Transcription of Cauliflower Mosaic Virus DNA: Detection of Promoter Sequences, and Characterization of Transcripts

Hubert Guilley, Roderick K. Dudley,
Gérard Jonard, Erwin Balàzs and
Kenneth E. Richards
Laboratoire de Virologie
Institut de Biologie Moléculaire et
Cellulaire du CNRS
15 rue Descartes
67084 Strasbourg Cédex, France

Summary

Four RNA transcripts encoded by cauliflower mosaic virus DNA have been detected in the polyadenylated RNA from virus-infected turnip leaves. Two of these transcripts, the major 35S and the 8S species, have the same 5′ termini, at nucleotide 7435. A viral DNA fragment encompassing this region directs transcription initiation at this point in vitro. The 5′ terminus of the 19S transcript is at nucleotide 5764, and a corresponding viral DNA fragment also directs transcription initiation in vitro. The major 35S RNA is a complete transcript of the circular viral genome, and is 3′-coterminal with 19S RNA at nucleotide 7615. The 8S RNA has its 3′ extremity at Δ1, the single-stranded interruption in the transcribed strand of virion DNA. A minor 35S RNA has also been detected that has its 5′ and 3′ termini at Δ1.

Introduction

Cauliflower mosaic virus (CaMV) is one of only a few double-stranded DNA plant viruses (for review see Shepherd, 1979), and as such it is a potential vector for the introduction of foreign DNA into plants. The viral genome is circular and about 8000 bp in length; the virion DNA of most strains contains three single-stranded breaks of unknown function, two in one strand of the DNA and one in the other (Volovitch et al., 1978). The single break (Δ1) in the so-called α strand is defined as the zero point of the standard map of the viral genome. The entire sequence of CaMV DNA is known for three strains of the virus, Cabb-BS (Franck et al., 1980), CM1841 (Gardner et al., 1981) and D/H (Balàzs et al., 1982), with about 5% sequence difference between strains. Only the α strand of DNA is transcribed (Howell and Hull, 1978). The sequence complementary to the α strand (that is, the sequence possessed by viral RNA transcripts) contains six open reading frames that together cover about 85% of the genome (Figure 1A).

Extraction of RNA from virus-infected plants has so far demonstrated the presence of two major viral RNA species (Covey et al., 1981; Odell et al., 1981). The smaller of these, termed 19S RNA, directs synthesis of a 62,000–66,000 dalton polypeptide in an in vitro translation system (Al Ani et al., 1980; Odell and

Howell, 1980; Covey and Hull, 1981). The 5′ end of this transcript was originally mapped by hybrid-arrested translation and electron microscopy to the junction between Eco RI fragments C and E (Figure 1), and the 3′ end was found to lie near Δ1 (Odell et al., 1981). Thus, 19S RNA encompasses open reading frame VI (Figure 1A). More recently, the S1 nuclease mapping procedure of Berk and Sharp (1978) and primer-extension techniques have been used to position the 5′ terminus of this transcript at nucleotide 5765 (of the Cabb-BS sequence), and the 3′ terminus at 0.95 map units (Covey et al., 1981; Dudley et al., 1982).

The other transcript, which we shall refer to as 35S RNA, hybridizes with all Eco RI fragments on Northern blots and has a size similar to that of the α strand of the viral DNA. On the basis of ultraviolet mapping experiments, Howell (1981) suggested that synthesis of 35S transcript begins within the long intergenic region (Figure 1) counterclockwise to Δ1. Low-resolution S1 nuclease mapping has since confirmed this suggestion, placing the 5′ extremity of 35S transcript at 0.93 map units—that is, 600 nucleotides upstream from Δ1 (Covey et al., 1981; Dudley et al., 1982). The 3′ terminus of 35S transcript was shown to be approximately coterminal with that of the 19S species at 0.95 map units (Covey et al., 1981), which implies that the 5′ and 3′ extremities of the major 35S transcript overlap by about 200 nucleotides.

The studies described above were performed with CaMV isolates B-JI (Covey et al., 1981) and CM4-184 (Dudley et al., 1982), for which extensive sequence data are not yet available. This is no drawback at the level of resolution that has been achieved so far because of the likely close similarity in the genetic organization of different CaMV strains, but it does represent an obstacle to precise localization of transcriptional control elements by in vitro transcription experiments. We present a high-resolution S1 nuclease mapping analysis of the 5′ and 3′ extremities of the 19S and major 35S transcripts of the Cabb-BS strain of CaMV, with results that confirm and further refine the already published mapping data. We have also detected and mapped two other virus-specific RNA species: a minor 35S transcript that begins and ends at Δ1, and an 8S RNA species of about 600 nucleotides with the same 5′ terminus as the major 35S transcript and a 3′ terminus at Δ1.

A HeLa-cell lysate, prepared so as to contain endogenous RNA polymerase II activity (Manley et al., 1980), or supplemented with purified enzyme (Wasylyk et al., 1980), can accurately initiate transcription of many animal and animal virus RNAs (for references see Breathnach and Chambon, 1981). We show that such a system can specifically initiate transcription on CaMV DNA restriction fragments containing the major 35S and 19S RNA start points, indicating that both species are primary transcripts. This observation also

suggests that the signals governing transcription of RNAs from animals and from higher plants are probably very similar.

Results

Detection of Cauliflower Mosaic Virus Transcripts

Total RNA was prepared from CaMV-infected turnip leaves 14 days after inoculation, several days after the appearance of severe symptoms. The bulk of the nonpolyadenylated RNA was eliminated by two passages through an oligo(dT) column, and the poly(A)[+] RNAs, denatured by glyoxylation, were separated from one another by electrophoresis through an agarose gel. After transfer to diazobenzyloxymethyl (DBM) paper, viral RNAs were revealed by hybridization with nick-translated CaMV DNA. The 35S and 19S CaMV transcripts described by others (Covey et al., 1981; Dudley et al., 1982) are readily visible in an autoradiograph of the DBM paper (Figure 2). We have also detected a smaller viral RNA species, referred to as 8S RNA. The heterogeneous material, indicated by brackets in Figure 2, migrating between the 35S and

19S transcripts has not yet been characterized, but is thought to represent degradation products of the 35S RNA.

Prior to S1 nuclease mapping, the various CaMV transcripts were separated from one another by sucrose gradient centrifugation. Fractions of the gradient were analyzed by agarose gel electrophoresis, transferred to DBM paper and hybridized with nick-translated CaMV DNA. The autoradiograph (Figure 2) shows that the 8S, 19S and 35S RNAs are well separated from one another on the gradient, eliminating the possibility of cross-contamination in the mapping experiments.

The 5' and 3' Extremities of the 35S and 19S Transcripts

To map the extremities of the various viral transcripts, we have employed the Berk and Sharp (1978) S1 nuclease mapping procedure as modified by Weaver and Weissmann (1979), using 5'- or 3'-^{32}P-end-labeled restriction fragments, first approximately, with reference to length standards, and then more precisely, by comparison with a sequence ladder of the appropriate end-labeled DNA fragment.

