VIROLOGY 177, 717–726 (1990)

Recombination Sites in Cauliflower Mosaic Virus DNAs: Implications for Mechanisms of Recombination[1]

VALERIE RAY VADEN[2] AND ULRICH MELCHER[3]

Department of Biochemistry, Oklahoma State University, Stillwater, Oklahoma 74078-0454

Received December 14, 1989; accepted April 11, 1990

Pairs of mutant cauliflower mosaic virus (CaMV) DNAs readily recombine in plants. Five plasmid clones of CaMV DNAs resulting from infection of turnips with pairs of mutant DNAs from different isolates were obtained. Restriction analysis and nucleotide sequencing identified deletions in two cloned recombinants, VR1249 and VR244B. The sequence missing in the former was consistent with its deletion by splicing of an RNA intermediate. These DNAs were not infectious in turnips. VR1243, VR244A, and VR246 induced in turnips disease symptoms that were mixtures of those produced by the parental isolates. Junctions between sequences of the parental isolates were identified by restriction fragment analysis. Three cloned chimeras resulted from multiple recombination events. Nucleotide sequencing identified more precisely the junctions in the five cloned chimeras and in three chimeras previously characterized. Consistent with a model in which reverse transcription plays a major role in generating recombinants, six chimeras had junctions at or near the site for initiation of DNA (−) strand synthesis, three had junctions near the initiation site of 35 S RNA transcription, and one junction was found near the initiation site of 19 S mRNA transcription. Junctions were also found in regions not bearing any obvious relation to DNA (−) strand synthesis by reverse transcription, suggesting that recombination of double-stranded DNAs may also generate CaMV DNA recombinants.

INTRODUCTION

Cauliflower mosaic virus (CaMV) is a 50-nm isometric retroid virus of crucifer plants. The genome is an 8-kbp double-stranded circular DNA with three single strand discontinuities (gaps) (Covey, 1985). Two major RNA transcripts and six large open reading frames are encoded by the DNA. Virion DNA is synthesized in the cytoplasm of infected turnip leaf cells by a process that includes reverse transcription (Mason *et al.*, 1987) of the 35 S RNA transcript. Inoculation of turnip plants with two mutant CaMV DNAs, neither capable of establishing infection by itself, frequently results in infection (Howell *et al.*, 1981; Lebeurier *et al.*, 1982; Daubert *et al.*, 1983; Choe *et al.*, 1985). Neither of the mutations is detected in the DNA of progeny virions indicating that infection is a result of recombination rather than complementation between the mutant DNAs. Three mechanisms for the generation of recombinant CaMV have been proposed: ligation of the linear DNAs provided in the inoculum (Geldreich *et al.*, 1986); double-stranded DNA recombination in the nucleus (Choe *et al.*, 1985); and template switching by reverse transcriptase during synthesis of DNA (−) strands (Grimsley *et al.*, 1986b).

That isolates of CaMV differ by about 5% in nucleotide sequence (Balazs *et al.*, 1982) makes it possible to analyze the parental origin of DNA sequences recombined *in vivo* (Dixon *et al.*, 1986; Stratford and Covey, 1989; Geldreich *et al.*, 1986; Grimsley *et al.*, 1986b; Walden and Howell, 1982; Choe *et al.*, 1985). To identify possible mechanisms of recombination we have analyzed *in vivo* recombinants by nucleotide sequencing. The analysis is consistent with roles for both reverse transcription and double-stranded DNA recombination in the generation of CaMV DNA recombinants.

MATERIALS AND METHODS

Plasmids used

Infectious CaMV DNAs of the CabbS, CM4-184, NY8153, and W isolates were obtained from plasmids Ca37 (Lebeurier *et al.*, 1980), pLW414 (Howell *et al.*, 1980), pCMS31 (Armour *et al.*, 1983), and pLW303X (Walden and Howell, 1982; Choe *et al.*, 1985), respectively. The following previously described plasmids containing mutant CaMV DNA were used (with the open reading frame (ORF) affected by the mutation indicated in parentheses): CabbS mutants pUM124 (ORF I), pUM24 (ORF VI) (Choe *et al.*, 1985); CM4-184 mutants pLW214 (ORF I) and pLW76 (ORF I) (Howell *et al.*,

[1] Sequence data from this article have been deposited with the EMBL/GenBank Data Libraries under Accession Nos. M32808, M32809, M32810, M32811, M32812, M32813, M32814, and M32815.

[2] Present address: Department of Biochemistry, Biophysics and Genetics, University of Colorado Health Sciences Center, Denver CO 80262.

[3] To whom requests for reprints should be addressed.

0042-6822/90 $3.00

1981; Melcher *et al.*, 1986b); NY8153 mutant pDLS19 (ORF IV) (Melcher *et al.*, 1986b); and W mutant pIC23 (ORF VI) (Choe *et al.*, 1985). Plasmids pIC141, pIC143, and pIC148, containing CaMV DNA recombined from CabbS ORF III mutant pUM41 and W ORF VI mutant pIC11 were previously described (Choe *et al.*, 1985). Limited nucleotide sequencing established that pLW76 (Howell *et al.*, 1981) differed from its parent pLW414 by a 432-bp in-frame deletion in ORF I (from positions 777 to 1208). Nucleotide sequence position numbers used in this article are those of the CabbS isolate (Franck *et al.*, 1980) irrespective of the isolate involved. All experiments involving recombinant DNA-containing organisms were performed in compliance with relevant NIH guidelines.

Growth and inoculation

Turnip plants (*Brassica rapa* L., cv. Just Right) were grown and inoculated as previously described (Sun *et al.*, 1988) with plasmid DNAs at a total concentration of 17 mg/liter. Prior to inoculation, plasmid DNAs were digested with a restriction enzyme to release linear CaMV DNA from the plasmid vector (*Sal*I for all except pLW303X and pIC23, for which *Xho*I was used). Plants were checked periodically over a 30-day postinoculation period for symptoms of CaMV infection. CaMV DNA was isolated from infected plants as described by Gardner and Shepherd (1980). CaMV DNA was cloned in pBR322 via common *Sal*I sites as described previously (Armour *et al.*, 1983). Plasmid DNAs were isolated from bacterial cultures by alkaline sodium dodecyl sulfate lysis and CsCl–ethidium bromide centrifugation. DNA samples were digested with restriction enzymes according to instructions of the suppliers (BRL, Boehringer-Mannheim, Promega, and New England Biolabs). Digests were analyzed by agarose or polyacrylamide gel electrophoresis as appropriate to the size range of fragments expected. Digests of recombinant CaMV plasmid DNAs were compared side by side with digests of the appropriate parental plasmid DNAs.

Nucleotide sequencing

Both chemical cleavage (Maxam and Gilbert, 1980) and chain termination (Sanger, 1981) methods were used to determine nucleotide sequences of DNA. Some nucleotide sequences of the W isolate were determined from shotgun inserts of *Sau*3AI fragments of pLW303S (Walden and Howell, 1982) in pGEM3Z (Promega) using T7 and SP6 primers (Promega). For nucleotide sequencing by chain termination of selected regions of CaMV DNA borne in plasmid DNAs, appropriate primers were synthesized (Applied Biosystems).

