980). Arrows indicate the positions of other 3'-termini deleted (Richards et al., 1981).

Four discrete species of capped and polyadenylated viral RNA could be detected (Condit and Maegher, 19883; Covery and Hull, 1981; Dudley et al., 1982; Guilley et al., 1982; Howell and Hull, 1978).

- A major 35S transcript corresponding to the transcription of the entire viral DNA plus a terminal repeat of 180 nucleotides. It initiates at nucleotide 7435 and terminates at nucleotide 7615.
- A minor 35S RNA starts at nucleotide 8017, but contains no terminal repeat.
- A subgenomic messenger called 19S RNA starts at nucleotide 5764 and terminates like the major 35S RNA at nucleotide 7615: thee two species (major 35S and 19S) terminate 18 nucleotides downstream of the polyadenylation signal sequence AATAAA.
- A minor 8S RNA is always detected in infected cells. Its role is still unknown, but it could correspond to an aborted transcription of the major 35S RNA started on a template possessing the $\Delta 1$ interruption (nucleotide position 8024) where it would stop (see Fig. 2).

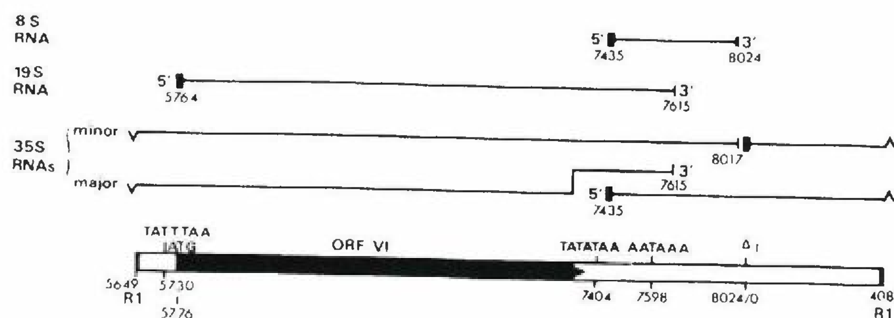

Figure 2. Genetic map of the CaMV DNA with the position of the four known polyadenylated viral transcripts. Stippled circle represents the DNA with position of Δ1 as indicated. Major coding regions are represented by open arrows inside the circle and transcripts by arrows outside the circle (according to Richards et al., unpublished results).

The transcription of the 35S major and the 19S RNAs are apparently under the control of classical TATA boxes located 36 nucleotides upstream of the cap site of these two RNAs, and their promoter sequences are very similar to those already described in other eukaryotes.

3. Translation of the Viral Proteins

Attempts to translate *in vitro* these different kinds of mRNAs failed except for the 19S RNA which is generated by the direct transcription of ORF VI and codes for the protein that constitutes the matrix of the inclusion bodies.

One of the major issues of the molecular biology of CaMV remains the mode of expression of the viral genes: a new hypothesis suggests that for example ORFs IV and V could possibly be expressed as a long precursor subsequently cleaved to generate functional proteins.

In fact transcription is fully asymmetric with only the α strand transcribed to generate the major 35S RNA. Although we do not exactly understand the translation process, we suppose that the full 35S RNAs migrate to the cytoplasm of infected cells to carry out this translation. Such a 35S RNA contains the 6 to 8 potential open reading frames (ORFs) corresponding to 6 major proteins and may be two minor ones (Fig. 3).

