Excision and Episomal Replication of *Cauliflower mosaic virus* Integrated into a Plant Genome

Julie Squires, Trudi Gillespie, James E. Schoelz, and Peter Palukaitis

Transgenic Arabidopsis (*Arabidopsis thaliana*) plants containing a monomeric copy of the cauliflower mosaic virus (CaMV) genome exhibited the generation of infectious, episomally replicating virus. The circular viral genome had been split within the nonessential gene II for integration into the Arabidopsis genome by *Agrobacterium tumefaciens*-mediated transformation. Transgenic plants were assessed for episomal infections at flowering, seed set, and/or senescence. The infections were confirmed by western blot for the CaMV P6 and P4 proteins, electron microscopy for the presence of icosahedral virions, and through polymerase chain reaction across the recombination junction. By the end of the test period, a majority of the transgenic Arabidopsis plants had developed episomal infections. The episomal form of the virus was infectious to nontransgenic plants, indicating that no essential functions were lost after release from the Arabidopsis chromosome. An analysis of the viral genomes recovered from either transgenic Arabidopsis or nontransgenic turnip (*Brassica rapa* var *rapa*) revealed that the viruses contained deletions within gene II, and in some cases, the deletions extended to the beginning of gene III. In addition, many of the progeny viruses contained small regions of nonviral sequence derived from the flanking transformation vector. The nature contained deletions within gene II, and in some cases, the deletions extended to the beginning of gene III. In addition, many of the progeny viruses contained small regions of nonviral sequence derived from the flanking transformation vector. The nature of the nucleotide sequences at the recombination junctions in the circular progeny virus indicated that most were generated by nonhomologous recombination during the excision event. The release of the CaMV viral genomes from an integrated copy was not dependent upon the application of environmental stresses but occurred with greater frequency with either age or the late stages of plant maturation.

Members of the virus family Caulimoviridae are ideal subjects for the study of recombination in plants. Several members of the Caulimoviridae have been found integrated into plant genomes, including *Banana streak virus* in banana (*Musa* spp.), *Petunia vein clearing virus* (PVCV) in petunia (*Petunia* spp.), *Tobacco vein clearing virus* in multiple *Nicotiana* species, and *Dahlia mosaic virus* in dahlia (*Dahlia variabilis*; for review, see Harper et al., 2002; Staginnus and Richert-Pöggeler, 2006; Pahalawatta et al., 2008). These integrated viruses have been called plant endogenous pararetroviruses (EPRVs) to distinguish them from pararetroviruses that never have been found integrated into host genomes in nature (Staginnus and Richert-Pöggeler, 2006). None of the episomal forms of these viruses are transmitted through seed, so the processes that lead to episomal infections occur with every generation of plants.

The nucleotide sequences of the integrated viral forms have revealed that the mechanisms that lead to activation of the EPRVs and subsequent episomal infections are complex. Activation of the EPRVs is thought to be associated with epigenetic changes that could (1) allow the production of a greater than full-length transcript or (2) lead to homologous recombination and circularization of the ERPV genome (Staginnus and Richert-Pöggeler, 2006). For example, the PVCV genome is integrated in a complete, continuous form in tandem arrays; activation could involve the production of a greater than full-length transcript that could be reverse transcribed into DNA (for review, see Harper et al., 2002; Staginnus and Richert-Pöggeler, 2006). On the other hand, the activation of *Banana streak virus* and *Tobacco vein clearing virus* would have to involve recombination between at least two integrated viral DNA segments integrated into different loci (Staginnus and Richert-Pöggeler, 2006). Furthermore, activation of the viral DNA and subsequent episomal replication appears to be triggered by changes in daylength or abiotic stresses such as drought or heat stress (Staginnus and Richert-Pöggeler, 2006). These host/virus combinations are valuable because they serve as exquisitely sensitive biosensors for recombination events that occur during the life of the plant.
since a single virus genome excised from its chromosomal location could be capable of replication and amplification within the plant.

We have sought to develop a system based on Cauliflower mosaic virus (CaMV) that could be used to examine the release of integrated caulimovirus sequences from host chromosomes and subsequent episomal infections. Such a system could be used to identify the genetic and environmental parameters for development of the caulimovirus episomal infections in plants. Although CaMV has never been shown to exist as a natural integrant in its hosts, its complete genome has been introduced into plant chromosomes through Agrobacterium tumefaciens-mediated transformation (Young et al., 1987; Gal et al., 1991), and progeny plants were examined for recombinants. In one study, a greater than full-length copy of the CaMV genome integrated into a host chromosome was able to be excised and to replicate episomally (Gal et al., 1991). The duplicated region of CaMV DNA was 989 bp, which provided a suitable substrate for the release of an episomal virus. In this instance, the mechanism for release was thought to involve the production of the full-length 35S RNA and its subsequent reverse transcription into DNA. Similarly, infectious CaMV also was generated when Escherichia coli plasmid DNA containing greater than full-length CaMV genomes were directly introduced into plant cells (Grimmery et al., 1986; Stratford and Covey, 1989; Vaden and Melcher, 1990). In another study (Young et al., 1987), only a single copy of the CaMV genome, interrupted within the essential gene V (Fig. 1), was integrated into the host. Although the gene VI protein product was detected, due to the activity of the CaMV 19S RNA promoter, no episomal infections arose from the insertion of this single CaMV genomic copy. Similarly, E. coli plasmids containing only a single copy of the CaMV genome were not infectious after inoculation to plants unless the flanking plasmid vector sequences were removed first, regardless of whether the viral genome was split within the essential gene V or the nonessential gene II (Walden and Howell, 1983).

Taken together, these studies suggested that episomal forms of CaMV would develop readily in transgenic plants if a 35S RNA transcript could be produced that could serve as a substrate for reverse transcription. However, in plants that contained a single copy of the CaMV genome, an infectious viral DNA would not be able to be excised from flanking nonviral DNA sequences, or the efficiency of nonhomologous recombination would be so low as to be undetectable. The recombination studies with transgenic plants that contain CaMV genomes stand in stark contrast to what occurs in nature with most plant EPRVs, in which the viral genomes are divided among several loci, yet episomal infections are common.