TABLE 1

INFECTIVITY OF COMBINATIONS OF MUTANT CaMV DNAs IN TURNIPS

Mutant pair	No. infected plants/No. plants inoculated	Cloned recombinants[a]
pUM124 pDLS19	1/6	pVR1249 (–)
pUM124 pIC23	2/62	pVR1243 (+)
pUM24 pLW214	2/6	pVR244A (–) pVR244B (–)
pUM24 pLW76	1/11	pVR246 (+)
pDLS19 pLW76	3/12	—

[a] Names of plasmids resulting from cloning of CaMV DNA from infected plants and orientations of the CaMV DNA inserts relative to Ca 37 (+, same; –, opposite) as determined by restriction analysis with *Pst*I.

The primers were used to prime synthesis from *Sal*I-digested or alkali-denatured supercoiled (Chen and Seeburg, 1985) plasmid DNAs. The Klenow fragment of DNA polymerase I or Sequenase (USB; Tabor and Richardson, 1987) was used to polymerize nucleotides. Sequences were manipulated and restriction fragment lengths predicted with the "DNA Strider" program (Marck, 1988).

RESULTS

Creation and characterization of new recombinant CaMV DNAs

Some plants became diseased when inoculated with certain combinations of noninfectious mutant DNAs (Table 1). CaMV DNA recovered from some of the diseased plants was cloned into pBR322 (Table 1). After linearization by *Sal*I digestion, three of the five recombinant CaMV DNAs (pVR1243, pVR246, and pVR244A) infected turnips. Plants infected with VR1243 had stunted leaves with mottled pigmentation, as did companion plants infected with either parental isolate, CabbS or W. Several leaves were highly rugose with pronounced vein banding (characteristics seen in CabbS-infected, but not W-infected companions). Vein clearing, prominent on leaves infected with CM4-184 or VR246, was not noted on leaves of plants infected with VR244A or CabbS. Stunted leaves, a characteristic of VR246 and CabbS infections, were not found on plants infected with CM4-184 or VR244A.

Noninfectivity due to deletion

CaMV DNAs in pVR1249 and pVR244B were not infectious. Restriction analysis of pVR1249 (Table 2) revealed that *Pst*I, *Hin*dIII, *Msp*I, *Eco*RI, and *Bgl*I fragments containing the ORF I region of CaMV DNA were

TABLE 2

RESTRICTION SITE ALLELES IN RECOMBINANT CaMV DNAs

Position	Enzyme	Alleles in Recombinant DNAs[a]				
		VR1243	VR1249	VR246	VR244A	VR244B
15	*Hin*fI	CS (+)	[b](–)	[b](–)	[b](–)	[b](–)
197	*Alu*I	[b](–)	x	CS (–)	CS (–)	CS (–)
830	*Hin*fI	[b](+)	del	CS (0)	CS (+)	CS (+)
904	*Acc*I	W (0)	del	CS (0)	[b](+)	[b](+)
1170	*Hae*III	[b](+)	del	CS (0)	[b](+)	[b](+)
1218	*Hin*fI	W (–)	del	CS (+)	CS (+)	CS (+)
1513	*Hin*dIII	[b](+)	[b](+)	CS (+)	CS (+)	CS (+)
1571	*Hin*fI	[b](+)	x	CS (+)	CS (+)	CS (+)
1643	*Taq*I	[b](+)	[b](+)	CS (+)	CS (+)	CS (+)
1781	*Hin*fI	[b](+)	x	CS (+)	CS (+)	CS (+)
1783	*Cla*I	[b](+)	x	CS (+)	CS (+)	CS (+)
1784	*Taq*I	[b](+)	x	CS (+)	CS (+)	CS (+)
2236	*Msp*I	CS (–)	x	CM (+)	CS (–)	CM (+)
2367	*Taq*I	CS (–)	x	CM (+)	CS (–)	CM (+)
2680	*Cla*I	[b](–)	CS (–)	CM (+)	CS (–)	CM (+)
3092	*Hha*I	x	x	CM (+)	CS (–)	CM (+)
3231	*Pst*I	CS (–)	CS (–)	CM (+)	CS (–)	CS (–)
3249	*Hin*dIII	[b](+)	CS (0)	[b](+)	[b](+)	[b](+)
3360	*Msp*I	CS (–)	x	CM (+)	CS (–)	CS (–)
3414	*Bgl*I	[b](+)	CS (0)	[b](+)	[b](+)	[b](+)
3426	*Pst*I	CS (–)	CS (–)	CM (+)	CS (–)	CS (–)
3635	*Hin*fI	x	x	CM (+)	CS (–)	CS (–)
3772	*Hin*dIII	CS (–)	CS (–)	CM (+)	CS (–)	CS (–)
3928	*Eco*RI	CS (+)	CS (+)	CM (–)	CS (+)	CS (+)
4133	*Hin*fI	[b](–)	x	CM (+)	CS (–)	del
4149	*Msp*I	[b](–)	CS (–)	CM (+)	CS (–)	CS (–)
4660	*Msp*I	[b](–)	CS (–)	CS (–)	CS (–)	CS (–)
4752	*Hae*III	CS (–)	x	CS (+)	CS (+)	CS (+)
4766	*Alu*I	x	x	CS (+)	CS (+)	CS (+)
4921	*Hin*fI	x	x	CS (–)	CS (–)	CS (–)
5541	*Taq*I	CS (+)	x	CS (+)	CS (+)	CS (+)
5943	*Hae*III	CS (–)	x	CS (–)	CM (+)	CM (+)
6045	*Eco*RI	CS (+)	CS (+)	CM (–)	CM (–)	CM (–)
6068	*Hae*III	CS (+)	x	CM (–)	CM (–)	CM (–)
6280	*Hha*I	x	x	CM (–)	CM (–)	CM (–)
6330	*Acc*I	CS (+)	CS (+)	CM (0)	CM (0)	CM (0)
6411	*Hae*III	CS (0)	x	[b](+)	[b](+)	[b](+)
6467	*Hae*III	CS (–)	x	CM (+)	CM (+)	CM (+)
6509	*Msp*I	CS (–)	CS (–)	CM (+)	CM (+)	CM (+)
6654	*Hae*III	CS (–)	x	CM (+)	CM (+)	CM (+)
6655	*Bgl*I	[b](–)	[b](–)	CM (+)	CM (+)	CM (+)
7118	*Msp*I	CS (+)	CS (+)	CM (–)	CM (–)	CM (–)
7224	*Hae*III	CS (+)	CS (+)	CM (–)	CM (–)	CM (–)
7794	*Hin*fI	[b](+)	CS (+)	[b](+)	[b](+)	[b](+)
7980	*Cla*I	W (–)	[b](+)	[b](+)	[b](+)	[b](+)

Note. Loci not analyzed are indicated by "x" and loci absent due to deletion in the recombinant, by "del." Symbols in parentheses indicate whether the allele is cut (+) or not cut (–) by the indicated restriction enzyme or whether it was absent from one parent due to mutation (0).

[a] Parental isolate providing the detected allele: CS, CabbS; W, W; CM, CM4-184; NY, NY8153.

[b] Both parents.

800 to 900 bp smaller than expected. The deletion was localized by restriction analysis to between positions 124 and 1514 because of the presence of an *Hpa*II fragment of 1030 bp (positions 7114 to 124) and *Hin*dIII fragments of 2.8 kbp (positions 5850 to 1513, including deletion) and 436 bp (positions 1513 to 1949). When restricted, the other plasmid DNA containing noninfectious CaMV DNA, pVR244B, produced *Acc*I, *Cla*I, and

*Hae*III fragments that were 140 ± 100 bp shorter than the respective expected sizes of 1074, 1504, and 1430 bp. An expected *Hin*fI fragment of 392 bp was replaced in pVR244B by a fragment of approximately 900 bp, suggesting a small deletion that includes a *Hin*fI site at 4110 (Table 2).