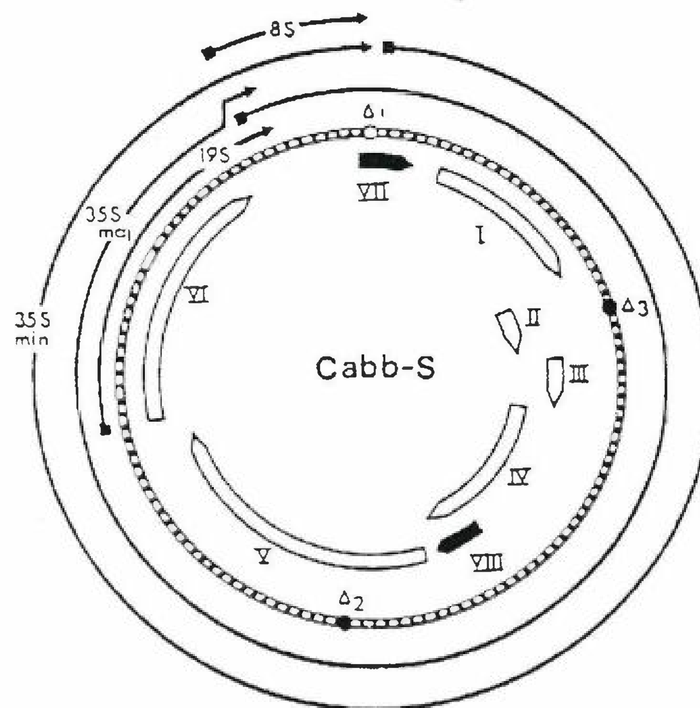

Figure 3. Schematic representation of the various transcripts isolated from cauliflower mosaic virus infected plants. The numbers indicate the start points (5'OH) and the end points (3'OH) of the three main transcripts. The lower scheme indicates the position of the TATA boxes for the 19 S (nucleotide 5730), the 8 S RNA and the 35 S major RNA (nucleotide 7404). The polyadenylated signal of the transcript is indicated as AATAA. No TATA box has been detected for the minor 35 S RNA. According to Guilley et al. (1982) the 8 S RNA starts at the same point as the major 35 S RNA but stops at the Δ1 interruption (see text for explanation).

The products of genes VII, I and VIII have not yet been identified, though gene I is known to be required for virus replication.

By contrast, gene II is dispensable since CM4-184 (Howarth et al., 1981), a natural isolate of CaMV, which bears a deletion in gene II, is perfectly viable, though no longer transmissible by aphids (Hull and Howell, 1978). Moreover aphid transmissibility requires the formation of a functional complex between the product of gene II and the major viroplasm protein (Givord et al., 1984).

The protein corresponding to the expression of gene III is detected in inclusion bodies (Xiong et al., 1984) and at the surface of virus particles by immunoelectromicroscopy using in both cases antibodies directed against synthetic peptide spanning the N-terminal part of gene III. The exact role of this protein is still unknown but is seems to be able to react with the virion subunits and to be a "DNA binding protein" (Giband et al., 1986).

Gene IV is known to correspond to the coat protein of the virus (Franck et al., 1980) ; it is obviously expressed although it is found in the virions in a processed form derived from a 57 Kd precursor. Proteolytic cleavage converts this precursor into the two protein species found in the mature virions: the 42 Kd protein could be used as a scaffolding protein around which the 37 Kd species would organize (Kruse et al., 1986).

Recently, CaMV virions were shown to contain a kinase activity which is able to phosphorylate the serine and threonine residues of various proteins. In the case of CaMV, the 57 and 42 Kd proteins could be phosphorylated, but not the 37 Kd species (Ménissier-de Murcia et al., 1986). The role of this kinase activity is still unknown, but it could possibly be related to the maturation of the coat protein precursor in the virion.

Antibodies raised against synthetic peptide spanning the C-terminal part of gene V have permitted the detection of this gene product in inclusion bodies (Ziegler et al., 1985). On the other hand the sequence homologies detected between ORF V and the "pol" gene of retroviruses such as Moloney murine leukemia virus, Human T cell leukemia virus, Rous sarcoma virus (Toh et al., 1983), strongly suggest that gene V corresponds to the reverse-transcriptase used by CaMV to amplify viral DNA during the replication cycle.

The final open reading frame VI encodes the 19S subgenomic RNA. This mRNA directs *in vivo* and *in vitro* the synthesis of the major protein of the inclusion bodies also called "viroplasms" (Xiong et al., 1982). These viroplasms are characteristic of caulimoviruses; they have a proteinaceous structure; they are found in the cytoplasm of infected cells and contain the virus particles.

4. Replication of the Viral Genome

Although we still do not understand the mode of expression of the viral genes, a *viral replication strategy*, much reminiscent of that of the retroviruses, has been proposed in 1983 by a group of our laboratory and was confirmed by others (e.g., Guilley et al., 1983; Hull and Cover, 1983; Kruse et al., 1986; Ménissier et al., 1983, 1984, 1986; Pfeiffer and Hohn, 1983). The current model is explained in Fig. 4. Synthesis of the negative DNA strand is always initiated with a host transfer RNA (t-RNA^{Met}) primer, where the 3' end is hydrogen-bonded to a 14 base sequence near the 5'end of the 35S RNA. When the template is exhausted about 600 bases from the initiation site, the first jump occurs to form a bridge between the new DNA segment R' and the R-RNA segment at the 3' end of one of the 35S RNA. Once safely repositioned, the negative DNA nascent strand is extended apparently continuously over the major expanse of the 35S RNA molecule upto the 5' end sequence of the t-RNA^{Met}, which is complementary of its 3' end sequence. During the extension of the negative DNA strand, the RNase H enzyme digests the 35S RNA starting near the complex formed between t-RNA^{Met} and the complementary sequence localized on the 35S RNA. Two GC rich segments are not digested by RNase H and are used as "primer" by the enzyme to begin the extension of the positive nascent DNA strand. These 2 GC segments define the position of the interruptions $\Delta 2$ and $\Delta 3$ while the interruption $\Delta 1$ corresponds to the position of the complex between the t-RNA^{Met} and the complementary sequence on the template 35S RNA. The viral double-stranded DNA nascent molecules are encapsidated by protein subunits present in viroplasms of the infected cells (Fig. 5).