To model the occurrence of episomal infections in plants, we transformed Arabidopsis (Arabidopsis thaliana) plants with a single copy of the CaMV genome split within gene II (Fig. 1). Since gene II is not required for replication or movement within the plant, recombination events confined within gene II would not abolish the infectivity of the virus. These transgenic plants then were used to assess whether recombination events could indeed lead to episomal replication of CaMV, whether the environment influenced the development of episomal infections, and what mechanism for recombination was used. We found that episomal infections readily developed in the transgenic plants, even when they were not subjected to any

Figure 1. Genome organization and expression strategy of CaMV. A, The circular DNA genome of CaMV showing the six characterized genes of CaMV (genes I–VI) as well as the major functions of the encoded proteins. In addition, two mRNAs are transcribed from the CaMV genome: The 19S RNA encodes the gene VI product, whereas the 35S RNA is a polycistronic mRNA for expression of genes I to V. The 35S RNA is also the template for reverse transcription of the CaMV genome into double-stranded DNA. B, The linear structure of the CaMV genome, cloned at its unique XhoI site into the plant transformation vector pGreen. The positions of the PCR primers used to detect episomal replication of the circular genome after excision and ligation are indicted by arrowheads in both A and B.
discernible stress. The results of our study show that the threshold for release of caulimovirus sequences integrated into plant genomes is surprisingly low.

RESULTS

Detection of Excision and Episomal Replication of CaMV

Transgenic, T2 generation Arabidopsis plants transformed with a full-length copy of the circular CaMV genome split at the XhoI site within gene II (Fig. 1B) were used in these experiments. These plants contained a single copy of the CaMV genome, as determined from a combination of the segregation frequency of the Basta-resistant plants (data not shown) and product analysis of PCR used to verify that the integration event at one locus was due to a single copy of the CaMV DNA and not to tandem inserts (Supplemental Fig. S1). Since the plants in the T2 generation still segregated for the presence of the CaMV genome, a leaf was removed at an early growth stage to identify transgenic plants. Plants from line 316 identified as transgenic were propagated under environmentally controlled conditions and analyzed for the presence of episomal infections at 30 d post planting. The environmental conditions included drought, heat, high light intensity, and infection by another viral pathogen, Cucumber mosaic virus (CMV). CMV is a plant virus containing an RNA genome and, unlike CaMV, does not have a nuclear phase in its replication strategy, although its 2b gene product does enter the nucleus (for review, see Palukaitis and García-Arenal, 2003). Episomal infections were assessed by western-blot analysis (Fig. 2) for viral proteins P6 (transactivator protein) and P4 (coat protein), by PCR for viral DNA, and by immunocapture electron microscopy of CaMV virions (Fig. 3).

A western-blot analysis for the presence of the P6 protein revealed that it was detected in the majority of the transgenic plants (Fig. 2A; data not shown). The CaMV P6 product is expressed from the 19S RNA promoter (Fig. 1); consequently, a baseline of expression would be expected to be present in all transgenic plants. However, we found that there was considerable variation in its concentration. In at least one transgenic plant (CMV-infected no. 3), the level of P6 protein was comparable to its expression in transgenic plants driven by a 35S RNA promoter (D4-2 samples 1 and 2). Multiple bands were detected for P6, which is in agreement with previous studies indicating the presence of the full-length P6 product and several breakdown products (Daubert and Routh, 1990; Schoelz et al., 1991). In contrast, the P4 protein is translated from the polycistronic 35S RNA. Given the structure of the CaMV insert in the T-DNA (Fig. 1), the P4 product could be expressed only after excision and episomal replication of the CaMV viral genome. The P4 protein was detected in fewer transgenic plants (Fig. 2B; data not shown), but the plants that expressed high levels of P4 protein also expressed high levels of the P6 protein (Fig. 2A and B). Taken together, the western blots for the CaMV P6 and P4 proteins indicated that episomal replication of CaMV had occurred in at least some of the transgenic plants.

Episomal replication of the viral DNA was confirmed by PCR (Fig. 2C), using primers that flanked the insertion of CaMV sequences into the T-DNA of the Agrobacterium vector (Fig. 1B). Consequently, a PCR band could be generated only after a recombination event that resulted in recirculation of the viral DNA.
to trigger the excision and episomal replication of CaMV DNA. Hence, in subsequent experiments, the plants were not prescreened for the presence of the transgene and they were assessed at different stages of development for episomal replication of CaMV. Consequently, some of the plants would not be expected to carry the CaMV transgene due to segregation of the trait. In one such experiment, episomal replication again was detectable in the line 316 control plant samples (no applied stress) taken at different times after germination (Table I). In general, the number of plants showing evidence of episomal replication increased as the age of the plants increased (Table I). This may have been due to the effects of plant development, aging, or possibly the occurrence increasing with time and number of plant genome replication cycles.

To examine whether environmental stresses were required to induce episomal replication, plants of line 316 also were propagated under various environmental stress conditions, including high light intensity, high temperature (heat), drought, and infection by CMV. Episomal replication was detected in plants subjected to each of these environmental stresses, but it was also detected in a high percentage of the control plants that were not subjected to any applied stress (Table I). In fact, the percentage of control plants in which recombinants could be detected by PCR either equaled or exceeded the percentage in plants that had been subjected to stress. A statistical analysis of the data in Table I, using a generalized linear model regression analysis, indicated that none of the environmental stress conditions appeared to increase episomal infections. In this experiment, the choice of assay method was not significant at the 5% level but was significant at the slightly weaker 7.6% level. From this analysis, episomal P4 protein was predicted to be observed in 35.7% ± 11.3% of plants, episomal P6 protein was predicted to be observed in 90.8% ± 9.1% of plants, and 83.7% ± 10.6% of plants were estimated to show episomal replication by PCR. Stress treatments were only significantly different at the 16.8% level. As with the nonstressed plants, the frequency of detection of episomal replication appeared to be higher in older plants, especially those reaching senescence. Thus, while environmental stress did not enhance the frequency of excision and episomal replication, the age of the plants or perhaps a maturation-related stress may have had some effect.