In preliminary experiments, 5'-^{32}P-labeled Eco RI fragments A–E were used to determine the approximate positions of 5' termini present in the 35S RNA fraction. Cloned DNA was used in preference to virion DNA in all the S1 nuclease mapping experiments. The
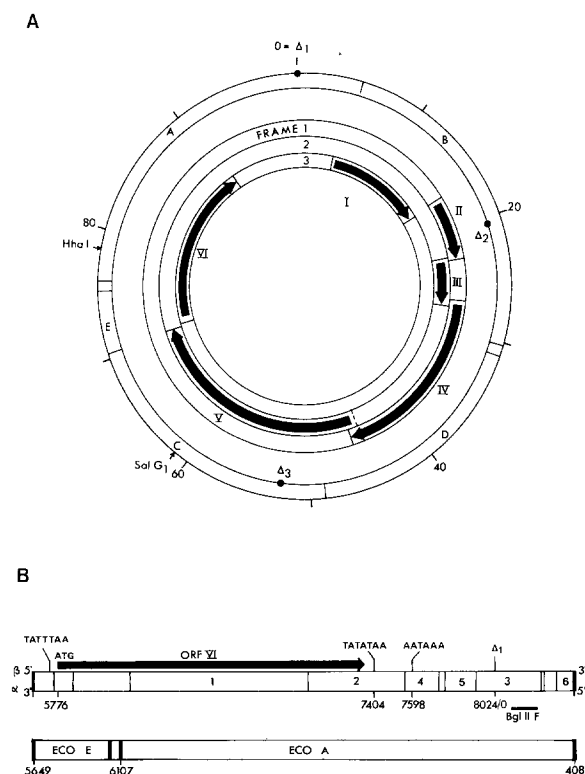

Figure 1. Map of Cauliflower Mosaic Virus

(A) Coding regions on CaMV DNA. (Inner circles) Positions on the DNA α strand of the six long open reading frames (I–VI) described by Franck et al. (1980). (Outer circle) Positions of Eco RI fragments A–E and the three discontinuities (Δ1–Δ3) in virion DNA.

(B) Expansion of the region containing Eco RI fragments A and E. (Top) The coordinates of Dde I fragments 1–6 and Bgl II fragment F within Eco RI fragment A. ORF: open reading frame. (Bottom) The coordinates of Eco RI fragments A and E.
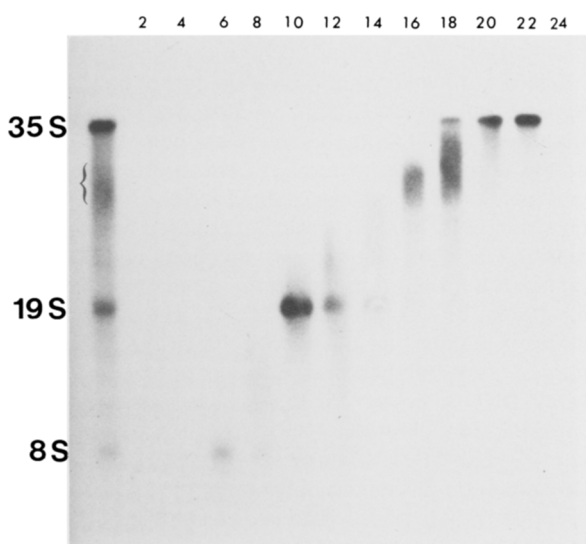

Figure 2. Separation of the 8S, 19S and 35S Viral Transcripts Present in Polyadenylated RNA from CaMV-Infected Turnip Leaves by Sucrose Gradient Centrifugation

A portion of each even-numbered gradient fraction (numbers at top) was denatured by glyoxylation, separated by electrophoresis through a 1.1% agarose gel and transferred to diazobenzyloxymethyl paper. Immobilized viral RNAs were detected by hybridization with CaMV DNA probe ^{32}P-labeled by nick translation. (Left lane) Sample of the RNA prior to gradient centrifugation. Bracket: uncharacterized heterogeneous viral RNA (see text). The RNA in gradient fractions 6 (8S), 10 (9S) and 22 (35S) was used for further studies.

two strands of each restriction fragment were separated from one another by polyacrylamide gel electrophoresis after denaturation, and the strand corresponding to the α strand, identified by sequence analysis, was incubated with gradient-purified 35S RNA under conditions favoring RNA–DNA hybridization. After elimination of nonhybridized nucleic acid with S1 nuclease, the nuclease-resistant RNA–DNA hybrids were denatured and run on a 5% polyacrylamide gel beside length markers. The result of experiments with Eco RI fragments B, C, D and E was a single S1-nuclease-resistant band corresponding to complete protection of the full-length single-stranded DNAs (data not shown). The same experiment performed with Eco RI fragment A gave rise to two major and one minor S1-nuclease-resistant bands (Figure 3A). The slowest-moving band corresponds to the full-length single-stranded DNA fragment, representing complete protection by 35S RNA. Evidence will be presented below that the 35S RNA fraction does not contain continuous sequences totally spanning the

Figure 3. S1 Nuclease Mapping of Gradient-Purified 35S RNA

The RNA was hybridized with 5′-end-labeled (A) or 3′-end-labeled (B) Eco RI fragment A (α strand). RNA–DNA hybrids were formed as described in the Experimental Procedures, and unhybridized nucleic acid was trimmed away with S1 nuclease. The hybrids were denatured, and the protected end-labeled DNA fragments were characterized by electrophoresis through a urea-containing 5% polyacrylamide gel. (Lanes 1) No S1 nuclease; (lanes 2) 300 U S1 nuclease; (lanes 3) 600 U S1 nuclease; (lane 4) 3000 U S1 nuclease. Length markers (bars; same in A and B) were single-stranded 5′-end-labeled Eco RI and Dde I fragments of CaMV DNA. MAJ and MIN: S1-nuclease-shortened fragments arising from the major and minor 35S transcripts (see text).

Eco RI A fragment, implying that protection of the full-length fragment is artifactual. The lengths of the two shorter S1-nuclease-resistant bands were 1000 and 400 nucleotides, respectively, suggesting that there are two species of 35S transcript, a major species with a 5′ end around nucleotide 7400 and a minor species with a 5′ end near Δ1. Control hybridization experiments (data not shown) carried out with a 35S RNA fraction from uninfected plants or with the β strand of Eco RI fragment A gave no bands resistant to S1 nuclease. Similar S1 nuclease mapping experiments with 35S RNA with the α strand of 3′-^{32}P-labeled Eco RI fragment A as probe also gave a major and a minor shortened S1-nuclease-resistant DNA fragment in addition to the full-length Eco RI fragment (Figure 3B). The lengths of the shortened fragments correspond to 35S RNA 3′ termini at about nucleotide 7600 and Δ1, respectively, for the major and minor species.

The precise sequence coordinates of the two 35S RNA species were established by S1 nuclease mapping with shorter 5′- and 3′-end-labeled restriction fragments encompassing the 5′ and 3′ extremities that had been approximately located in the experiments shown in Figure 3 (for details see Experimental Procedures; legend to Figure 4). Autoradiographs of polyacrylamide gels showing the S1-nuclease-resistant DNA from RNA–DNA hybrids beside sequence ladders from the appropriate restriction fragment are shown in Figure 4; the results are summarized in Figure 7. The nucleotide coordinates of the 3′ and 5′ termini given in Figure 7 correspond in each case to the major S1-nuclease-resistant fragment detected in the mapping experiment. Often, the major band was accompanied by minor bands migrating 1 or 2 nucleotides faster or slower. It is not known whether these minor bands reflect length microheterogeneity of the extremities or incomplete S1 nuclease digestion (or overdigestion) of the RNA–DNA hybrids.

Gradient-purified 19S RNA was also subjected to S1 nuclease mapping with the α strands of a 5′-labeled Hind III fragment (nucleotides 5376–5851) and of 3′-labeled Dde I fragment 4. The results (data not shown) place the 5′ extremity of this transcript at nucleotide 5764, 12 nucleotides upstream from the first potential translational initiation codon of open reading frame VI. This is in close agreement with the work of Covey et al. (1981), who located the 19S RNA start point at nucleotide 5765 by primer extension. The 3′ extremity of 19S transcript was mapped to nucleotide 7615, again confirming the conclusion of Covey et al. (1981), based on low-resolution S1 nuclease mapping data, that the major 35S transcript and the 19S transcript are 3′-coterminal.