The positions of the deletions in pVR1249 and pVR244B were defined more precisely by nucleotide sequencing of the regions surrounding the deletions and comparison of the sequence with those of the parental CabbS and NY8153 (Fig. 1) isolates. The deletion in pVR244B extended 47 bp in ORF V from position 4096 ± 1 to 4142 ± 1. The region surrounding and including the deletion does not contain large inverted complementary repeats or RNA splicing signals. The deleted sequence is flanked by an imperfect direct repeat of 13 residues. The deletion in pVR1249 extended 851 bp from position 657 to 1508. With the exception of two nucleotides (TC, positions 652–653) not found in either parent, the sequences on the ORF I side of the deletion were derived from the NY8153 isolate while those on the ORF III side were from the CabbS isolate. The region deleted has sequences at the 5′ and 3′ end (of the putative transcript) that correspond to previously identified (Hirochika *et al.,* 1985) RNA splice donor and acceptor sites.

Restriction analysis of recombinants

Sites for restriction by the enzymes *Pst*I, *Hha*I, *Acc*I, *Bgl*I, *Cla*I, *Eco*RI, *Hin*dIII, *Alu*I, *Hae*III, *Hin*fI, *Msp*I, and *Taq*I which differ between parental isolates and whose presence or absence can be readily scored by gel electrophoretic analysis of restriction fragments were identified (Table 2). Sites for CM4-184 and CabbS isolates were those predicted from the nucleotide sequences (Dixon *et al.,* 1986; Franck *et al.,* 1980), while those for NY8153 and W were obtained from restriction mapping (Gardner *et al.,* 1980; Vaden and Melcher, unpublished) and, where available, from nucleotide sequences (Melcher *et al.,* unpublished). Analysis of the diagnostic restriction sites in the cloned recombinant CaMV DNAs (Table 2) revealed junctions between CM4-184 and W sequences in pVR244A between positions 5541 to 5943 and 7224 to 197. In addition to the pVR244A crossover regions, pVR244B contained CM4-184–CabbS junctions between positions 1784 to 2236 and 3092 to 3231. Four junctions between sequences of parental isolates were found in pVR246 (between positions 7224 to 197, 1784 to 2236, 4149 to 4660, and 5943 to 6045) and in pVR1243 (between positions 15 to 904, 1218 to 2236, 7224 to 7980, and 7980 to 15). No junctions were detected by restriction in pVR1249.

Nucleotide sequence near DNA (−) strand start sites

The nucleotide sequence near DNA (−) strand start sites was determined for each of the recombinants, for three cloned recombinant CaMV DNAs previously characterized by restriction pIC141, pIC143, pIC148 (Choe *et al.,* 1985), and for the DNA of the W (Fig. 2) and NY8153 (Fig. 1) isolates. For VR246, VR1249, VR1243, IC141, IC143, and IC148, the sequences were identical to those of one of the two parents on one side of the start site, and identical to those of the other parent on the other side of that site (Fig. 3). The region sequenced in VR244A and VR244B contained no residues diagnostic of the CM4-184 parent. The VR1249 sequence suggests the presence of additional junctions not detected by restriction analysis: that at the DNA (−) strand start site (Fig. 3) and a junction between positions 98 to 2680.

Nucleotide sequences near other crossover sites

The positions of other junctions between parental sequences were also determined by nucleotide sequencing. Junctions in pVR1243, pVR244B, and pVR244A (Fig. 4) spanned the site of initiation of 35 S RNA synthesis. Junctions in four of the eight recombined DNAs (IC143, IC148, VR244B; VR246) occurred in the 131 nucleotide region between positions 2101 to 2232 (Fig. 5). However, only the IC143 crossover could have occurred at precisely the same position as one of the others. In addition, junctions in VR1243 and IC141 occurred within 440 nucleotides of this region (Figs. 6B and 6C). A junction in VR246 between positions 4410 to 4435 occurred within 200 bp, but not at, one site of initiation of DNA (+) strand synthesis (Fig. 6F). A junction in VR244A and VR244B spanned the site of initiation of 19 S RNA synthesis (Fig. 6H). One junction in IC143 spanned the position of the *Sal*I site used to create one, but not the other, of the inoculum DNA molecules (Fig. 6G). The remaining junctions detected were scattered in other regions of the genome (Figs. 6A, 6D, 6E, and 6H): positions 342 to 494 (VR1243), 3019 to 3231 (VR244B), 3350 to 3354 (IC143), and 5947 to 6025 (VR246). Eight of the 11 junctions that could not be related easily to known features of the DNA sequence occurred near TATA or TAATA sequences either 5′ of the junction in the (+) strand or within the junction. The significance of these sequences is uncertain since there are 45 independent occurrences of TATA and TAATA in CabbS DNA.

The precision of the identification of these junctions ranged from 4 to 157 nucleotides (VR244B, Fig. 5 and VR244A, Fig. 6H, respectively) because of differing distances between diagnostic positions. In all cases, no

```
624
   ATTATCA GTGGACATTC ACGATGCCAC AGGTAAGGTA TATCTTCCTT TAATCACTAA AGAGGAGATA AATAAAA
                                                                              697

1474
   AATTAAT CATCTTAACA ACCTCAATGA GATTGTAGGA AGAAGCTTAC TCGGAATATG GAAGATCAAC TCATACTTCG GACTAAGCAA AGACCCTTCG 1570
GAGTCCAAAT CAAAAAACCC GTCAGTTTTT AATACTGCAA AAACCATTTT TAAGAGTGGG GGGGTTGATT ACTCGAGCCT AAAGGAAATA AAATCCCTTT 1673
TAGAAGCTCA AAACACTAGA ATTAAAAGTC TA
                                  1705

7945
   TGCTTGT ATTTTACCCT ATATACCCTA ATAACCCCTT ATCGATTTAA AGAAATAATC CGCATAAGCC CCCGCTTAAA AATTGGTATC AGAGCCATGA   16
ATCGGTTTAA AGACCAAACT CAAGAGGGTA AAACCTCATC AAAATACGAA AGAGTTCTTA ACTCTAAAGA TAAAAGATCT T
                                                                              97
```

FIG. 1. Nucleotide sequences of selected regions of NY8153 CaMV DNA. Numbering of positions is that of the CabbS isolate (Franck *et al.*, 1980).

interspersion of alleles was detected. The absence of allele interspersion near junctions in IC143 (Fig. 6G) and IC141 (Fig. 6C) conflicts with previous interpretations of restriction enzyme digestions of the respective plasmid DNAs (Choe *et al.*, 1985). Those interpretations were based on limited restriction and sequence analysis that suggested the cloned DNA resembled that of the CM1841 DNA. Nucleotide sequencing of the W isolate suggested that the cloned DNA resembled CabbB-JI most closely (see below). A reevaluation of the restriction analysis of IC141 and IC143 revealed that the observed restriction fragments were consistent with the W and CabbB-JI nucleotide sequences.