Deletions and Insertions in the Sequences of the Progeny Viral Genomes

PCR products obtained from selected plants of line 316 or line 318 were cloned, and two clones were sequenced from each PCR product. In most but not all cases, the sister clones had identical sequences. Analysis of the sequences revealed several trends (Fig. 4; Supplemental Fig. S2). The sequences of the PCR products indicated that all of the viruses contained a

Figure 3. Electron micrographs of CaMV particles. Anti-CaMV particle antibody was used to coat electron microscopy grids. Virus extract was made using phosphate buffer, pH 6.5, and a standard antibody-coated grid protocol was used. The grids were stained with ammonium molybdate, pH 6.5. Images are virions recovered from Arabidopsis containing CaMV E316-10 (A; smaller PCR product from Fig. 2C) or a control CaMV-H7-infected turnip (B). Bars = 50 nm.

Cumulative Development of Episomal Infections in Stressed and Nonstressed Plants

It is conceivable that the removal of a leaf for analysis of transgene status might be a sufficient biotic stress genome. Interestingly, PCR products were amplified from every plant except one (Drought no. 2), in contrast to the western blots for the P4 protein. Furthermore, several of the PCR-derived bands were larger in size than the bands derived from CaMV strains H7 and W260, an indication that the viral DNA had either acquired foreign DNA sequences or sustained a rearrangement.

To further confirm the presence of episomally replicating viruses, leaf dips from individual plants were examined by immunocapture electron microscopy, revealing the presence of icosahedral CaMV particles (Fig. 3). Typical CaMV particles were detected in leaf dips of nonstressed plants and each type of applied-stressed plants (Fig. 3; data not shown).

To verify the reproducibility of the above results and to demonstrate that they were not limited to one transgenic line (line 316), a second transgenic line (line 318) was evaluated under two environmental conditions for episomally replicating CaMV at 25 d post planting and again at flowering. Both the western blots for the P4 protein and PCR across the gene II/III junction confirmed the detection of episomal replication of CaMV (data not shown).
deletion within gene II, and in several viruses, the deletions extended into gene III (see clones E318-4, E318-5, and E318-7 in Fig. 4). Although gene II can be deleted without affecting infectivity (Howarth et al., 1981), gene III has an essential role in cell-to-cell movement (Stavolone et al., 2005). Consequently, viruses that sustained a deletion within gene III were likely dysfunctional. Some of the progeny viruses obtained from different plants had identical deletions (see E318-1F and E318-6Fa as well as E318-5F and E318-6Fb), an indication that specific sequences might facilitate recombination events. Interestingly, in some cases, the viruses present in samples extracted before flowering had larger deletions than those taken from the same plants after flowering (see E318-4 versus E318-4Ft/m/b and E318-5 versus E318-5Ft); in other cases, this could not be ascertained, since the plants either were negative by PCR for the presence of episomal virus before flowering (E318-1, E318-3, and E318-6) or died after flowering (E318-7). These examples are likely indicative of new recombination-excision events in a single plant, but we cannot rule out that they may have been generated by recombination of a defective virus (lacking large parts of genes II and III) with the homologous, intact CaMV genome sequences. It is important to note that such recombination events would not result in a wild-type virus.

Many of the PCR products also contained insertions of DNA sequences that varied in length from 41 to 135 bp and were derived from the downstream, flanking transformation vector that had been incorporated into the plant genome during the transformation of Arabidopsis (Fig. 4; Supplemental Fig. S2). The presence of the vector DNA sequences was not correlated with environmental stress, as these sequences were recovered from plants that had not been subjected to stress (Fig. 4; e.g. E316-10, E316-11, E318-F, and E318-4) as well as from stressed plants (e.g. E316-6, E316-7, and E318-1). Some PCR products also contained small regions (6, 12, or 22 bp) derived from the upstream, flanking vector (Fig. 4; yellow regions in samples E316-6, E316-7, and E318-4ma). Finally, some viruses contained additional vector DNA that was not present in viruses amplified from the same plant at an earlier time point (Fig. 4; see E318-4Ft and the two clones of E318-4Fm versus E318-4). This is further evidence that multiple recombination-excision events occurred in a single plant. Although a number of discreet deletion variants of CaMV were obtained, since CaMV is not seed transmissible (Tompkins, 1937; Squires et al., 2007), these variants would have to be created anew in every generation of plants.

Recombination junctions were identified in the 28 clones recovered from transgenic Arabidopsis plants; 12 were characterized by a region of microhomology of four to six nucleotides (Supplemental Fig. S2), whereas no evidence for microhomology existed in the other 16 clones. No consensus sequence emerged within the 12 junctions that exhibited microhomology. In seven of these clones, the recombination ends were imperfect, but some degree of homology could be seen. For example, in two of the clones with imperfect recombination ends (E318-4F-ba and E318-4F-bb), the sequence on one end of the recombination junction was TTACT, whereas the sequence on the other end was TTAATT. In the five other clones with imperfect recombination ends (E318-6Fb, E318-5a, E318-5b, E318-5Fa, and E318-5Fb), the sequence AGTA was found on one end, whereas on the other end, AGTA was located two nucleotides from the recombination junction. In five clones, the ends of the recombination junctions were identical: GGGT in clones E316-7a and E316-7b, TCCT in clones E318-3Fa and E318-3Fb, and CCGGG in clone E318-4Fma.