Mapping the 5′ Extremities of 35S Transcripts by Primer Extension

The fact that a large fraction of the end-labeled Eco RI fragment A used in the mapping experiments shown
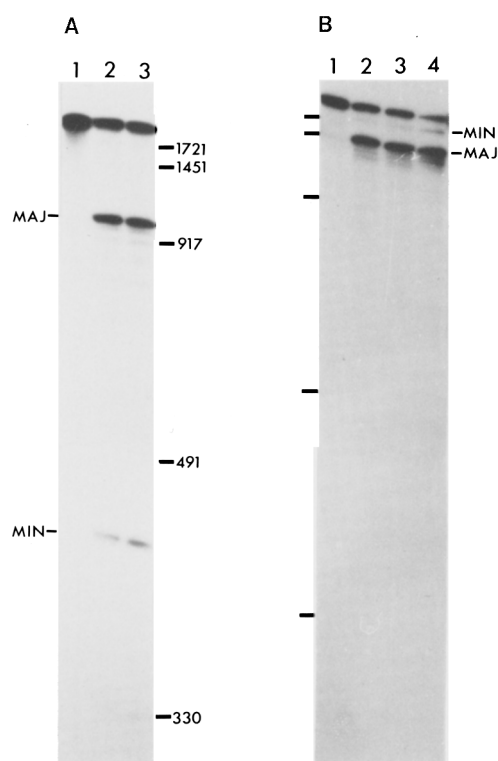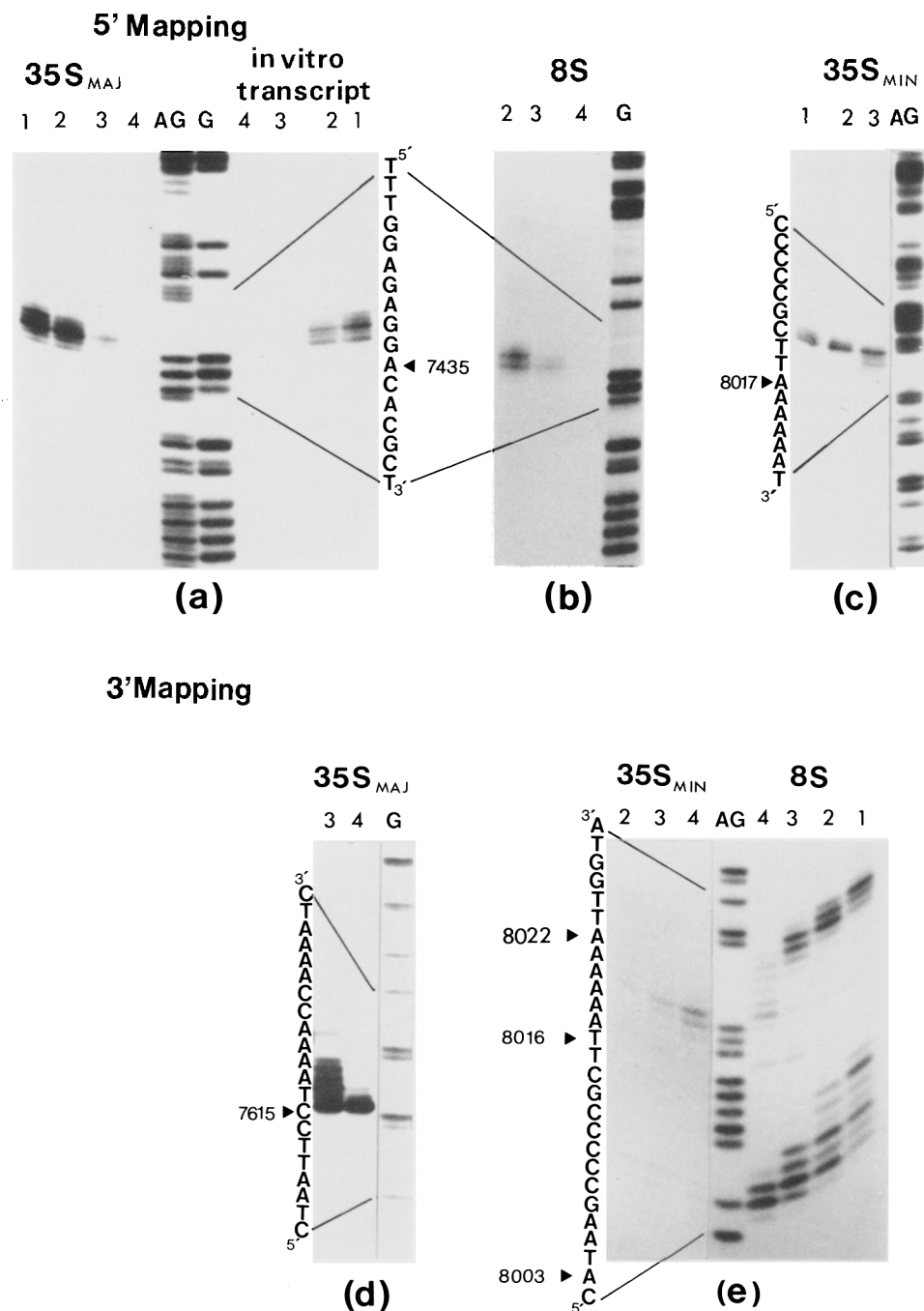
Figure 4. High-Resolution S1 Nuclease Mapping of the 5′ and 3′ Extremities of CaMV RNAs, and Localization of the 5′ Terminus of the RNA Transcribed in Vitro from the Major 35S RNA Promoter

The α strand of the appropriate 5′- or 3′-^{32}P-labeled restriction fragment was annealed with sucrose-gradient-purified RNA or the in vitro–synthesized runoff transcript (see Experimental Procedures), and after S1 nuclease digestion, the protected DNA was characterized by electrophoresis through an 8% sequencing gel beside a sequence ladder of the same DNA fragment. The hybrid was digested with 300 U S1 nuclease (lanes 1), 600 U S1 nuclease (lanes 2), 3000 U S1 nuclease (lanes 3) or 6000 U S1 nuclease (lanes 4). In each case the nucleotide sequence given is that of the complementary β strand, possessing the same polarity as viral RNA. Thin lines: principal 5′ or 3′ terminus of each transcript. The restriction fragments used for mapping were: Dde I fragment 2 (a and b), Bgl II fragment E (nucleotides 7664–7692; c), Dde I fragment 4 (d) and the small Cla I–Bst E2 fragment (nucleotides 7982–8024, 0–127; e). The runoff transcript used in the in vitro transcription experiment in (a) was synthesized from Eco RI fragment A (see text). Length markers were single-stranded 5′-end-labeled Eco RI and Dde I fragments of CaMV DNA. 35S$_{MAJ}$ and 35S$_{MIN}$: major and minor 35S transcripts.

in Figure 3 remained full-length after S1 nuclease digestion suggests the possible existence of other types of 35S RNA, either with extremities outside of Eco RI fragment A or possessing no extremities at all (that is, circular molecules). To test this possibility, we have used reverse transcriptase to synthesize DNA copies of 35S RNA within the Eco RI fragment A region. The α strand of Bgl II fragment F (Figure 1B), ^{32}P-labeled at its 5' terminus, was used as primer. After annealing it to 35S RNA, we extended the primer with reverse transcriptase in the presence of actino-mycin D, and analyzed the cDNA products by poly-acrylamide gel electrophoresis. Autoradiography of the gel (Figure 5) revealed two cDNA species migrat-ing more slowly than the unextended primer of 129 nucleotides: a major species of approximately 800 nucleotides, and a minor species of 230 nucleotides. The major and minor cDNAs were present in relative proportions of 15:1, as determined by counting the excised gel bands. Their estimated lengths corre-spond closely to expectation for cDNAs terminating by runoff at nucleotide 7435 and Δ1, respectively, thus substantiating our hypothesis that 35S RNA pos-sesses 5' termini at these two points. The results also suggest that the 35S RNAs are not spliced, at least in this region of the genome, in agreement with electron microscopic examination of R loops between 35S RNA and CaMV DNA (Odell et al., 1981) and with the findings of Covey et al. (1981). Finally, the fact that essentially no cDNA longer than 800 nucleotides can be detected rules out the existence of circular RNA transcripts or molecules with 5' termini outside of Eco RI fragment A in the 35S fraction.