Parental DNA sequences

Since CM4-184 DNA appears to be a chimera, generated through template switching by the reverse transcriptase, between sequences of an unknown isolate and CM1841 DNA (Dixon *et al.*, 1986), we analyzed the distribution of nucleotide sequence differences in W and NY8153 DNA. W DNA in the 590 nucleotide large intergenic region between the initiation site of 35 S RNA transcription and the (−) strand start site (Fig. 2) differed in sequence from CabbB-JI (Stanley, unpublished), CabbS (Franck *et al.*, 1980), CM1841 (Gardner *et al.*, 1981), and D/H (Balazs *et al.*, 1982)

```
        7152
           CTGTCACTT CATCGAAAGG ACAGTAGAAA AGGAAGGTGG CTCCTACAAA TGCCATCATT GCGATAAAGG AAAGGCTATC ATTCAAGATG 7240
CCTCTGCCGA CAGTGGTCCC AAAGATGGAC CCCCACCCAC GAGGAGCATC GTGGAAAAAG AAGACGTTCC AACCACGTCT TCAAAGCAAG TGGATTGATG 7340
TGACATCTCC ACTGACGTAA GGGATGACGC ACAATCCCAC TATCCTTCGC AAGACCCTTC CTCTATATAA GCAAGCTCAT TTCATTTGGA GAGGACACGC 7440
TGAAATCACC AGTCTCTCTC TACAAATCTA TCTCTCTCTC TATTCTCCAT AATAATGTGT GAGTAGTTCC CAGATAAGGG AATTAGGGTT CTTATAGGGT 7531
TTCGCTCATG TGTTGAGCAT ATAAGAAACC CTTAGTATGT ATTTGTATTT GTAAAATACT TCTATCAATA AAATTTCTAA TTCCTAAAAC CAAAATCCAG 7631
TACTAAAATC CAGATCTCCT AAAGTCCCTA TAGATCTTTG TCGTGAATAT AAACCAGACA CGAGACGACT AAACCTGGAG CCCAGACGCC GTTTGAAGCT 7731
AGAAGTACCG CTTAGGCAGG AGGCCGTTAG GGAAAAGATG CTAAGGCAGG GTTGGTTACG TTGACTCCCC CGTAGGTTTG GTTTAAATAT GATGAAGTGG 7831
ACGGAAGGAA GGAGGAAGAC AAGGAAGGAT AAGGTTGCAG GCCCTGTGCA AGGTAAGAAG ATGGAAATTT GATAGAGGTA CGTTACTATA CTTATACTAT 7931
ACACTAAGGG ATGCTTGTAT TTTACCCTAT ACCCCCTAAT AACCCCTTAT CGATCTAAAG AAATAATCCG CATAAGCCCC CGCTTAAAAA ATTGGTATCA    7
GAGCCATGAA TAGGTCTATG ACCAAAACTC AAGAGGATAA AACCTCACCA AAATACGAAA GAGTTCTTAA CTCTAAAGAT AAA
                                                                          90

         322
           CTCTAACGA GTATCCACAG AAAAAATAAC CTTCTGTGTT GAGATGGATT TGTATCCAGA AGAAAATACC CAAAGCGAGC AATCGCAGAA  410
TTCTGAAAAT AATATGCAAA TATTTAAATC AGAAAATTCG GATGGATTCT CCTCCGATCT AATGATCTCA AACGATCAAT TAAAGAATAT CTCTAAAACC  510
CAATTAACCT TGGAAAAAGA AAAGATATTC AA
                                  542
         1583
            AAAAACCC GTCAGTTTTT AATACTGCAA AAACCATTTT TAAGAGTGGG GGGGTTGATT ACTCGAGCCA ACTAAAAGAA ATAAAATCTC 1670
TTTTAGAATC TCAAAATACT AGAATTAAAA GTCTAGAAAA AGCAATTCAA TCCTTAGAGC GTAAGATTGA ACCAGAGCCC TTAACTAAAG AAGAAGTTAA 1770
AGAGCTTAAA GAATCGATTA ACTCGATCAA AGAAGGATTA AAGAATATTA TTGGCTAAAA TGGCTAATCT TAATCAAATC CAAAAAGAAG TCTCTGAAAT 1870
CCTCAGTGAC CAAAAATCCA TGAAAGCGGA TATAAAAGCT ATCTTAGAAC TATTAGGATC CCAAAATCCT ATTAAAGAAA GCTTAGAAGC TGTTGCAGCG 1970
AAAATCGTTA ATGACTTAAG CAAGCTCATC AATGATTGCC CCTGTAACAA AGAGATATTA GAAGCCTTAG GCAATCAGCC TAAAGAGCAA CTAATAGAAC 2070
AACCTAAAGA GAAAGGCAAA GGCCTTAATC TAGGAAAATA TTCTTACCCC AATTACGGAG TAGGAAATGA AGAATTAGGA TCCTCTGGAA ACCCTAAAGC 2170
TTTAACCTGG CCCTTCAAAG CTCCAGCAGG ATGGCCGAAT CAATTTTAGA CAGGACCATT AACCGGTTCT GGTATAATCT GGGAGAAGAT TGTCTCTCAG 2270
AAAGTCAATT TGACCTTATG ATAAGGTTAA TGGAAGAGTC CCTTGACGGG GACCAAATTA TTGATCTAAC CTCTCTACCT AGTGATAATT TGCAGGTCGA 2370
ACAGGTTATG ACAACTACCG AAGACTCGAT CTCGGAAGAA TCAGAATTCC TTCTAGCAAT AGGAGAAACA TCTGAAGACG AAAGCGATTC AGGAGAAGAA 2473
CCTGAATTCG AACAAGTTCG AATGGATCGA ACAGGAGGAA CGGAGATTCC CAAAGAAGAA GATGGTGAAC CATCTAGACA CAATGAGAGA AAGAGAAAGA 2576
CCCCGGAAGA CCGGTACTTT CCAACTCAAC CAAAGACCAT TCCAGGACAA AAACAAACGT CTATGGGAAT GCTCAACATC GACTGCCAAA CCAATCGAAG 2676
AACCTTAATC GACGACTGGG CAGCAGAAAT CGGACTG
                                        2713

    3268
       ATG CAAGAAAACA TCCAAGAAGA AGTATCATAA GCGATACAAG AAGAAATATA AGGTCTATAA ACCTTATAAG AAGAAGAAGA AATTCCGATC 3363
CGGAAAATAC TTC
             3376

      4630
         A GAAGATCTTC TCTTCCTTCG ACTGCAAGTC AGGATTCTGG CAAGTTCTGC TAGATCAAGA ATCAAGACCT CTAACGGCAT TCACATGTCC 4720
CCAAGGTCAC TACGAATGGA ATGTGGTCCC TTTCGGCCTA AAGCAGGCTC CATCCATATT CCAAAGACAC ATGGACGAAG CATTTCGTGT GTTCAGAAAG 4820
TTCTGTTGCG TTTATGTCGA CGACATTCTC GTATTCAGTA ACAACGAAGA AGATCACCTA CTTCATGTAG CGATGATCTT ACAAAAGTGC AATCAACATG 4920
GAATCATCCT TTCCAAAAAG AAAGCACAAC TCTTCAAGAA GAAGATAAAC TTCCTTGGTC TAGAAATAGA TGAAGGAACA CATAAGCCTC AAGGACACAT 5020
CTTGGAACAC ATCAACAAGT TCCCAGATAC CCTTGAAGAC AAGAAGCAAC TTCAGAGATT CTTAGGCATA CTAACA
                                                                        5096
```

FIG. 2. Nucleotide sequences of selected regions of W CaMV DNA. Numbers refer to the nucleotide sequence positions of the CabbS isolate (Franck *et al.*, 1980).