Although the model described in Figs. 4 and 5 has not yet been fully proven experimentally, it is substantiated by a whole set of circumstantial pieces of evidence. Indeed, attempts to isolate the viral replication complexes showed that the viroplasms are not only involved in the assembly of the virus, but also in its replication (Volovitch et al., 1983). A DNA polymerase activity has been detected in virions, suggesting that synthesis of the viral DNA may be completed in immature virions during morphogenesis (Ménissier et al., 1984).

5. *In vivo* Viral Genome Recombination

It has been firstly demonstrated that the co-inoculation, onto-sensitive plants, of non infectious complementary mutants induced only the multiplication of wild-type virus (Walden and Howell, 1982). Similarly, co-inoculation of bacteria plasmids containing CaMV DNA inserted into various restriction sites led to expression of the inserted genome owing to recombinations occurring at the plasmid level: the more distant the restriction sites used for insertion of the viral DNA were, the higher the recombination frequency (Lebeurier et al., 1982). Recent observations showed that upon co-

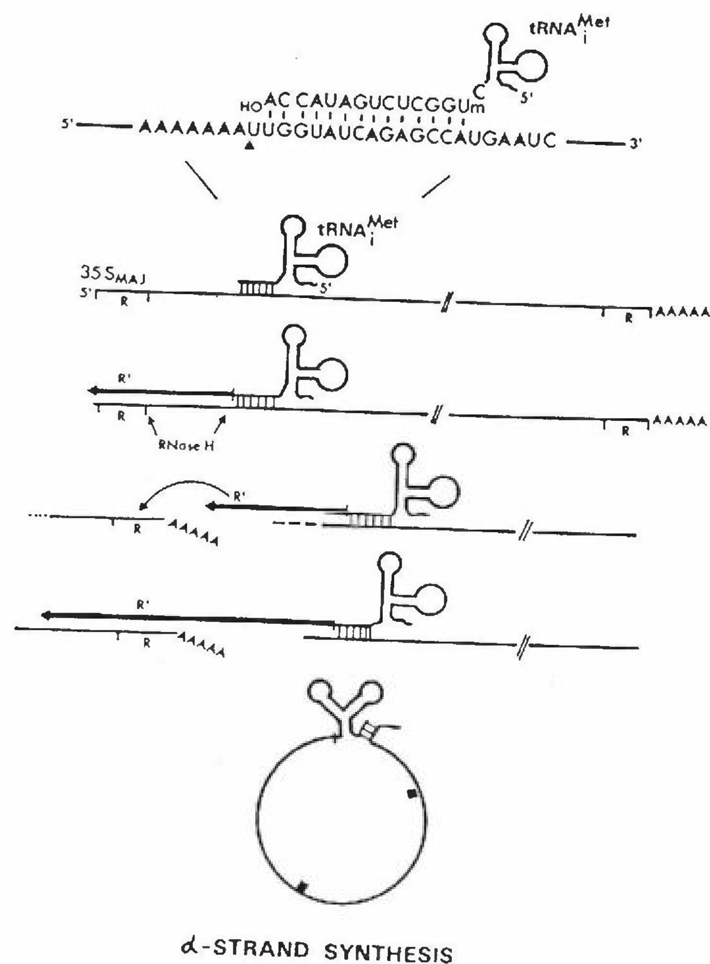

Figure 4. A hypothetical model for the replication of CaMV. The t-RNA^{Met} is used as a primer by the viral reverse transcriptase encoded by ORF V to copy the 5' region of the 35 S RNA into complementary single-stranded DNA. The direct terminal repeat in the 35 S RNA allows the newly synthesized R' DNA sequence to recognize the complementary R at the 3' end of the same RNA molecule (or possibly of another 35 S RNA molecule). This generates a newly synthesized α DNA strand which ends with an open complementary sequence. The 35 S RNA is digested by the RNase H activity associated with the viral reverse transcriptase which spares the polypurine tracts that will serve in turn as primers for synthesis of the β DNA strand. The synthesis of this second strand would then proceed until the next primer is reached: the advancing reverse transcriptase displaces it and synthesis stops a few nucleotides beyond the last ribonucleotide, thereby generating the triple-stranded regions which may retain a few ribonucleotides at their 3' end as a remnant of incompletely digested primer.