Characterization of the 8S Transcript

In addition to the 19S and 35S transcripts described above, the poly(A)[+] RNA fraction from CaMV-infected cells contains a small virus-specified RNA that sedi-ments at about 8S in a sucrose gradient (Figure 2). Upon electrophoresis through a polyacrylamide gel under denaturing conditions, the virus-specific RNA in the 8S fraction migrates as a rather broad band cor-responding to chain lengths of about 610–680 nu-cleotides (data not shown). Preliminary experiments with nick-translated CaMV Eco RI fragments A–E as radioactive probes showed that only Eco RI fragment A hybridizes with the 8S transcript. ^{32}P-labeled cDNA prepared from 8S RNA with oligo(dT) as a primer hybridized specifically with Dde I fragments 2, 3, 4 and 5 (immobilized on DBM paper) but not with Dde I fragments 1 and 6 (data not shown). Thus the se-quence corresponding to the 8S transcript must lie in the large intergenic region between the end of open reading frame VI and Δ1.

More precise mapping of the 8S transcript was achieved by primer extension. The α strands of Dde I fragments 2, 3, 4 and 5, each 5'-^{32}P-end-labeled,
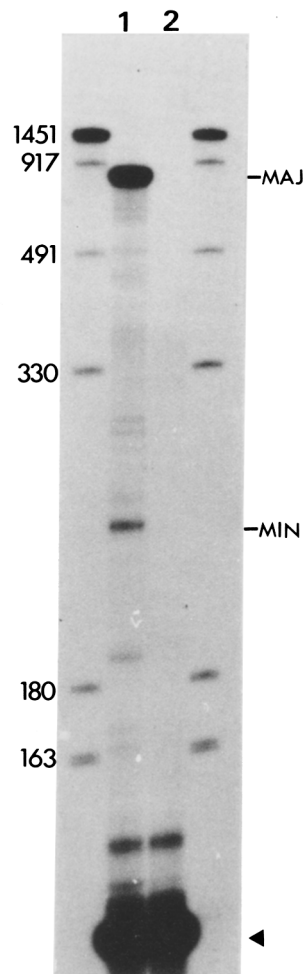

Figure 5. Mapping the 5' Termini of 35S RNAs by Primer Extension

The primer was the 5'-^{32}P-labeled α strand of Bgl II fragment F (see Figure 1B). After we annealed the primer with gradient-purified 35S RNA from virus-infected plants (lane 1) or healthy plants (lane 2), we extended it with reverse transcriptase. The resulting cDNAs were analyzed on a denaturing 6% polyacrylamide gel. (Right and left lanes) Length markers, 5'-end-labeled Eco RI and Dde I fragments of CaMV DNA. Arrowhead: position of unextended primer. MAJ and MIN: cDNAs terminating at the 5' extremities of the major and minor 35S RNAs.

were annealed to gradient-purified 8S fraction and extended with reverse transcriptase. Analysis of the resulting cDNAs on a polyacrylamide gel revealed a prominent cDNA band when Dde I fragment 3, 4 or 5 was used as primer (Figure 6), but no such prominent band was visible for the reaction primed by Dde I fragment 2.

For the reverse-transcription reactions primed by Dde I fragments 3 and 5, the major cDNA species was essentially the only elongated product detected, while in the case of Dde I fragment 4 the prominent cDNA was accompanied by numerous minor bands (see Figure 6a). We suggest that this background was
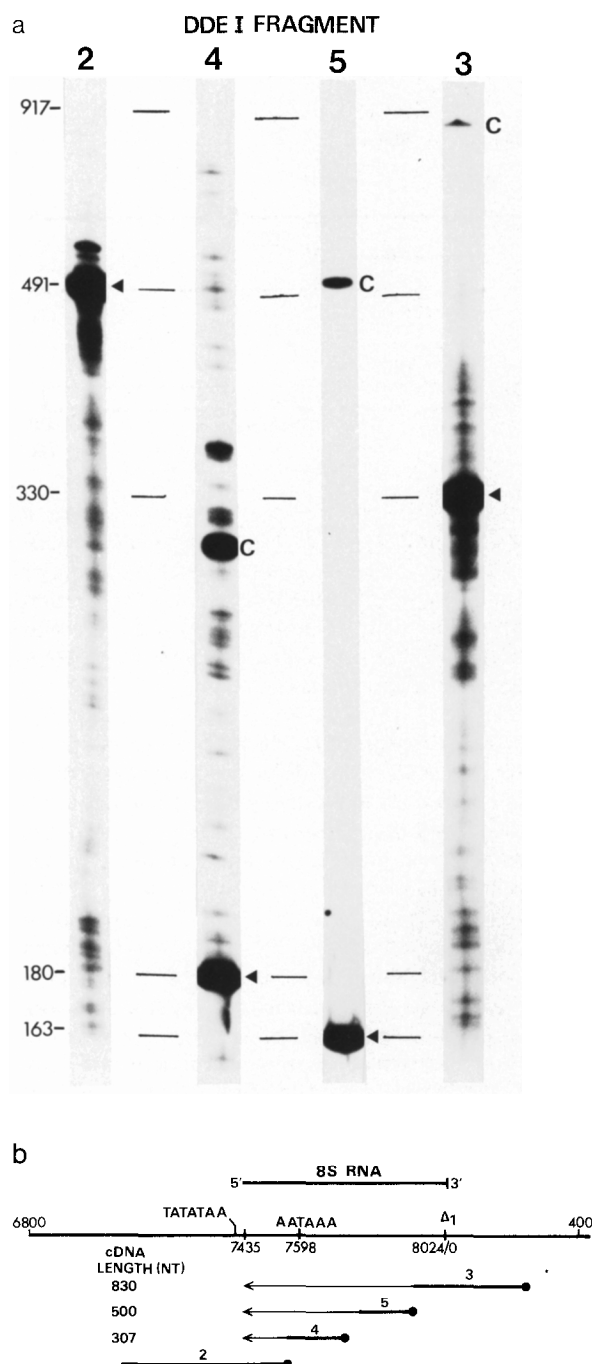
a

DDE I FRAGMENT

b

Figure 6. Mapping the 5' Terminus of 8S RNA by Primer Extension

(a) The primers were 5'-^{32}P-labeled α strands of Dde I fragments 2, 3, 4 and 5. cDNAs were characterized by electrophoresis on a denaturing 6% polyacrylamide gel. Length markers (bars) were single-stranded 5'-end-labeled Eco RI and Dde I fragments of CaMV DNA. Arrowheads: positions of unextended primers. (b) Expected lengths of each major cDNA species (830, 500 and 307 necleotides [NT]; C in [a]). Thin arrows: extension of primer.

caused by contamination of the 8S fraction with short polyadenylated degradation products of the major 35S and 19S RNAs. Such degradation products would not be eliminated by oligo(dT)-cellulose chromatography or sucrose gradient centrifugation. Dde I

fragments 2 and 4 (but not fragments 3 and 5) would anneal to them and yield short cDNAs of diverse length upon extension.