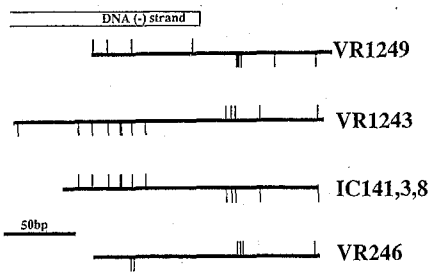

FIG. 3. Diagrams of nucleotide sequence alleles in the region of the 5′ end of virion DNA (−) strands (open bar) of CaMV recombinants. Bars above the horizontal sequence lines identify CabbS-specific nucleotides. Those below the lines identify nucleotides specific to the other parents as follows (nucleotide positions sequenced in parentheses): for VR1249 (7945 to 97), NY8153; for VR1243 (7887 to 90), W; for IC141, IC143, and IC148 (whose sequences were identical in the 7923 to 84 region), W; and for VR246 (7945 to 86), CM4-184.

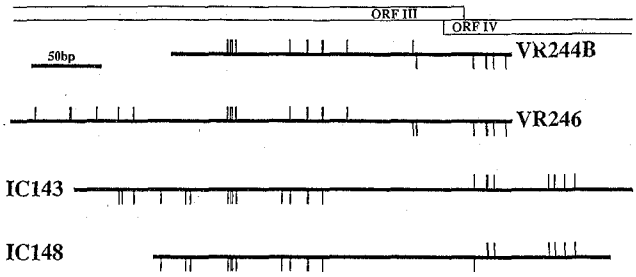

FIG. 5. Diagrams of nucleotide sequence alleles of recombinant CaMV DNAs in regions of crossover near the ORF III-IV (open bar) boundary. Bars above the horizontal lines identify CabbS-specific nucleotides. Those below the lines identify nucleotides specific to the other parents as follows (nucleotide positions sequenced in parentheses): for VR244B (2003 to 2250), CM4-184; for VR246 (1880 to 2250), CM4-184; for IC143 (1927 to 2340), W; and for IC148 (1986 to 2322), W.

DNAs at 18, 14, 17, and 23 positions, respectively. In 2258 positions in other regions of CaMV DNA (Fig. 2), the W sequence differed from that of CabbB-JI in only 5 positions, whereas there were 99, 91, and 121 differences when the W sequences were compared with those of CabbS, CM1841, and D/H isolates, respectively. The high similarity of CabbB-JI and W sequences was evident in both regions flanking the large intergenic region. Restriction fragments of pLW303X were inconsistent (Choe *et al.*, 1985) with one published restriction map of CabbB-JI DNA (Hull, 1980), but were consistent, except for *Pst*I (Table 2) and *Pvu*II fragments, with another (Delseny and Hull, 1983). Restriction of pLW303X (Choe *et al.*, 1985) with *Pvu*II revealed a single site in the CaMV DNA. Nucleotide sequencing established that the site was due to a C to G transversion at position 6325 relative to the CabbB-JI sequence (Stanley, unpublished). All other restriction digests analyzed were consistent with W DNA being a chimera of a variant CabbB-JI sequence (as defined by the nucleotide sequence) and a sequence in the large intergenic region of another isolate.

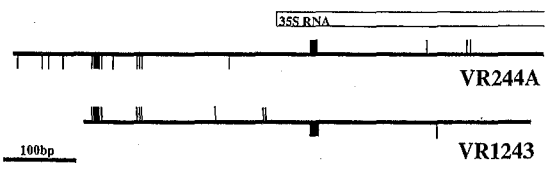

FIG. 4. Diagrams of nucleotide sequence alleles around the initiation site for 35 S RNA (open bar) transcription of recombinant CaMV DNAs VR244A (the sequence of VR244B in this region was identical) and VR1243. Bars above the horizontal lines identify CabbS-specific nucleotides. Those below the lines identify nucleotides specific to the other parents as follows: for VR244A (7051 to 7832), CM4-184; and for VR1243 (7152 to 7815), W.

The NY8153 sequence differed from CabbB-JI, CabbS, CM1841, and D/H DNA sequences in the 468 positions of the regions shown in Fig. 1 by 24, 20, 22, and 35 residues, respectively. In the region on the ORF I side of the DNA (−) strand start site, the large intergenic region, and the fragments of ORF I and ORF II, none of the four sequences compared to the NY8153 sequence had significantly fewer differences relative to NY8153 than did any of the other three suggesting that NY8153 is not closely related to the sequenced isolates. The CM4-184 sequence represented by the plasmid clone pLW414 differed slightly from the sequence of CM4-184 DNA reported by Dixon *et al.* (1986). In VR246, pLW76 (the CM4-184 parent of VR246), and pLW414 (from which pLW76 was constructed) residue 4323 was an A instead of a G found in published sequences of both CabbS and CM4-184 DNAs. Residue 5754 of pVR246 and pLW414 was an A instead of the C reported for CM4-184 DNA. Further, pLW414 lacked an *Hpa*I site predicted from the sequence to occur at position 5983. These differences suggest that the CM4-184 DNA cloned in pLW414 differs slightly from that previously sequenced (Dixon *et al.*, 1986).

Restriction fragment patterns for digestion of pVR1243, pVR246, pVR244A, and pVR244B by the enzymes used were predicted using the assumption that all sequence junctions had been located by the above analysis. The predicted pattern agreed with the pattern observed except for the *Taq*I digest of pVR246 and pVR244B. The exception was due to the absence of a site predicted at nucleotide 2964. The *Taq*I site at the equivalent position was also missing in pLW76, the parent of pVR246, thus identifying a further difference between the CM4-184 cloned in pLW414 and that sequenced by Dixon *et al.* (1986).