### Table 1. Expression of episomal replication following stress treatment of individual plants of transgenic line 316

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Flowering</th>
<th>Seed Set</th>
<th>Senescence</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>P4b</td>
<td>P6b</td>
<td>PCR</td>
</tr>
<tr>
<td>Control</td>
<td>0/4</td>
<td>3/4</td>
<td>4/4</td>
</tr>
<tr>
<td>Light</td>
<td>1/4</td>
<td>4/4</td>
<td>1/4</td>
</tr>
<tr>
<td>CMV</td>
<td>2/4</td>
<td>1/4</td>
<td>1/4</td>
</tr>
</tbody>
</table>

*Plant incubated under stress conditions: high light intensity, heat drought, or by infection with CMV. †Detection of CaMV-encoded proteins P4 and P6 by western blotting. ‡Detection of replicating virus by PCR. §Standard incubation conditions of 22°C day/18°C night, 250 μmol m⁻² s⁻¹, and watered twice a day. ‖Number of plants showing P4, P6, or replicating virus over number of plants tested. ¶Light intensity increased to 450 μmol m⁻² s⁻¹. ⌂32°C day and night. ❋Watersed every other day. ❖Infected by CMV.

### Recovery and Analysis of CaMV after Passage to Nontransgenic Plants

The deletions detected in the excised CaMV variants might affect their infectivity, since CaMV has very strict requirements for translation of its 35S RNA. To investigate whether the excised CaMV variants were
able to replicate and spread in nontransgenic plants, plants of transgenic lines 316 and 318 were grown for 45 d, and plants were bulked for isolation of CaMV virions. CaMV virions were partially purified as described previously (Schoelz et al., 1986) and were used for inoculation of 15 turnip (Brassica rapa var rapa ‘Just Right’) plants. An ELISA for the presence of CaMV virions revealed that at 21 d after inoculation, nine of the 15 plants were positive for CaMV infections (data not shown); initial evidence was that the episomal CaMV genomes that originated in Arabidopsis were infectious.

Viral DNAs were isolated from four of the turnip plants that tested positive for CaMV by ELISA, and an EcoRI digest revealed that all four plants contained the major DNA bands characteristic of a full-length viral genome (Fig. 5A). A PCR analysis of CaMV sequences flanking the putative recombination junction revealed that each of the plants contained multiple viral genomes, as viral DNA isolated from individual turnip plants yielded PCR products that varied in size from approximately 250 to 760 bp (Fig. 5B). In contrast, the CaMV full-length clone pCaMV10 (Gardner et al., 1981) consistently yielded a PCR product of 676 bp (Fig. 5B). Furthermore, noncloned viral DNA of CaMV strain H31 (Schoelz and Shepherd, 1988), which had been isolated from infected turnip, also yielded a PCR product of 676 bp (data not shown). This is significant because it showed that CaMV DNA isolated from turnip plants inoculated with an infectious CaMV clone did not contain size variations indicative of multiple insertions and deletions.

The nucleotide sequences of PCR clones isolated from the turnip plants had many similarities to the PCR clones isolated directly from the transgenic Arabidopsis. All clones sustained deletions within gene II (Fig. 6; Supplemental Fig. S3). Some clones (15-1 and 3-2) also contained insertions of vector sequences, which demonstrated conclusively that these clones were derived from the transgenic Arabidopsis plants. These two clones were very similar, but not identical, to the clones recovered directly from the transgenic Arabidopsis. For example, the upstream recombination junction of clones 15-1 and 3-2 was identical to clone E316-7, whereas on the downstream junction, 15-1 and 3-2 contained a larger segment of the Agrobacterium vector than E316-7 (Supplemental Fig. S3). Interestingly, the changes in all three of these clones retained the gene II reading frame.

Figure 4. Genomic structure of recombinant, episomally replicating progeny recovered from transgenic Arabidopsis plants. CaMV genes II and III are shown in purple and blue, respectively, whereas sequences derived from the Agrobacterium binary vector pGreen are illustrated in dark green (vector D for sequences downstream from the CaMV insertion point) and yellow (vector U for sequences upstream from the CaMV insertion point). XhoI sites that have been retained in the episomal viruses are shown in light green. Deleted sequences are illustrated by thin lines. The nucleotide sequences, presented in Supplemental Figure S2, were obtained from individual plants of two transgenic lines (316 and 318), with plant numbers shown after the line number. Two clones were sequenced from each PCR product obtained from each plant, and these are indicated with letter designations a or b, but they are only shown separately when the sequences were different from sister clones. In some cases, samples were collected after flowering (indicated by an F after the plant number). For samples from plant E318-4F, three PCR bands were obtained, and sequences from these bands are presented as t (top band), m (middle band), and b (bottom band).
The population of viruses recovered directly from Arabidopsis did differ in some ways from the population of viruses recovered from turnip. For example, none of the deletions in viruses recovered from turnip extended into gene III. Since gene III is essential for infection (Stavolone et al., 2005), this class of deletions might be expected to be eliminated upon passage to nontransformed plants. However, we did find one deletion (clone 5-2; Figure 6; Supplemental Fig. S3) in which the last 12 codons of gene I, encoding the cell-to-cell movement protein, were eliminated. The turnip plant from which clone 5-2 was derived also contained other viral genomes in which gene I was intact (represented in clones 5-3 and 5-18), so it is possible that they may have complemented 5-2 for cell-to-cell movement. Nevertheless, this experiment showed that CaMV virions generated in the transgenic Arabidopsis were competent to establish systemic infections in nontransgenic plants.

Of the nine clones recovered from turnip plants, the recombination junctions of two (5-3 and 5-18) were characterized by a region of microhomology six nucleotides in length (AAGGAT for clone 5-3 and GATCAA for clone 5-18; Supplemental Fig. S3). Interestingly, the recombination junction of clone 5-3 was identical to CM4-184 (Fig. 6; Supplemental Fig. S3), a naturally occurring CaMV deletion mutant that was characterized 30 years ago (Howarth et al., 1981). There was no evidence of any microhomology around the recombination junctions in the remaining seven clones (Fig. 3). Two of these clones (15-1 and 3-2) contained a single filler nucleotide, a thymidine, which could not be accounted for in either the vector DNA or the CaMV DNA sequence, whereas no filler sequences were found in any of the other junctions.