The estimated lengths of the major newly synthesized cDNAs primed by Dde I fragments 3, 4 and 5 correspond to termination of synthesis at approximately nucleotide 7430, near the transcriptional initiation point of the major 35S RNA. High-resolution S1 nuclease mapping of 8S RNA with 5'-end-labeled Dde I fragment 2 has confirmed that the 8S and major 35S RNAs are in fact exactly 5'-coterminal at nucleotide 7435 (Figure 4).

In view of the estimated length of the 8S transcript (610–680 nucleotides, including the poly[A] tail), its 3' extremity is expected to lie near Δ1. S1 nuclease mapping of 8S RNA with the 3'-end-labeled small Cla I–Bst E2 fragment produced a bimodal pattern of nuclease-resistant bands (Figure 4). One set of bands converged toward nucleotide 8003 with increasing S1 nuclease concentration; the other set of bands converged toward nucleotide 8022. At present we cannot judge whether the two sets of bands represent distinct 3' termini, reflecting 3'-terminal heterogeneity of the 8S RNA fraction, or whether one of the two sets of bands arises from overdigestion or underdigestion with S1 nuclease of the RNA–DNA hybrid. Nevertheless, the data do substantiate our conclusion that the 3' terminus of 8S RNA falls near Δ1.

The sequence coordinates of the various polyadenylated transcripts described above are summarized in Figure 7. In the diagram it has been assumed that the major 5' and 3' extremities (at nucleotides 7435 and 7615, respectively) detected in the 35S RNA fraction are associated with one and the same species, the so-called major 35S transcript, while the minor termini belong to a second species, the minor 35S transcript, beginning and ending at Δ1. It is difficult, however, to rule out the existence in small quantities of transcripts representing the permutations of this basic scheme— that is, with 5' extremity at Δ1 and 3' extremity at nucleotide 7615 or with 5' extremity at nucleotide 7435 and 3' extremity at Δ1. If neither of these forms is present, then the relative amounts of radioactivity present in the major and minor bands detected in the S1 nuclease mapping experiments with end-labeled Eco RI fragment A should be identical for 5'- and 3'-end-labeled Eco RI fragments. In fact, visual inspection of the original autoradiographs shown in Figure 3 suggests that relatively speaking, there was somewhat more of the minor band generated when the 5'-labeled Eco RI fragment A was used as probe in the S1 nuclease mapping experiments than when the 3'-labeled fragment was used, consistent with the existence in the 35S fraction of a significant amount of a long transcript starting at Δ1 and ending at nucleotide 7615. Alternatively, the departure from the ideal ratio may simply reflect differential sensitivity to S1 nuclease of the major and minor 5' and 3' termini.

The overlap between the 5' and 3' extremities of

A

B

```
                              A TTTTTTA-5'
                            A
virion  α   3'-GTATTCGGGGGCG ATTTTTTAAC-3'
DNA     β   5'-CATAAGCCCCGCTTAAAAAATTG-5-
                         |          |
                        8010       8020


                            3' 5'
                          V‾‾‾‾‾‾‾‾‾V
35S min  β  5'-CATAAGCCCCGCTTAAAAAATTG


                            ?
                          V‾‾‾‾‾‾‾‾‾V
8S       β  5'-CATAAGCCCCGCTTAAAAAATTG
```
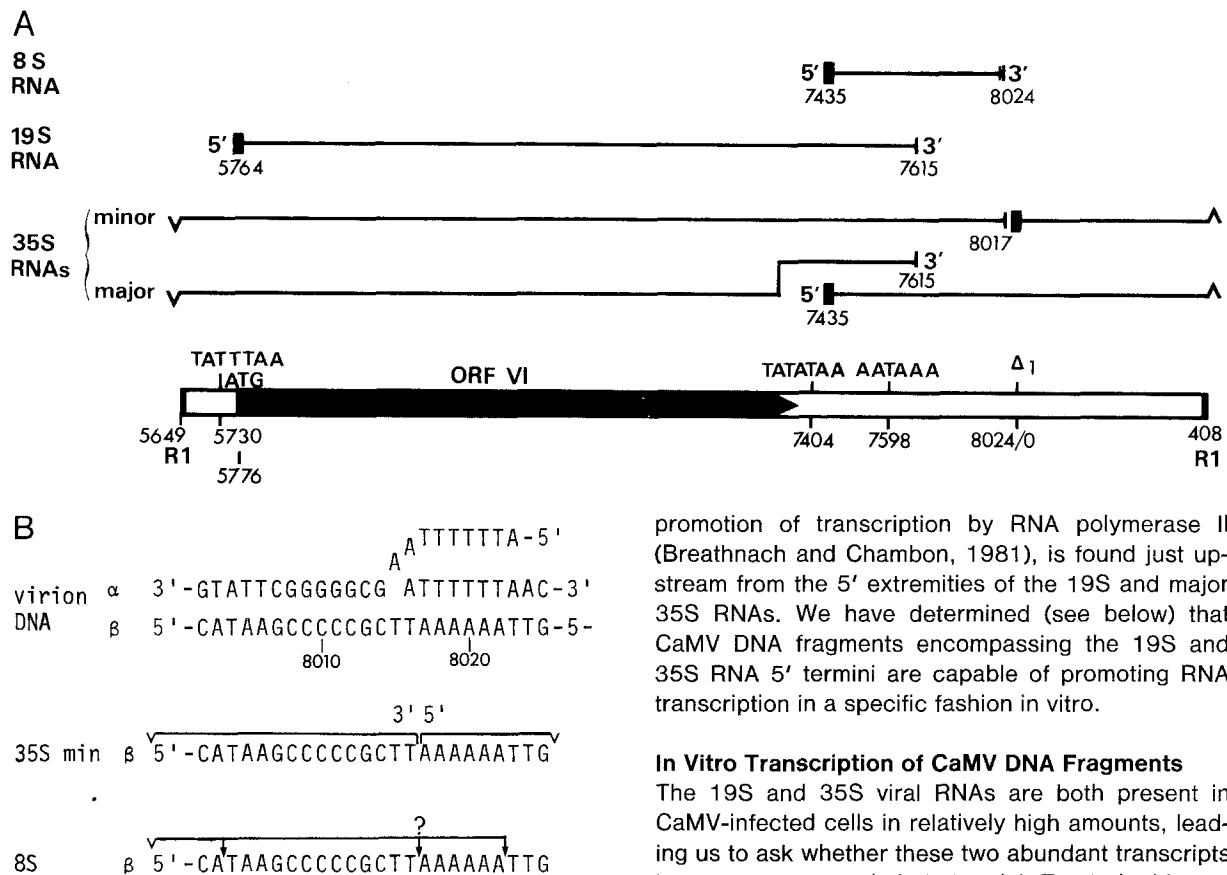
Figure 7. Sequence Coordinates of Various Polyadenylated Transcripts

(A) Coordinates of the four viral RNA transcripts. R1: Eco RI site. ORF: open reading frame. Thick bars: 5' ends. Thin bars: 3' ends. √‾: transcripts extend beyond the limits of the portion of the DNA sequence shown. (B) Sequence around the discontinuity Δ1 in virion DNA (Richards et al., 1981), showing the start and stop points of the minor 35S RNA, and the 3' termini (arrows) of 8S RNA.

major 35S RNA is presumably responsible for the persistence of some S1-nuclease-resistant full-length DNA fragments in the mapping experiments with 35S RNA and Eco RI fragment A (Figure 3). A single 35S RNA molecule might sometimes be expected to loop back upon itself and hybridize with both the 3' and 5' portions of the DNA fragment. S1 nuclease digestion would eliminate the single-stranded RNA tail corresponding to the sequence overlap, but the resulting RNA–DNA duplex, with only a single nick in the RNA chain, might be relatively resistant to further digestion.