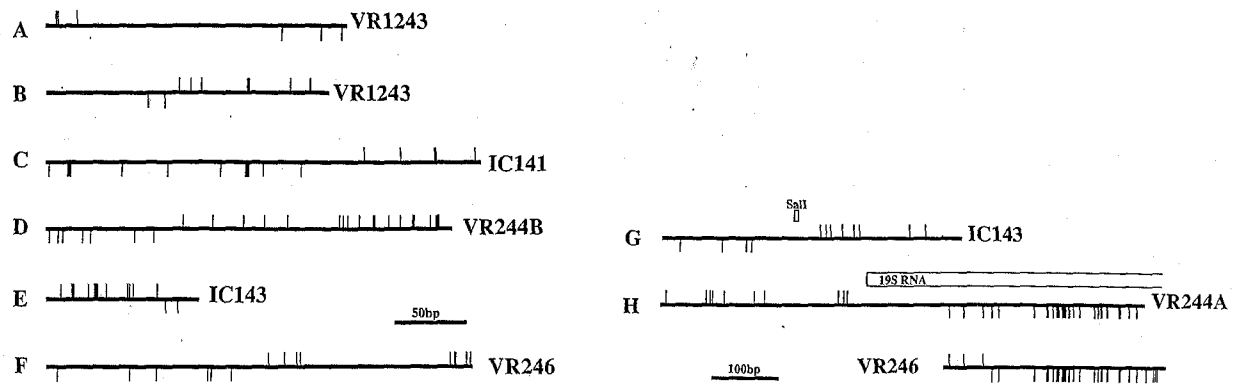

FIG. 6. Diagrams of nucleotide sequence alleles of recombinant CaMV DNAs in regions of crossover. Bars above the horizontal lines identify CabbS-specific nucleotides. Those below the lines identify nucleotides specific to the other parents (nucleotide positions sequenced in parentheses). Scale bar of 50 bp applies to A–F, of 100 bp to G and H. (A) VR1243 (322 to 542) compared with its parents, CabbS and W, near the start of ORF I. (B) VR1243 (1583 to 1780) compared with its parents, CabbS and W, in ORF II. (C) IC141 (2397 to 2713) compared with its parents, CabbS and W, in ORF IV. (D) VR244B (3019 to 3312) compared with its parents, CM4-184 and CabbS, in ORF IV. (E) IC143 (3268 to 3376) compared with its parents, CabbS and W, in ORF IV. (F) VR246 (4274 to 4583) compared with its parents, CabbS and CM4-184, in ORF V. (G) IC143 (4630 to 5096) compared with its parents, CabbS and W, in ORF V. Open bar locates the *Sal*I site. (H) VR244A (identical to VR244B in 5450 to 6197 region) and VR246 (5889 to 6235) compared with their parents, CabbS and CM4-184, in the ORF V to ORF VI region. Open bar identifies 5′ end of 19 S RNA.

DISCUSSION

The parental origins of DNA sequences in the eight recombinant CaMV DNAs analyzed by restriction enzyme digestion and nucleotide sequencing are summarized in Fig. 7. VR246 and VR244A DNAs, both recombinants between CabbS and CM4-184 isolates differ in the large intergenic region, in the ORF IV-ORF V region (positions 2177 to 4409), and in three nucleotides at the beginning of ORF VI. The ORF VI differences result in two amino acid substitutions in the predicted amino acid sequence of the CabbS inclusion body protein: serine to proline and glutamate to valine.

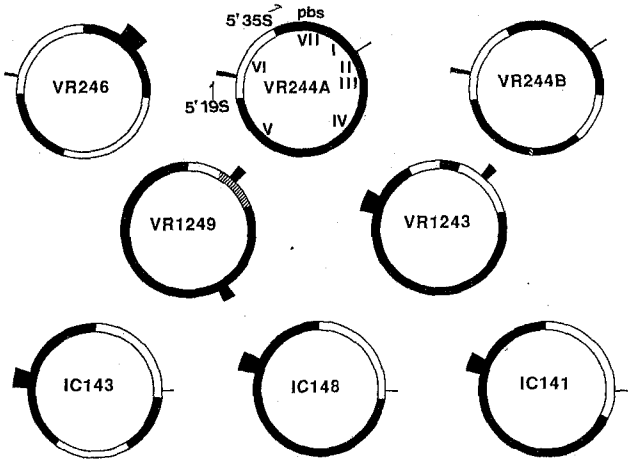

FIG. 7. Summary of structures of the DNA of eight CaMV recombinants. Sequences from CabbS parents are shown by filled arcs with sequence from the other parent shown by open arcs. Striped arcs represent regions deleted in the recovered recombinants. Sectors extending out from the circles show the positions and extent of the inactivating mutations in the parental DNAs. The positions of the 5′ ends of 19 S and 35 S RNAs and of the primer binding site are indicated on the VR244A map. Roman numerals show approximate locations of open reading frames.

That VR246 and CabbS stunted leaves, while VR244A and CM4-184 did not, suggests that leaf stunting is associated with the CabbS DNA region unique to VR246, the beginning of ORF VI. ORF VI has been previously implicated in leaf pigmentation patterns (Baughman *et al.*, 1988), systemic spread of infections, and local lesion types (Schoelz *et al.*, 1984, 1986). On the other hand, Stratford and Covey (1989), in analyzing symptoms induced by *in vivo* DNA recombinants of Baril and CabbB-JI isolates, concluded stunting was not caused by ORF VI. The stunting determinant of the CabbB isolate (Daubert *et al.*, 1984) is also not associated with the ORF VI region identified by VR244A and VR246. Stunting determinants may thus be located in entirely different regions in different isolates. This possibility is consistent with the suggestion that nucleic acid sequences play a role, other than via their protein coding function, in the production of some symptoms (Melcher, 1989). Confirmation of different locations of stunting determinants requires construction of appropriate *in vitro* recombinants since *in vivo* recombinants may contain small additional regions of exchange not detected by the restriction and nucleotide sequencing analysis performed. The necessity for construction of *in vitro* recombinants is underscored by

the presence of more than two sequence junctions in four of the eight recombinants studied (Fig. 7). The high frequency of multiple junction pairs is consistent with the observation of multiple recombinant forms in single plants (Fig. 7 and Choe *et al.*, 1985; Grimsley *et al.*, 1986b).

The ability of CaMV isolates to induce vein clearing in turnip leaves appears to be an active property (Melcher, 1989). VR246, but not VR244A, induced vein clearing characteristic of CM4-184, suggesting that vein clearing ability may be conveyed by the ORF IV-ORF V region and the large intergenic region. A study of the ability of mixtures of *in vivo* generated recombinants to cause vein clearing (Daubert *et al.*, 1984) suggested that two or more regions of the CaMV genome may be necessary for vein clearing. The large intergenic region of CM4-184 is almost identical in sequence to that of CabbS (Dixon *et al.*, 1986), making it unlikely that differences in this region contribute to symptom differences. However, isolate W does not induce vein clearing (Melcher, 1989) while CabbB-JI, which differs from W primarily in the large intergenic region, does (Stratford *et al.*, 1988).

One noninfectious chimeric CaMV DNA, pVR1249, was missing a nucleotide sequence whose ends correspond to splice donor and acceptor sequences. The possibility of a splicing-generated deletion is consistent with other observations suggesting reverse transcription of spliced CaMV RNAs (Melcher *et al.*, 1986b; Hohn *et al.*, 1986; Hirochika *et al.*, 1985). The splice donor site for the deletion observed in the DNA of the S isolate (Hirochika *et al.*, 1985) is not present in pVR1249 due to the substitution (of unknown origin) of TC for GT at positions 652 and 653. The VR1249 deletion corresponds to the deletion in an alternate S splicing product. The VR1249 deletion joined parental NY8153 and CabbS sequences making it impossible to determine whether the junction was formed by splicing *in trans* or whether the removed sequence contained a recombination junction.

The three suggested mechanisms for the generation of recombinant CaMV DNAs predict different crossover points in the resulting circular virion DNA. Template switching during (−) strand synthesis should result in a junction corresponding to the ends of the (−) strands, and a junction at the site of template switching. Ligation of linear inoculum DNAs should, in addition to the junction at the ligation site, produce a junction at the ends of the 35 S RNA transcript derived from the ligated molecule. Junctions at the 35 S RNA start site are also expected for any recombination event, including double-stranded DNA recombination, that occurs before reverse transcription.

Three recombinants (VR244A, VR244B, and VR1243) had junctions located at or near the start site of 35 S RNA transcription. That a second junction in VR244A and VR244B was at or near the start site of 19 S RNA transcription suggests that VR244A arose by two intermolecular template switches during reverse transcription. The first switch from the 5′ end of a CabbS 35 S RNA template was probably to a 19 S RNA of the mutant CM4-184 isolate. When the transcriptase reached the 5′ end of this template RNA (in the process incorporating a nonmutant copy of ORF VI into the DNA product) a second template switch likely occurred, this one to a CabbS-derived 35 S RNA. Since both 5′ and 3′ ends of the resulting VR244A DNA (−) strand would be derived from CabbS templates, no junction at the (−) strand start site is expected. A minor junction in progeny virions obtained by agroinfection (Grimsley *et al.*, 1986a) also mapped near the 5′ end of 19 S RNA (Grimsley *et al.*, 1986b) as did junctions in two recombinants analyzed by restriction (Stratford and Covey, 1989).