DISCUSSION

Recombination associated with CaMV has been investigated in great detail. Several studies showed that overlapping fragments of the viral genome, which between them contained the entire viral genome, delivered into the same cells could rapidly recombine to yield the infectious, wild-type genome (Howell et al., 1981; Lebeurier et al., 1982; Walden and Howell, 1983; Grimsley et al., 1986; Vaden and Melcher, 1990). In contrast, plasmids containing a single copy of the viral genome did not yield infectious viral progeny in inoculated plants (Walden and Howell, 1983). Furthermore, CaMV isolates containing neutral genetic markers were observed to undergo abundant recombination within a single plant; recombinants accounted for up to 50% of the viral population (Froissart et al., 2005). Later studies showed that recombination was possible between CaMV and transgenes derived from CaMV under strong and weak selection pressures (Schoelz and Wintermantel, 1993; Wintermantel and Schoelz, 1996). Furthermore, the release of infectious viral genomes from longer than full-length CaMV genomes integrated into the genome of Brassica napus (Gal et al., 1991) indicated that infectious virus could arise from plant genomes. In most of these studies, the mechanism of recombination could be explained by template switches from one RNA species to another during the reverse transcription process, rather than by a homologous recombination event. However, it is difficult to discern the precise mechanism involved in many of the recombinants, because the recombination events could be explained by either mechanism. For example, Vaden and Melcher (1990) examined the recombination junctions of eight recombinant viruses formed between different CaMV isolates and concluded that their evidence supported a model involving reverse transcription as well as recombination between double-stranded DNA. Our study provides the clearest evidence for a model involving non-

Figure 5. PCR analysis of episomal CaMV DNA isolated from turnip plants. A, An agarose gel containing an EcoRI digest of viral DNA isolated from individual, nontransgenic turnip plants. The lane labeled pCaMV10 consists of CaMV strain CM1841 DNA cloned at its unique SalI site into the E. coli plasmid pBR322. The arrows on the side illustrate the three largest products of the EcoRI digest of noncloned viral DNA. B, An agarose gel showing PCR products generated from episomally replicating CaMV DNAs recovered from individual, nontransgenic turnip plants. PCR primers spanned the recombination junction between genes II and III. The lane labeled pCaMV10 illustrates the expected 676-bp PCR product obtained from an infectious CaMV clone.
homologous recombination, with strand breakage of DNA, either flanking or within the CaMV sequences contained in the chromosomal DNA, followed by ligation of the CaMV genome into a circular form. Furthermore, our study illustrates how an integrated, monomer copy of a pararetrovirus could be activated from host chromosomes and converted into an infectious form.

Two mechanisms might explain the release of an infectious CaMV virus from a monomer-length transgene source. One mechanism involves strand breakage of DNA either flanking or within the CaMV sequences contained in the chromosomal DNA, followed by ligation of the CaMV genome into a circular form (Fig. 7). A second mechanism would require the synthesis of the CaMV reverse transcriptase (RT) and subsequent reverse transcription of viral RNAs into a DNA copy (Fig. 8). Both mechanisms could allow for the insertion of binary vector sequences into the CaMV genome, as well as deletions within gene II. The recircularized viral DNA would then be able to replicate episomally, and those viral DNAs with an intact gene III would also be able to infect the plant systemically (Jacquot et al., 1998; Kobayashi et al., 2002).