Figure 7 shows the position of several sequence motifs that we and others (Franck et al., 1980; Covey et al., 1981; Gardner et al., 1981; Howell, 1981; Dudley et al., 1982) have suggested may play a role in initiation and termination of transcription. The sequence AATAAA, which is found about 20 nucleotides upstream from the poly(A) tail of most eucaryotic RNAs (Proudfoot and Brownlee, 1976), occurs 18 nucleotides before the end of the CaMV 19S and major 35S transcripts. A sequence resembling the eucaryotic TATA box, believed to be important for

promotion of transcription by RNA polymerase II (Breathnach and Chambon, 1981), is found just upstream from the 5' extremities of the 19S and major 35S RNAs. We have determined (see below) that CaMV DNA fragments encompassing the 19S and 35S RNA 5' termini are capable of promoting RNA transcription in a specific fashion in vitro.

In Vitro Transcription of CaMV DNA Fragments

The 19S and 35S viral RNAs are both present in CaMV-infected cells in relatively high amounts, leading us to ask whether these two abundant transcripts have promoters at their 5' termini. To study this, we have used the in vitro transcription system described by Manley et al. (1980). Fragments of the viral genome containing the 5' ends of the 35S and 19S RNAs were cloned into pBR322, and the plasmid DNA was prepared. The inserted DNA, or portions thereof, was excised with the appropriate restriction enzyme(s), and the insert and plasmid were separated from one another on sucrose gradients. When the purified Eco RI fragment A was added to the in vitro system prepared from HeLa cells, a transcript of approximately 1 kb was synthesized (Figure 8A). In an identical reaction containing α-amanitin (1 μg/ml) this band was not synthesized, demonstrating that it is a product of transcription by RNA polymerase II. Presuming that this band at 1 kb represents a true runoff fragment, this positions the 5' end of the in vitro–synthesized transcript at approximately nucleotide 7440.

To map exactly the 5' end of the transcript, we used RNA synthesized in vitro off the Eco RI fragment A in an S1 mapping experiment. The RNA synthesized in vitro in a reaction lacking labeled nucleotides was hybridized to the α strand of Dde II fragment 2 (Figure 1B) that was labeled at its 5' end with ^{32}P. After digestion with S1 nuclease, the products were analyzed beside a sequence ladder of the Dde I fragment and the S1 nuclease products of the same reaction with 35S RNA isolated from CaMV-infected plants instead of RNA synthesized in vitro. Figure 4 shows that a very similar pattern of S1-resistant bands was
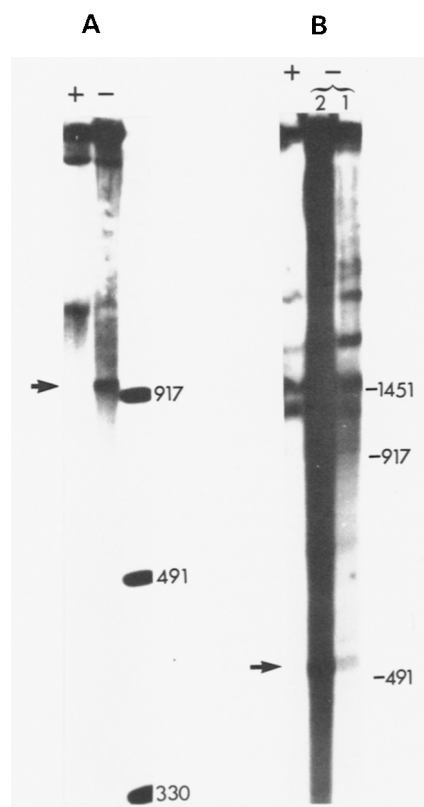
Figure 8. In Vitro Transcription of Restriction Fragments from Cloned CaMV DNA Containing the 35S and 19S Promoters

(A) Transcription of Eco RI fragment A in the HeLa-cell lysate in the presence (lane +) or absence (lane −) of α-amanitin. Concentration of added DNA was 80 μg/ml. (B) Transcription of the small Sal I–Hha I fragment in the presence (lane +) or absence (lanes −) of α-amanitin. Concentration of added DNA was 80 μg/ml (lanes + and 1) or 120 μg/ml (lane 2). Arrows: runoff transcripts described in the text.

produced with the in vitro–synthesized RNA and the RNA isolated from CaMV-infected plants, suggesting strongly that the in vitro– and in vivo–synthesized RNAs have the same 5' end. In a parallel experiment in which α-amanitin (1 μg/ml) was added to the in vitro transcription reaction, no S1-nuclease-resistant bands were detected after hybridization to Dde I fragment 2. Thus our results show that the HeLa-cell system faithfully transcribes the viral DNA in this region of the genome, and that the HeLa-cell RNA polymerase II has recognized a sequence that acts to promote transcription of the major 35S RNA. In contrast, the transcription start signal for minor 35S RNA synthesis does not appear to function in the in vitro system with cloned Eco RI fragment A, since no runoff transcript of the predicted size of 415 nucleotides could be detected.

To test for promoters in the vicinity of the 19S RNA 5' end, we added the small Sal I–Hha I fragment of CaMV DNA (nucleotides 4837–6283) to the in vitro transcription system. An α-amanitin-sensitive tran-

script of 520 nucleotides was synthesized that, assuming this is a true runoff fragment, corresponds to a 5' terminus at approximately nucleotide 5750 (Figure 8B). Experiments designed to locate the 19S in vitro transcription start point more precisely by S1 nuclease mapping have not been successful, presumably due to the low level of in vitro transcription from this promoter.

Discussion

Using an in vitro transcription system prepared from HeLa cells, we have obtained evidence for at least two RNA polymerase II promoters in the CaMV genome. S1 nuclease mapping of in vitro–synthesized transcripts has shown that one of these promoters is active in eliciting synthesis of a transcript with a 5' terminus identical to that of the major 35S and 8S RNAs found in CaMV-infected plants. Thus our observations suggest that the signals governing transcriptional initiation in plants are likely to be similar to those in other eucaryotic systems, and that the components of the HeLa-cell lysate will recognize plant DNAs as well as DNAs derived from animal cells or viruses. Examination of the sequence around the transcription start point reveals a "TATA-box" sequence, TATATAA, situated 31 nucleotides upstream from the 5' end of the 35S and 8S RNAs. It is well known that such sequences are generally found 26–34 nucleotides upstream from polymerase II transcriptional initiation sites in eucaryotes (Breathnach and Chambon, 1981). In fact, the major 35S RNA TATA box conforms closely to the consensus sequence

$$T_{49}A_{58}T_{56}A_{51}T_{22}^{A38} A_{53}T_{20}^{A30}$$

(subscripts refer to the number of times in 60 cases that the indicated base is found in that position), defined by comparison of a large number of such sequences (Breathnach and Chambon, 1981).

The other major viral transcript, 19S RNA, also has an AT-rich sequence, TATTTAA, 33 nucleotides upstream from its 5' end. Our in vitro transcription studies have shown that there is promoter activity associated with this region, thus tending to rule out models of CaMV transcription in which the 19S RNA is processed from the 3' end of the major 35S RNA. Although a systematic comparison has not yet been carried out, the 19S RNA promoter appears to be much less active in the in vitro system than is that of the 35S RNA. It will be interesting to determine whether this difference in activity is linked to the sequence difference between the two TATA boxes. A useful material to investigate this question will be the DNA of CaMV strain D/H, in which the T in position 4 of the 19S RNA TATA box is replaced by A (Balàzs et al., 1982).