Intermolecular template switches may occur at 5′ ends of 35 S RNAs. Differences between W and CabbB-JI isolates are concentrated in the region between the 35 S RNA transcription initiation site and the DNA (−) strand start site (Figs. 3 and 4), as are differences between nucleotide sequences of CM1841 and CM4-184, the latter a DNA chimera of the CM1841 isolate and an isolate related to CabbS (Dixon *et al.*, 1986). W and CM4-184 were obtained without selection for recombination. Thus, intermolecular template switches at 5′ ends of 35 S RNAs may occur with appreciable frequency and play a role in evolution of caulimoviral genomes. On the other hand, only VR244A of the four recombinants with a single pair of crossovers (Fig. 7) and one of six recombinants analyzed by restriction by Stratford and Covey (1989) show evidence of an intermolecular template switch. Thus, the majority of template switches from the 5′ end to the 3′ end of 35 S RNA may be intramolecular.

The presence of multiple junctions hinders confident assignment of the processes that led to their production. VR244B, having been obtained from the same plant as VR244A, probably derived from VR244A by a further recombination event and a deletion. The origin of others is less clear. It is possible that the junction spanning the site of initiation of 35 S RNA synthesis in VR1243 resulted from intermolecular template switching, followed by replacement of the defective with a wild-type ORF I region. Alternatively, the product of a double-stranded DNA recombination event near position 1675 could have generated a transcript that when reverse transcribed resulted in a nondefective DNA with junctions at the start site of 35 S RNA synthesis

and near position 1675. A strand switch near position 400 during reverse transcription of the chimeric 35 S RNA could then generate VR1243.

Ligation of linear inoculum DNAs (Geldreich *et al.*, 1986) is not a feasible explanation for the creation of the observed recombinants between mutant CaMV DNAs. No identified crossover regions (Fig. 7) spanned the *Xho*I site used to linearize pIC11 and pIC23 for inoculation. Only one of the 24 junctions analyzed (in IC143) was at or near the position of the *Sal*I site used to produce all other inoculum DNAs. These *Sal*I ends could not have participated in creating IC143 directly by ligation of inoculum DNAs since the two inoculum DNAs had ends in different regions of the genome (ORF II and ORF V for *Xho*I and *Sal*I sites, respectively). If the positioning of the IC143 junction at the *Sal*I site was not fortuitous, the inoculum DNA must have participated in this recombination event as a double-stranded linear DNA. We suggest that the 3′ end of the (−) strand (or the 5′ end of the (+) strand) initiated recombination by invading a double-stranded IC11 DNA. Completion of the double-stranded DNA recombination followed by transcription and reverse transcription yielded a chimeric CaMV DNA in which DNA from the start site of 35 S RNA transcription clockwise to the *Sal*I site in ORF V were derived from IC11 DNA. Two further events, replacement of IC11 sequences by those of UM41 in the large intergenic region and in ORF IV, must have occurred to obtain the recombinant observed.

Six of the eight recombinants analyzed had junctions at or near the DNA (−) strand start site suggesting that these recombinants were generated by template switching during synthesis of DNA (−) strands. A third of the recombinants analyzed by restriction by Stratford and Covey (1989) could have had similar junctions. An alternative explanation for the preponderance of crossovers at gap 1 is that it is a source of free DNA ends suitable for initiating double-stranded DNA recombination. However, no junctions were associated with gap 2 or 3, equally likely sources of free DNA ends, and CaMV DNA in nuclei, where recombination is postulated to occur, is not gapped.

Splicing, 35 S and 19 S RNA 5′ end template switching, or DNA (−) strand initiation and the ends of linear inoculum molecules were not directly responsible for 11 of the 24 junctions analyzed (Fig. 7). Possibly, DNA (−) strand synthesis was initiated on minor RNA templates that had their 5′ ends within the junction regions. Minor RNA species that could be templates for reverse transcription have been isolated from infected turnip leaves (Condit *et al.*, 1983; Plant *et al.*, 1985). A particularly prominent minor species has its 5′ end near the ORF III–ORF IV boundary, where several junctions have been located (Fig. 5 and Stratford and Covey, 1989).

Minor differences in nucleotide sequence were found between two cloned representatives of CM4-184 (this work and Dixon *et al.*, 1986), between the nonintergenic regions of W and CabbB-JI (this work and Stanley, unpublished), and between the nonintergenic regions of CM1841 and CM4-184 (Dixon *et al.*, 1986). However, no evidence of sequence heterogeneity was encountered during the nucleotide sequencing of CabbS (Franck *et al.*, 1980) and D/H (Balazs *et al.*, 1982) viral DNAs. These results suggest that a CaMV isolate as obtained from a plant consists of a majority of one predominant nucleotide sequence and several small populations of minor sequence variants. The high error rates of reverse transcriptase-mediated DNA replications would lead one to expect a much greater heterogeneity of sequence. It is possible that the frequent exchange of segments by template switching and DNA recombination observed here couple with the strong ability of one sequence variant to dominate over another during multiplication in plants (Melcher *et al.*, 1986a; Zhang and Melcher, 1989) to limit CaMV sequence heterogeneity and rate of evolution. The possibility is supported by the observation that the divergence from a common ancestor of the amino acid sequences of caulimoviral ORF I proteins is significantly lower than that of corresponding proteins of RNA viruses (Melcher, 1990).

ACKNOWLEDGMENTS

We thank A. Williams and M. Schatz for technical assistance, K. Hooper for assistance in nucleotide sequencing, and J. Stanley for providing the CabbB-JI DNA sequence prior to its publication. This work was supported by the Herman Frasch Foundation and the Oklahoma Agricultural Experiment Station of which this is Journal Article J-5722.

REFERENCES

ARMOUR, S. L., MELCHER, U., PIRONE, T. P., LYTTLE, D. J., and ESSENBERG, R. C. (1983). Helper component for aphid transmission encoded by region II of cauliflower mosaic virus DNA. *Virology* **129**, 25–30.

BALAZS, E., GUILLEY, H., JONARD, G., and RICHARDS, K. (1982). Nucleotide sequence of DNA from an altered-virulence isolate D/H of the cauliflower mosaic virus. *Gene* **19**, 239–249.

BAUGHMAN, G. A., JACOBS, J. D., and HOWELL, S. H. (1988). Cauliflower mosaic virus gene VI produces a symptomatic phenotype in transgenic tobacco plants. *Proc. Natl. Acad. Sci. USA* **85**, 733–737.

CHEN, E. Y., and SEEBURG, P. H. (1985). Supercoil sequencing: A fast and simple method for sequencing plasmid DNA. *DNA* **4**, 165–170.

CHOE, I. S., MELCHER, U., RICHARDS, K., LEBEURIER, G., and ESSENBERG, R. C. (1985). Recombination between mutant cauliflower mosaic virus DNAs. *Plant Mol. Biol.* **5**, 281–289.

CONDIT, C., HAGEN, T. J., MCKNIGHT, T. D., and MEAGHER, R. B. (1983). Characterization and preliminary mapping of cauliflower mosaic virus transcripts. *Gene* **25**, 101–108.