It is important to illustrate the steps that would be necessary for RT-mediated release to evaluate its viability as a model for the release of CaMV in our study (Fig. 8). For example, there are two prerequisites for the RT-mediated mechanism of recombination to occur: one is the production of a polycistronic mRNA through the action of a fortuitously placed Arabidopsis promoter, and the second is the production of the CaMV RT (P5). Of the two, the greatest hurdle involves the synthesis of the RT protein, because the RT cistron is in the third position on the CaMV portion of a putative polycistronic transcript (Fig. 8). Any start codons in the Arabidopsis or pGreen portions of the transcript would further impede translation of the RT. Consequently, the RT could be produced only through the action of the CaMV translational transactivator (TAV; P6), which would be expressed in transgenic plants from the monocistronic 19S mRNA. In principle, the TAV could redirect ribosomes to act on the polycistronic mRNA for synthesis of the RT. However, the evidence of Young et al. (1987) suggests that the RT would not be produced in this scenario. Young et al. (1987) generated transgenic plant tissue that contained the full genome of CaMV and evaluated which viral proteins were synthesized. They found that neither viral protein P1 nor P4 was produced from a polycistronic transcript, even though the presence of the TAV protein was verified. The implication is that there may be a fundamental difference between translation of a polycistronic mRNA from a transgenic source and the translation of the polycistronic 35S RNA in the context of a CaMV infection. Indeed, there is little evidence for the effectiveness of the CaMV P6 protein as a TAV on polycistronic transgenes beyond its action on a single dicistronic transgene (Zijlstra and Hohn, 1992). If the CaMV RT is not produced from a polycistronic mRNA through the action of a fortuitously placed Arabidopsis promoter, and the second is the production of the CaMV RT (P5). Of the two, the greatest hurdle involves the synthesis of the RT protein, because the RT cistron is in the third position on the CaMV portion of a putative polycistronic transcript (Fig. 8). Any start codons in the Arabidopsis or pGreen portions of the transcript would further impede translation of the RT. Consequently, the RT could be produced only through the action of the CaMV translational transactivator (TAV; P6), which would be expressed in transgenic plants from the monocistronic 19S mRNA. In principle, the TAV could redirect ribosomes to act on the polycistronic mRNA for synthesis of the RT. However, the evidence of Young et al. (1987) suggests that the RT would not be produced in this scenario. Young et al. (1987) generated transgenic plant tissue that contained the full genome of CaMV and evaluated which viral proteins were synthesized. They found that neither viral protein P1 nor P4 was produced from a polycistronic transcript, even though the presence of the TAV protein was verified. The implication is that there may be a fundamental difference between translation of a polycistronic mRNA from a transgenic source and the translation of the polycistronic 35S RNA in the context of a CaMV infection. Indeed, there is little evidence for the effectiveness of the CaMV P6 protein as a TAV on polycistronic transgenes beyond its action on a single dicistronic transgene (Zijlstra and Hohn, 1992). If the CaMV RT is not produced from a polycistronic mRNA through the action of a fortuitously placed Arabidopsis promoter, and the second is the production of the CaMV RT (P5). Of the two, the greatest hurdle involves the synthesis of the RT protein, because the RT cistron is in the third position on the CaMV portion of a putative polycistronic transcript (Fig. 8). Any start codons in the Arabidopsis or pGreen portions of the transcript would further impede translation of the RT. Consequently, the RT could be produced only through the action of the CaMV translational transactivator (TAV; P6), which would be expressed in transgenic plants from the monocistronic 19S mRNA. In principle, the TAV could redirect ribosomes to act on the polycistronic mRNA for synthesis of the RT. However, the evidence of Young et al. (1987) suggests that the RT would not be produced in this scenario. Young et al. (1987) generated transgenic plant tissue that contained the full genome of CaMV and evaluated which viral proteins were synthesized. They found that neither viral protein P1 nor P4 was produced from a polycistronic transcript, even though the presence of the TAV protein was verified. The implication is that there may be a fundamental difference between translation of a polycistronic mRNA from a transgenic source and the translation of the polycistronic 35S RNA in the context of a CaMV infection. Indeed, there is little evidence for the effectiveness of the CaMV P6 protein as a TAV on polycistronic transgenes beyond its action on a single dicistronic transgene (Zijlstra and Hohn, 1992). If the CaMV RT is not produced from a polycistronic mRNA through the action of a fortuitously placed Arabidopsis promoter, and the second is the production of the CaMV RT (P5). Of the two, the greatest hurdle involves the synthesis of the RT protein, because the RT cistron is in the third position on the CaMV portion of a putative polycistronic transcript (Fig. 8). Any start codons in the Arabidopsis or pGreen portions of the transcript would further impede translation of the RT. Consequently, the RT could be produced only through the action of the CaMV translational transactivator (TAV; P6), which would be expressed in transgenic plants from the monocistronic 19S mRNA. In principle, the TAV could redirect ribosomes to act on the polycistronic mRNA for synthesis of the RT. However, the evidence of Young et al. (1987) suggests that the RT would not be produced in this scenario. Young et al. (1987) generated transgenic plant tissue that contained the full genome of CaMV and evaluated which viral proteins were synthesized. They found that neither viral protein P1 nor P4 was produced from a polycistronic transcript, even though the presence of the TAV protein was verified. The implication is that there may be a fundamental difference between translation of a polycistronic mRNA from a transgenic source and the translation of the polycistronic 35S RNA in the context of a CaMV infection. Indeed, there is little evidence for the effectiveness of the CaMV P6 protein as a TAV on polycistronic transgenes beyond its action on a single dicistronic transgene (Zijlstra and Hohn, 1992). If the CaMV RT is not produced from a polycistronic mRNA...
in transgenic tissue, then there is little possibility that recombination could be mediated by template switching during replication.

In contrast, there are good precedents for the strand breakage and religation model. The potential for homologous recombination has been evaluated in plants by using a transgene cassette that consists of two nonfunctional copies of the luciferase gene that contained 1,146 bp of overlapping sequence (Gorbunova et al., 2000; Kovalchuk et al., 2003). The restoration of luciferase function was visualized as bright spots on the leaves and indicated that a homologous recombination event had led to the excision of the duplicated portion of the luciferase gene. The analysis of recombination junctions in our study indicates that the excision would be mediated by nonhomologous recombination and would lead to numerous products, some of which are capable of episomal replication.

It is unclear at this point what factors initiated the recombination event. Stress caused by heat, drought, high light intensity, or infection with another virus did not appear to enhance the incidence of episomal replication of integrated CaMV. In contrast, environmental stresses do enhance the excision and subsequent episomal replication of PVCV (Richert-Poggeler et al., 2003; Noreen et al., 2007), and pathogen-induced stress also increased the frequency of homologous recombination and restoration of luciferase function in transgenic plants (Kovalchuk et al., 2003). Since CaMV is able to replicate itself upon excision from the host chromosome, it may be difficult to detect subtle differences in recombination frequency. For example, a baseline level of recombination that led to restoration of luciferase function was detected in transgenic plants even in the absence of any identifiable stress (Kovalchuk et al., 2003). When the outcome of recombination is an infectious virus, even a low baseline level of recombination might be sufficient for all plants to eventually become infected.

In addition, it might be impossible to completely eliminate pathogen stress as a contributor to recombination in our study. The CaMV P6 product is expressed as a monocistronic mRNA from the 19S RNA promoter, and several studies have shown that expression of P6 in transgenic Arabidopsis plants elicits virus-like symptoms (Zijlstra and Hohn, 1992; Cecchini et al., 1997; Yu et al., 2003). It is possible that the expression of P6 might trigger the plant genome instability and the excision of the infectious CaMV DNA. We also noted that the incidence of episomal replication of CaMV increased with time or at later stages of maturation. It seems less likely that this was due simply to increased probability associated with an increased number of cellular divisions, since there was much less cellular division that would have taken place between flowering and senescence than in the growth stages up to flowering. Thus, it seems more likely that some event or process associated with age or maturation triggered the events leading to DNA strand breakage, subsequent religation, and episomal replication. Ultimately, the factor(s) that contributed to excision and episomal replication, whether it is