An unusual feature of the major 35S transcript is the 180 nucleotide sequence overlap between its 5' and 3' extremities (Figure 7). We have no data on the mechanism by which the RNA polymerase II molecule,

during transcription of the major 35S RNA, passes a termination signal at the position of the 3' end of the transcript on its first circuit of transcription but stops at the same termination signal on its second circuit. One possibility is that the termination signal is leaky, in which case we would also expect to find viral transcripts longer than 35S, as well as an RNA of 180 nucleotides corresponding to transcription of the overlapping portion of the 35S RNA from the 5' terminus to the 3' terminus. Another possibility is that the transcriptional termination signal involves participation of sequences or secondary structure on the nascent RNA transcript that lie upstream from the 35S RNA initiation site. A similar suggestion has been put forth to explain readthrough of the termination signal in the lefthand long terminal repeat of integrated retrovirus DNA (Benz et al., 1980).

The discontinuity in the α strand of virion DNA at $\Delta 1$ should present an obstacle to synthesis of the major 35S RNA. Recently, however, Ménissier et al. (1982) have demonstrated the existence of supercoiled CaMV DNA in virus-infected cells. We regard it as likely that this supercoiled form, which must lack the single-stranded interruptions characteristic of virion DNA, is the template for major 35S RNA synthesis. The 8S transcript is 5'-coterminal with major 35S RNA, and its synthesis is presumably controlled by the same promoter. We suggest, however, that unencapsidated DNA of the interrupted virion form rather than supercoiled DNA serves as template for 8S RNA transcription. Termination of its synthesis may occur by runoff at the $\Delta 1$ discontinuity or by detachment of the polymerase a few nucleotides upstream from the interruption. It is noteworthy that both 8S and minor 35S RNAs are apparently polyadenylated, in spite of the absence of an AATAAA signal in the vicinity of their termination site at $\Delta 1$.

There is no TATA box sequence near $\Delta 1$, and we have failed to detect promoter activity on this region of the uninterrupted cloned DNA used in the in vitro transcription experiments. This need not preclude transcriptional initiation of minor 35S RNA at $\Delta 1$ on the interrupted form of CaMV DNA, however, since RNA polymerase II has been shown to be capable of preferentially starting RNA synthesis at a discontinuity (Lewis and Burgess, 1980; Cooke et al., 1981).

The picture of CaMV RNA transcription emerging from our work and that of others is still incomplete. In particular, it remains to be seen whether separate viral RNAs corresponding to coding regions I-V exist, perhaps in the nonpolyadenylated RNA fraction, or whether synthesis of these polypeptides is directed by the 35S transcript. Thus far, however, our efforts to translate 35S RNA in the reticulocyte cell-free system have been unsuccessful. The significance of the 8S and minor 35S RNAs likewise remains unclear. In so far as the 8S and minor 35S RNAs arise by transcription of the interrupted rather than the super-

coiled form of CaMV DNA, it is possible that they are without function, resulting from the adventitious action of polymerase on the pool of interrupted CaMV DNA existing free in the infected cell (Ménissier et al., 1982). If this is so, however, it is somewhat difficult to explain why no viral transcripts originating from $\Delta 2$ and $\Delta 3$, the two discontinuities in the β strand, are detected. Thus we feel that the alternative hypothesis also merits consideration: that the 8S and minor 35S transcripts, as well as the 19S and major 35S transcripts, all function in the viral infectivity cycle, and that generation of the single-stranded discontinuity at $\Delta 1$ represents a mechanism for switching from one class of transcript to the other.

Experimental Procedures

Preparation and Analysis of Polyadenylated CaMV RNAs
Young turnip plants (variety Just Right) were mechanically inoculated with CaMV (Cabb-S), and the virus was allowed to multiply for 14 days under controlled conditions (23°C, 2000 lux for 16 hr each day). Total nucleic acid was extracted from those leaves showing pronounced symptoms by a modification of the method of Howell and Hull (1978). Fifty grams of leaf tissue was quick-frozen in liquid nitrogen and then immediately ground at high speed in a Waring Blender with 75 ml TNS buffer plus 1% diethylpyrocarbonate and 50 ml phenol per gram of tissue (TNS buffer is 1% sodium triisopropylnaphthalene sulfonate, 6% p-amino-salicylic acid, 50 mM NaCl, 10 mM Tris–HCl [pH 7.5]). After low-speed centrifugation, the aqueous phase was reextracted with phenol and then three times with ether. Nucleic acid was ethanol-precipitated and collected by centrifugation, and the pellet was resuspended in 25 ml ice-cold 3 M sodium acetate (pH 6). The high molecular weight RNA, which does not dissolve, was collected by centrifugation and washed two more times with 3 M sodium acetate. Finally, the residual RNA was dissolved in water and ethanol-precipitated. RNA was prepared from healthy turnip plants in a similar manner to serve as a control.

Polyadenylated RNA was separated from nonpolyadenylated molecules by two cycles of chromatography on oligo(dT)-cellulose (P-L Biochemicals) and then concentrated by ethanol precipitation. Sucrose gradient centrifugation was carried out for 16 hr at 30,000 rpm in a Beckman SW41 rotor. The gradient was 5%–20% sucrose made up in 10 mM Tris–HCl (pH 7.5), 1 mM EDTA, 0.1% SDS. The polyadenylated RNA sample, no more than 200 μg, was heated to 50°C for 2 min in the gradient buffer and then quick-chilled before being layered on the gradient. Gradient fractions of 0.5 ml were ethanol-precipitated with 20 μg carrier tRNA, and the precipitate, after washing with 70% ethanol, was taken up in 30 μl sterile water.

CaMV-specific RNAs were detected by hybridization to a nick-translated CaMV DNA probe. RNA (1 μl of each gradient fraction) was denatured by glyoxylation and subjected to electrophoresis through a 1% agarose gel as described by McMaster and Carmichael (1977). The separated RNAs in the gel were transferred by blotting to freshly prepared DBM paper (Alwine et al., 1977). Hybridization of the paper with nick-translated CaMV DNA was carried out as described by Alwine et al. (1977).

Preparation of CaMV DNA Restriction Fragments
Starting material was either Cabb-S CaMV cloned via its unique Sal I site into the Sal I site of pBR322, or the CaMV Eco RI fragment A cloned into the pBR322 Eco RI site. Cloned CaMV Eco RI fragments B, C, D and E were also used in preliminary experiments. For in vitro transcription experiments, the desired restriction fragment was excised from the cloned DNA with the appropriate restriction enzyme or enzymes and separated from other fragments by sucrose gradient centrifugation.

Single-stranded 5'-^{32}P-labeled DNA fragments for S1 nuclease mapping, primer extension and sequencing were prepared from

cloned CaMV DNA essentially as described by Franck et al. (1980). 3'-end-labeled DNA fragments were prepared by filling in the restriction site of the double-stranded restriction fragment (Smith and Calvo, 1980) with reverse transcriptase (provided by J. Beard) and the appropriate α-^{32}P-deoxynucleotide triphosphate. Preparation of single-stranded DNA fragments from the double-stranded material then followed the procedures outlined by Franck et al. (1980).

In Vitro Transcription

HeLa-cell lysates for use in in vitro transcription were prepared exactly as described by Manley et al. (1980). In vitro transcription was carried out as described by Wasylyk et al. (1980), except that no RNA polymerase was added to the lysate and DNA fragments were added at concentrations ranging from 50 to 100 μg/ml. For the preparation of runoff transcripts for use in S1 nuclease mapping experiments, radioactive α-^{32}P-CTP was replaced by nonradioactive CTP and the reaction volume was doubled.