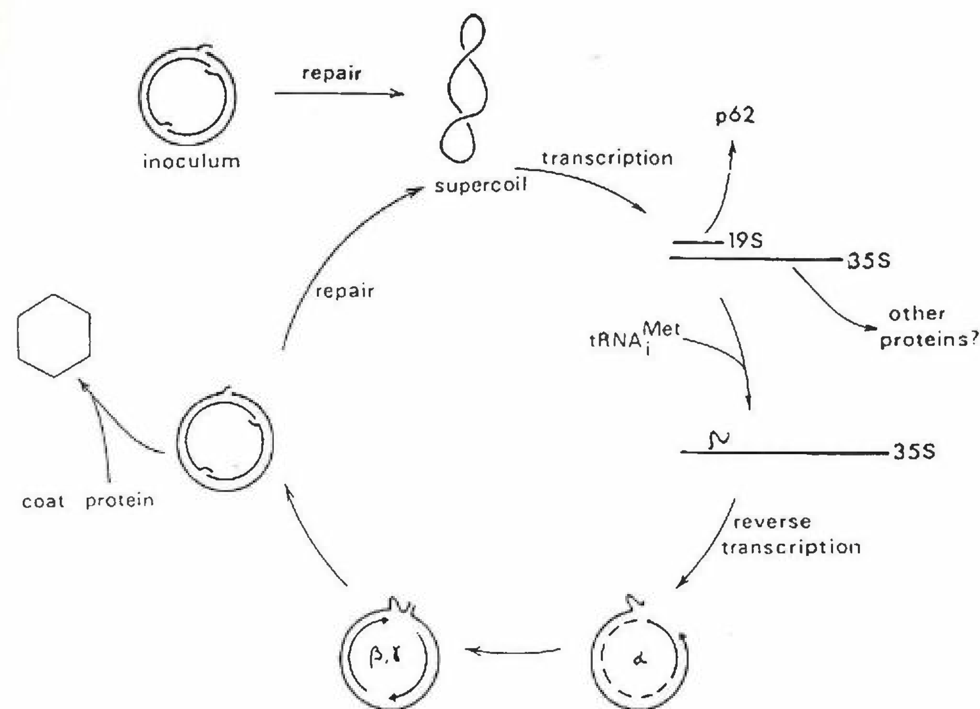

Figure 5. Cauliflower mosaic virus potential vector gene, genome organization and replication.

inoculation of two strains, one deleted and the other complete but linearized with the same restriction enzyme, recombinants between the two strains were generated through a dimeric intermediate formed at the level of the regions cut by the restriction enzyme. The dimeric molecule is formed in the inoculated cells. Then RNA polymerase II replicates a 35S RNA, using dimeric molecules as template. Such recombinants have a genome corresponding to the replication of a dimeric DNA molecule by an intermediate 35S RNA (Geldreich et al., 1986).

6. Cauliflower Mosaic Virus as a Gene Vector

Ten years ago it was thought that CaMV could be a vector for the introduction of foreign genes into higher plants. But unfortunately such experiments demonstrate that recombination processes are dominant between viral DNA molecules and that the foreign gene inserted in viral genome is eliminated. One exception was the insertion of methotrexate resistant gene into the localization of gene II (Brisson et al., 1984). In conclusion, there was no success until now in the use of CaMV as vector. This is why another strategy has

been tried: the T1 plasmid of *Agrobacterium tumefaciens* has been used to mediate integration of CaMV DNA: the CaMV genome has been inserted into the T-DNA and integrated into the nuclear genome of various plants. These plants express two transcripts corresponding to the 19S and 35S RNA, but of variable levels and relative amounts (Shewmaker et al., 1985). The last strategy is to place a foreign gene under the control of CaMV promoters and to introduce both by direct gene transfer in cell cultures of a monocotyle plant which is not sensitive to *Agrobacterium* infection (Fromm et al., 1985). Therefore certain sequences of CaMV DNA can be successfully used for expression of foreign genes in a wide range of higher plants. But still much work has to be done until CaMV can be used as a good vector.