---

**Figure 7.** Strand breakage and religation model for excision of CaMV from Arabidopsis chromosomes. CaMV gene II is highlighted in purple, whereas the upstream and downstream regions of the *Agrobacterium* vector pGreen are highlighted in yellow and brown, respectively. In the recombination model, the CaMV genome is hypothesized to form a loop, juxtaposing sequences within gene II and/or the vector sequences. If nonhomologous recombination occurs between the upstream and downstream regions of pGreen, the outcome is an episomal CaMV genome with an insertion of pGreen sequences. If nonhomologous recombination occurs between sequences within gene II, the outcome is an episomal CaMV genome that has sustained a deletion within gene II.
The release of CaMV from host chromosomes also contributes to an understanding of what sequences in plants should or should not be designated recombination hotspots. The use of the CaMV 35S promoter for the expression of transgenes in plants has attracted controversy because it is considered a recombination hotspot (Ho et al., 1999, 2000; Kohli et al., 1999). The evidence for this hotspot was based on a limited number of intrachromosomal recombination events (Kohli et al., 1999), as recombination involving the 35S RNA promoter was detected in four of the 12 transgenic rice lines. Others have used this study to suggest that recombination associated with the 35S RNA promoter could lead to untenable risks to human health and to the environment. These risks include the accidental activation of plant genes or endogenous viruses that could somehow promote cancer in humans, or that the 35S RNA promoter might recombine with mammalian viruses such as Human immunodeficiency virus, with unexpected consequences (Ho et al., 1999, 2000). The uncertainty associated with recombination involving the 35S RNA promoter has created an environment of fear within the general public that has damaged the credibility of genetically modified plant products. However, since the initial publication describing the CaMV 35S RNA promoter as a hotspot, new information has emerged regarding the potential for recombination in plants (Gorbunova et al., 2000; Kovalchuk et al., 2003), indicating that the concept of the recombination hotspot as defined by Kohli et al. (1999) should be reexamined.

In our study, we found that recombination centered around gene II nucleotide sequences could be detected by PCR in 29 of 48 transgenic plants (Table I). There was no evidence for the involvement of microhomology in the recombination junctions of the majority of

**Figure 8.** A model for the release of CaMV DNA through template switches during reverse transcription of RNA intermediates.

Step 1. For an infectious CaMV virus to be produced from a RT-mediated event, three RNAs would have to be produced from the full-length CaMV DNA insert. The 3' ends of the mRNAs are represented by the arrows. One monocistronic mRNA (19S RNA) would be produced from the CaMV 19S RNA promoter, and this would lead to the production of the P6 TAV protein. A second polycistronic mRNA (mRNA A) would be initiated from the CaMV 35S RNA promoter, although it is unclear where this transcript would terminate, whether in the vector sequence or the Arabidopsis sequence flanking the CaMV insert. A third transcript (mRNA B) would have to be generated from a putative Arabidopsis promoter. This transcript would initiate in the Arabidopsis DNA and then continue through the vector sequence to CaMV gene II, ultimately ending at the 35S RNA termination sequence.

Step 2. The CaMV RT could only be produced through the action of the CaMV translational TAV, which would be synthesized in the transgenic plants through the action of the CaMV 19S RNA promoter. In principle, the TAV might redirect ribosomes to act on the polycistronic mRNA for synthesis of the RT. Step 3. Nonetheless, if RT was produced, it might initiate synthesis of the first strand of CaMV DNA within mRNA A. This mRNA contains a nucleotide sequence complementary to a Met tRNA, which serves as the primer binding site for first-strand CaMV DNA synthesis by the CaMV RT in natural infections. Step 4. When the CaMV RT reaches the 3'-end of mRNA A, it would then need to switch templates to the 3'-end of mRNA B. This step also corresponds to the replication process of CaMV in natural infections. Step 5. As the RT reaches the end of CaMV sequences within gene II on mRNA B, it would need to initiate a second template switch back to the 3'-end of mRNA A. This additional template switch would be necessary for the production of an infectious virus by reverse transcription.
clones we examined (23 of 37 clones; Supplemental Figs. S2 and S3). In addition, recombination occurred at multiple sites within gene II; no single recombination junction predominated in the clones sequenced. Consequently, it would be incorrect to state that the gene II locus contains a recombination hotspot. Instead, it appears that strand breakage and religation involving CaMV sequences at a specific locus might occur at least once within the life of a plant and may occur more than once in some plants. Indeed, Kovalchuk et al. (2003) found that uninfected, transgenic Nicotiana tabacum plants had an average of 8.4 luciferase spots, indicating that homologous recombination leading to restoration of luciferase function had occurred multiple times within a single plant. The background levels of intrachromosomal recombination observed in our study and by Kovalchuk et al. (2003) exceed those observed for recombination involving the 35S RNA promoter (Kohli et al., 1999), and collectively, these studies suggest that the 35S RNA promoter should not be considered a recombination hotspot.

MATERIALS AND METHODS

**Viruses Source, Maintenance, and Propagation**

A full-length, infectious clone of CaMV strain H7 (Daubert et al., 1984), a chimera generated between CaMV strains CM1841 and D4, was the source for transformation of Arabidopsis (Arabidopsis thaliana). Strain H7 elicits extremely mild symptoms in turnip (Brassica rapa var. rapa) and is not aphid transmissible (Daubert et al., 1984). The CaMV clones pCaMV10 and H31 have been described previously (Gardner et al., 1981; Schoelz and Shepherd, 1988). The CMV used for the environmental stress experiments was Fny-CMV, a transmissible (Daubert et al., 1984). The CaMV clones pSoup c58 (Hellens et al., 2000) and the P6 gene is under the control of the 35S RNA promoter from CaMV. The forward primer was 5′-GTGGGATTGTGCGTCA-3′ and the reverse primer was 5′-GATA-CCAGGATCTCAGGTA-3′. These particles were observed for samples E316-6, E316-10, and E317-11. Anti-CaMV particle antibody was used to coat the electron microscope grids. Virus extracts were made using phosphate buffer, pH 6.5, and a standard antibody-coated grid protocol was used. The grids were stained with ammonium molybdate, pH 6.5, or 2% uranyl acetate and viewed with the electron microscope (Phillips CM10).