COVEY, S. N. (1985). Organization and expression of the cauliflower mosaic virus genome. *Mol. Plant Virol.* **2**, 121–159.

DAUBERT, S. D., SCHOELZ, J., DEBAO, L., and SHEPHERD, R. J. (1984). Expression of disease symptoms in cauliflower mosaic virus genomic hybrids. *J. Mol. Appl. Genet.* **2**, 537–547.

DAUBERT, S., SHEPHERD, R. J., and GARDNER, R. C. (1983). Insertional mutagenesis of the cauliflower mosaic virus genome. *Gene* **25**, 201–208.

DELSENY, M., and HULL, R. (1983). Isolation and characterization of faithful and altered clones of the genomes of cauliflower mosaic virus isolates Cabb B-JI, CM4-184, and Bari I. *Plasmid* **9**, 31–41.

DIXON, L., NYFFENEGGER, T., DELLEY, G., MARTINEZ-IZQUIERDO, J., and HOHN, T. (1986). Evidence for replicative recombination in cauliflower mosaic virus. *Virology* **150**, 463–468.

FRANCK, A., GUILLEY, H., JONARD, G., RICHARDS, K., and HIRTH, L. (1980). Nucleotide sequence of cauliflower mosaic virus DNA. *Cell* **21**, 285–294.

GARDNER, C. O., JR., MELCHER, U., SHOCKEY, M. W., and ESSENBERG, R. C. (1980). Restriction enzyme cleavage maps of the DNA of two cauliflower mosaic virus isolates. *Virology* **103**, 250–254.

GARDNER, R. C., HOWARTH, A. J., HAHN, P., BROWN-LUEDI, M., SHEPHERD, R. J., and MESSING, J. (1981). The complete nucleotide sequence of an infectious clone of cauliflower mosaic virus by M13mp7 shotgun sequencing. *Nucleic Acids Res.* **9**, 2871–2888.

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

GELDREICH, A., LEBEURIER, G., and HIRTH, L. (1986). *In vivo* dimerization of cauliflower mosaic virus DNA can explain recombination. *Gene* **48**, 277–286.

GRIMSLEY, N., HOHN, B., HOHN, T., and WALDEN, R. (1986a). "Agroinfection," an alternative route for viral infection of plants by using the Ti plasmid. *Proc. Natl. Acad. Sci. USA* **83**, 3283–3286.

GRIMSLEY, N., HOHN, T., and HOHN, B. (1986b). Recombination in a plant virus: Template-switching in cauliflower mosaic virus. *EMBO J.* **5**, 641–646.

HIROCHIKA, H., TAKATSUJI, H., UBASAWA, A., and IKEDA, J.-E. (1985). Site-specific deletion in cauliflower mosaic virus DNA: Possible involvement of RNA splicing and reverse transcription. *EMBO J.* **4**, 1673–1680.

HOHN, B., BALAZS, E., RUEGG, D., and HOHN, T. (1986). Splicing of an intervening sequence from hybrid cauliflower mosaic viral RNA. *EMBO J.* **5**, 2759–2762.

HOWELL, S. H., WALKER, L. L., and DUDLEY, R. K. (1980). Cloned cauliflower mosaic virus DNA infects turnips (*Brassica rapa*). *Science* **208**, 1265–1267.

HOWELL, S. H., WALKER, L. L., and WALDEN, R. M. (1981). Rescue of *in vitro* generated mutants of cloned cauliflower mosaic virus genomes in infected plants. *Nature (London)* **293**, 483–486.

HULL, R. (1980). Structure of the cauliflower mosaic virus genome III. Restriction endonuclease mapping of thirty-three isolates. *Virology* **100**, 76–90.

LEBEURIER, G., HIRTH, L., HOHN, B., and HOHN, T. (1982). *In vivo* recombination of cauliflower mosaic virus DNA. *Proc. Natl. Acad. Sci. USA* **79**, 2932–2936.

LEBEURIER, G., HIRTH, L., HOHN, T., and HOHN, B. (1980). Infectivities of native and cloned DNA of CaMV. *Gene* **12**, 139–146.

MARCK, C. (1988). "DNA Strider": A "C" program for the fast analysis of DNA and protein sequences on the Apple Macintosh family of computers. *Nucleic Acids Res.* **16**, 1829–1836.

MASON, W. S., TAYLOR, J. M., and HULL, R. (1987). Retroid virus genome replication. *Adv. Virus Res.* **32**, 35–96.

MAXAM, A. M., and GILBERT, W. (1980). Sequencing end-labeled DNA with base-specific chemical cleavages. *In* "Methods in Enzymology" (L. Grossman and K. Moldave, Eds.), Vol. 65, pp. 499–560. Academic Press, San Diego.

MELCHER, U. (1989). Symptoms of cauliflower mosaic virus infection in *Arabidopsis thaliana* and turnip. *Bot. Gazette* **150**, 139–147.

MELCHER, U. (1990). Similarities between putative transport proteins of plant viruses. *J. Gen. Virol.*, in press.

MELCHER, U., CHOE, I. S., LEBEURIER, G., RICHARDS, K., and ESSENBERG, R. C. (1986a). Selective allele loss and interference between cauliflower mosaic virus DNAs. *Mol. Gen. Genet.* **203**, 230–236.

MELCHER, U., STEFFENS, D. L., CHOE, I. S., LYTTLE, D. J., LEBEURIER, G., LIN, H., and ESSENBERG, R. C. (1986b). Infectious and non-infectious mutants of cauliflower mosaic virus DNA. *J. Gen. Virol.* **67**, 1491–1498.

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.

SANGER, F. (1981). Determination of nucleotide sequences in DNA. *Science* **214**, 1205–1211.

SCHOELZ, J. E., DAUBERT, S. D., and SHERHERD, R. J. (1984). Gene VI of cauliflower mosaic virus controls systemic spread in solanaceous hosts. *Phytopathology* **74**, 861.

SCHOELZ, J., SHEPHERD, R. J., and DAUBERT, S. (1986). Region VI of cauliflower mosaic virus encodes a host range determinant. *Mol. Cell. Biol.* **6**, 2632–2637.

STRATFORD, R., and COVEY, S. N. (1989). Segregation of cauliflower mosaic virus symptom genetic determinants. *Virology* **172**, 451–459.

STRATFORD, R., PLASKITT, K. A., TURNER, D. S., MARKHAM, P. G., and COVEY, S. N. (1988). Molecular properties of Bari 1, a mild strain of cauliflower mosaic virus. *J. Gen. Virol.* **69**, 2375–2386.

SUN, T. J., MELCHER, U., and ESSENBERG, M. (1988). Inactivation of cauliflower mosaic virus by a photoactivatable cotton phytoalexin. *Physiol. Mol. Plant Pathol.* **33**, 115–126.

TABOR, S., and RICHARDSON, C. C. (1987). DNA sequence analysis with a modified bacteriophage T7 DNA polymerase. *Proc. Natl. Acad. Sci. USA* **84**, 4767–4771.

WALDEN, R. M., and HOWELL, S. H. (1982). Intergenomic recombination events among pairs of defective cauliflower mosaic virus genomes. *J. Mol. Appl. Genet.* **1**, 447–456.

ZHANG, X. S., and MELCHER, U. (1989). Competition between isolates and variants of cauliflower mosaic virus in infected turnip plants. *J. Gen. Virol.* **70**, 3427–3437.