As we have described above, CaMV uses a reverse transcriptase to replicate its genome. The same mechanism is also used by Retroviruses and Hepadnaviruses. Comparisons of the amino-acid sequences predicted from the known nucleotide sequences of members of the three virus groups have been published. Such studies showed striking similarities between the "pol" gene of retroviruses and an open reading frame of Hepadnaviruses and ORF V of Caulimoviruses. This is why Pfeiffer and Hohn (1983) have proposed the super group name of "retroid" viruses.

REFERENCES

Brisson, N., Paszkowski, J.R., Penswick, J.R., Gronenborn, B., Potrykus, I., and Hohn, T. 1984. Expression of a bacterial gene in plants by using a viral vector. *Nature* **310:** 511–514.

Condit, C., and Maegher, R.B. 1983. Multiple discrete 35 S transcripts of cauliflower mosaic virus. *J. Molec. Appl. Genet.* **2:** 301–314.

Covey, S.N. and Hull, R. 1981. Transcription of cauliflower mosaic encoding the virus inclusion body protein. *Virology* **111:** 463–474.

Dudley, R.K., Odel, J.T., and Howell, S.H. 1982. Structure and 5' termini of the large and 19 S RNA transcripts encoded by the cauliflower mosaic virus genome. *Virology* **117:** 19–28.

Franck, A., Guilley, H., Jonard, G., Richards, K., and Hirth, L. 1980. Nucleotide sequence of cauliflower mosaid virus DNA. *Cell* **21:** 285–294.

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.

Geldreich, A., Lebeurier, G., and Hirth, L. 1986. *In vivo* dimerization of cauliflower mosaic virus DNA can explain recombination. *Gene* (in press).

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**:no. 10 (in press).

Givord, L., Xiong, C., Giband, M., Koenig, I., Hohn, T., Leberier, G., and Hirth, L. 1984. A second cauliflower mosaic virus gene product influences the structure of the viral inclusive body. *EMBO J.* **3:** 1423–1427.

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.

Olszewski, N., Hagen, G., and Guilfoyle, T.J. 1982. A transcriptionally active covalently closed minichromosome of cauliflower mosaic virus DNA isolated from infected turnip leaves. *Cell* **29:** 395–402.

Pfeiffer, P. and Hohn, T. 1983. Involvement of reverse transcription in the replication of cauliflower mosaic virus: a detailed model and test of some aspects. *Cell* **33:** 781–789.

Richards, K.E., Guilley, H., and Jonard, G. 1981. Further characterization of the discontinuities in cauliflower mosaic virus DNA. *FEBS Lett.* **134:** 67–70.

Shewmaker, C.K., Caton, J.R., Houck, C.M., and Gardner, R.C. 1985. Transcription of cauliflower mosaic virus integrated into plant genomes. *Virology* **140:** 281–288.

Toh, H., Hayashida, H., and Miyata, T. f1983. Sequence homology between retroviral reverse transcriptase and putative polymerases of Hepatitis B virus and cauliflower mosaic virus. *Nature* **305:** 827–829.

Volovitch, M., Modjtahedi, N., Yot, P., and Brun, G. 1983. RNA dependent DNA polymerase activity in cauliflower mosaic virus infected leaves. *EMBO J.* **3:** 309–313.

Volovitch, M., Drugeon, G., and Yot, P. 1978. Studies on the single-stranded discontinuities of the cauliflower mosaic virus genome. *Nucleic Acids Res.* **5:** 2913–2925.

Walden, R.M. and Howell, S.H. 1982. Intergenomic recombination events among pairs of defective cauliflower mosaic virus genomes in plants. *J. Molec. Appl. Genet.* **1:** 447–456.

Xiong, C., Müller, S., Lebeurier, G., and Hirth, L. 1982. Identification by immunoprecipitation of cauliflower mosaic virus *in vitro* major translation product with a specific serum against viroplasm protein. *EMBO J.* **1:** 971–976.

Xiong, C., Lebeurier, G., and Hirth, L. 1984. Detection *in vivo* of a new gene product (gene III) of cauliflower mosaic virus. *Proc. Nato. Acad. Sci. USA* **1:** 6608–6612.

Ziegler, V., Laquel, P., Guilley, H., Richards, K., and Jonard, G. 1985. Immunological detection of cauliflower mosaic virus gene V protein produced in engineered bacteria. *Gene* **36:** 271–279.