S1 Nuclease Mapping

S1 nuclease mapping was carried out by a modification of the procedure of Weaver and Weissmann (1979). For experiments with 5'- or 3'-labeled single-stranded Eco RI fragments, RNA (4 μl of gradient fraction) and DNA were dissolved in hybridization buffer (0.4 M NaCl, 0.04 M PIPES [pH 6.8], 1 mM EDTA) containing 8 twice-recrystallized formamide, heated at 65°C for 5 min and then rapidly transferred to a water bath at 48°C for 4 hr. Following incubation, the reaction was diluted into S1 nuclease buffer (0.25 M NaCl, 0.03 M sodium acetate [pH 5.0], 0.001 M ZnSO$_4$) and treated with 300–3000 U S1 nuclease (Miles) at 37°C for 45 min. After extraction with phenol and precipitation with ethanol, the dried samples were dissolved in sequence gel buffer containing 80% deionized formamide, heated to 90°C for 1 min and loaded immediately onto 8% sequence gels. In experiments with shorter single-stranded DNA fragments, formamide was omitted from the hybridization buffer and the samples were not heated to 65°C prior to hybridization. In most instances control experiments were carried out in parallel with RNA from uninfected plants.

cDNA Synthesis

For primer extension, 1–3 μl gradient-purified 35S or 8S RNA fraction from healthy or virus-infected plants was incubated at 60°C for 20 min with a single-stranded 5'-^{32}P-labeled CaMV DNA fragment in a total volume of 20 μl of 0.6 M KCl, 50 mM Tris–HCl (pH 8). The mixture was then diluted to a final volume of 150 μl with addition of stock solutions of the following ingredients to give the indicated final concentrations: 50 mM Tris–HCl (pH 8), 15 mM MgCl$_2$, 10 mM dithiothreitol, 30 μg/ml actinomycin D and 1 mM of each deoxynucleotide triphosphate. The mixture was then incubated for 1 hr at 42°C with 12 U reverse transcriptase. The cDNA was phenol-extracted, dialyzed to eliminate the bulk of the nucleotide triphosphates and ethanol-precipitated with carrier tRNA. The precipitate was collected by low-speed centrifugation, dried and dissolved in 5 μl sequence gel buffer plus 80% deionized formamide. After heat denaturation (90°C for 2 min), the sample was loaded onto a 5% polyacrylamide sequence gel along with ^{32}P-labeled single-stranded DNA length markers. High specific activity cDNA for use as a probe was synthesized in the same fashion, except that 2 μg oligo(dT)$_{12-18}$ (P-L Biochemicals) was used as primer and the synthesis was carried out in the presence of 50 μCi α-^{32}P-dCTP with the concentration of nonradioactive dCTP reduced to 2 μM. After dialysis, the cDNA was heat-denatured in a boiling water bath and introduced directly into the hybridization buffer (Alwine et al., 1977).

References

Al Ani, R., Pfeiffer, P., Whitechurch, O., Lesot, A., Lebeurier, G. and Hirth, L. (1980). A virus-specified protein produced upon infection by cauliflower mosaic virus. Ann. Virol. 131E, 33–53.

Alwine, J. C., Kemp, D. J. and Stark, G. R. (1977). Method for detection of specific RNAs in agarose gels by transfer to diazobenzyloxymethyl-paper and hybridization with DNA probes. Proc. Nat. Acad. Sci. USA 74, 5350–5354.

Balàzs, E., Guilley, H., Jonard, G. and Richards, K. (1982). Nucleotide sequence of an altered virulence strain of cauliflower mosaic virus. Gene, in press.

Benz, E. W., Wydro, M. R., Nadal-Ginard, B. and Dina, D. (1980). Moloney murine sarcoma proviral DNA is a transcriptional unit. Nature 288, 665–669.

Berk, A. J. and Sharp, P. A. (1978). Spliced early mRNAs of simian virus 40. Proc. Nat. Acad. Sci. USA 75, 1274–1278.

Breathnach, R. and Chambon, P. (1981). Organization and expression of eucaryotic split genes coding for proteins. Ann. Rev. Biochem. 50, 349–383.

Cooke, R., Durand, R., Teissere, M., Penon, P. and Richard, J. (1981). Characterization of heparin-resistant complex formation and RNA synthesis by wheat germ polymerases I, II and III, in vitro on cauliflower mosaic virus DNA. Biochem. Biophys. Res. Commun. 98, 36–42.

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.

Covey, S. N., Lomonossoff, G. P. and Hull, R. (1981). Characterisation of cauliflower mosaic virus DNA sequences which encode major polyadenylated transcripts. Nucl. Acids Res. 9, 6735–6747.

Dudley, R. K., Odell, J. T. and Howell, S. H. (1982). Structure and 5'-termini of the large and 19S 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 mosaic virus DNA. Cell 21, 285–294.

Gardner, R. C., Howarth, A. J., Hahn, P., Brown-Leudi, M., Shepherd, R. J. and Messing, J. (1981). The complete nucleotide sequence of an infectious clone of cauliflower mosaic virus by M13mp7 shot gun sequencing. Nucl. Acids Res. 9, 2871–2887.

Howell, S. H. (1981). Ultraviolet mapping of RNA transcripts encoded by the cauliflower mosaic virus genome. Virology 112, 488–495.

Howell, S. H. and Hull, R. (1978). Replication of cauliflower mosaic virus and transcription of its genome in turnip leaf protoplasts. Virology 86, 468–481.

Lewis, M. K. and Burgess, R. R. (1980). Transcription of simian virus 40 DNA by wheat germ polymerase II. Priming of RNA synthesis by the 3' hydroxyl end of DNA at single-strand nicks. J. Biol. Chem. 255, 4928–4936.

Manley, J. L., Fire, A., Cano, A., Sharp, P. A. and Gefter, M. L. (1980). DNA-dependent transcription of adenovirus genes in a soluble whole-cell extract. Proc. Nat. Acad. Sci. USA 77, 3855–3859.

McMaster, G. K. and Carmichael, G. C. (1977). Analysis of single- and double-stranded nucleic acids on polyacrylamide and agarose gels by using glyoxal and acridine orange. Proc. Nat. Acad. Sci. USA 74, 4835–4838.

Ménissier, J., Lebeurier, G. and Hirth, L. (1982). Free cauliflower mosaic virus supercoiled DNA in infected plants. Virology 117, 322–328.

Odell, J. T. and Howell, S. H. (1980). The identification, mapping and characterization of mRNA for P66, a cauliflower mosaic virus-coded protein. Virology *102*, 349–359.

Odell, J. T., Dudley, R. K. and Howell, S. H. (1981). Structure of the 19S RNA transcript encoded by the cauliflower mosaic virus genome. Virology *111*, 377–385.

Proudfoot, N. J. and Brownlee, G. G. (1976). 3' noncoding region sequences in eucaryotic messenger RNA. Nature *263*, 211–214.

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

Shepherd, R. J. (1979). DNA plant viruses. Ann. Rev. Plant Physiol. *30*, 405–423.

Smith, D. R. and Calvo, J. M. (1980). Nucleotide sequence of the *E. coli* gene coding for dihydrofolate reductase. Nucl. Acids Res. *8*, 2255–2274.

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

Wasylyk, B., Kédinger, C., Corden, J., Brison, O. and Chambon, P. (1980). Specific *in vitro* initiation of transcription on conalbumin and ovalbumin genes and comparison with adenovirus-2 early and late genes. Nature *285*, 367–373.

Weaver, R. F. and Weissmann, C. (1979). Mapping of RNA by a modification of the Berk-Sharp procedure. The 5' termini of 15S β-globin mRNA precursor and mature 10S β-globin mRNA have identical map coordinates. Nucl. Acids Res. *7*, 1175–1193.