**Transformation and Propagation of Arabidopsis**

The D4-2 plants are homozygous for expression of P6 from the D4 strain of CaMV (Yu et al., 2003). The P6 gene is under the control of the 35S RNA promoter and rbcS terminator sequences. To insert the full-length CaMV genome into Arabidopsis, the cloned CaMV strain H7 was digested with SalI to release the vector pBR322 and the gel-purified viral DNA was ligated into a circular form. The full-length, circular viral DNA was subsequently digested with Xhol and ligated into the Xhol site of the Agrobacterium tumefaciens plasmid pGreen 0229 (Hellens et al., 2000). The pGreen plasmid containing the full-length, circular viral DNA was subsequently digested with XhoI and ligated into the XhoI site of the Agrobacterium tumefaciens plasmid pSoup c58 (Hellens et al., 2000). Arabidopsis ecotype C24 plants were transformed using the flower dip method (Clough and Bent, 1998). The T0 seeds were collected from the infiltrated plants and screened for the presence of the 35S RNA promoter from CaMV. The forward primer was 5′-GATCCTAATTCTTC-3′ and the reverse primer was 5′-CAAAGACCCTTCGGAGT-3′. The PCR amplified DNA was subsequently digested with Xhol and ligated into the Xhol site of the Agrobacterium tumefaciens plasmid pSoup c58 (Hellens et al., 2000). Arabidopsis ecotype C24 plants were transformed using the flower dip method (Clough and Bent, 1998). The T0 seeds were collected from the infiltrated plants and screened for the presence of the 35S RNA promoter from CaMV. The forward primer was 5′-GATCCTAATTCTTC-3′ and the reverse primer was 5′-CAAAGACCCTTCGGAGT-3′. The PCR amplified DNA was subsequently digested with Xhol and ligated into the Xhol site of the Agrobacterium tumefaciens plasmid pSoup c58 (Hellens et al., 2000). Arabidopsis ecotype C24 plants were transformed using the flower dip method (Clough and Bent, 1998). The T0 seeds were collected from the infiltrated plants and screened for the presence of the 35S RNA promoter from CaMV. The forward primer was

**Extraction of Plant Nucleic Acids from Arabidopsis and PCR Analysis**

For examination of episomal replication in Arabidopsis, DNA was extracted from total protein preparations used in parallel for western-blot analysis prepared as described below. DNA was precipitated with addition of potassium acetate, centrifugation, and addition of isopropanol to the supernatant and then used as a template for PCR. The forward primer for PCR was 5′-CAAAAGACCCCTCGGAGT-3′ and the reverse primer was 5′-CCACAGGATCTCAGGTA-3′, with an expected product size of 600 bp. PCR products were analyzed on 1% agarose gels stained with ethidium bromide. DNAs present in bands on the gel were extracted using the QIAquick gel extraction kit (Qiagen) and were cloned into pGEM-T-Easy (Promega). The nucleotide sequence was determined at the sequencing facility at the Scottish Crop Research Institute.

**Extraction of Proteins and Western-Blot Analysis**

For protein analysis, samples were extracted from leaf material by grinding in Tris-HCl/SDS buffer and were subjected to SDS-PAGE. Western-blot analysis was carried out following transfer to nitrocellulose membranes as described (Sambrook et al., 1989). The membranes were immuno-probed with antibodies against P6 and the viral coat protein of CaMV. The antibody to P6 was raised against a synthetic peptide that comprised the 10 amino acids on the C terminus of P6 (Schoelz et al., 1991), whereas the antibody to the viral coat protein were generated against CaMV virions (Anderson et al., 1991).

**Electron Microscopy**

Immune-specific electron microscopy was performed on tissues of specific transgenic plant lines testing positive for episomal replication using immunocapture onto antibody-labeled grids to confirm the presence of spherical particles of size appropriate for CaMV (50 nm). These particles were observed for samples E316-6, E316-10, and E317-11. Anti-CaMV particle antibody was used to coat the electron microscope grids. Virus extracts were made using phosphate buffer, pH 6.5, and a standard antibody-coated grid protocol was used. The grids were stained with ammonium molybdate, pH 6.5, or 2% uranyl acetate and viewed with the electron microscope (Phillips CM10).

**Recovery of Infectious CaMV Virions and Viral DNA from Turnip Plants**

Plants of transgenic lines 316 and 318 were grown for 45 d, and the Arabidopsis plants were bulked for isolation of CaMV virions. CaMV virions were partially purified as described previously (Schoelz et al., 1986) and used for inoculation of turnip plants. Turnip plants were evaluated for CaMV infection by ELISA as described (Schoelz et al., 1986) using polyclonal antibodies directed against CaMV virions (Anderson et al., 1991), and viral DNA was isolated from individual turnip plants as described by Gardner et al. (1981). CaMV DNA isolated from individual turnip plants was examined on 1% agarose gels stained with ethidium bromide. PCR was used to determine nucleotide sequence rearrangements that occurred within gene II of the infectious viral DNA. The forward PCR primer was 5′-CCGGGATCTCAGTCCTCTCAGT-3′ and the reverse primer was 5′-TAGGA- TTTGCCGATCTCAGT-3′. The PCR amplified DNA was subsequently cloned into pGEM-T-Easy (Promega), and the nucleotide sequence was determined at the DNA Sequencing Core at the University of Missouri in Columbia.

**Statistical Treatment of Data**

The statistical analysis for episomal replication was carried out using a generalized linear model with a logit link and dispersion estimation. Data analysis was done using the program GenStat Release 12.2 (VSN International).

**Supplemental Data**

The following materials are available in the online version of this article.

**Supplemental Fig. S1.** Verification of single copy insertion of the CaMV genome into the Arabidopsis genome.
Recombination and Episomal Replication

Supplemental Figure S2. Sequence analysis of PCR products of the CaMV gene II through the gene III junction region of episomally replicating CaMV DNAs.

Supplemental Figure S3. Sequence analysis of PCR products of the CaMV gene II through the gene III junction region of episomally replicating CaMV DNAs recovered from nontransgenic turnips.

ACKNOWLEDGMENTS

We acknowledge the technical advice of Tomas Canto regarding some of this work and J.W. McNicol for the statistical analysis. This paper is dedicated to the memory of Rob W. Goldhaber, whose pioneering work has inspired many.

Received December 21, 2010; accepted January 25, 2011; published January 28, 2011.

LITERATURE CITED


Cecchini E, Gong Z, Geri C, Covey SN, Milner JJ (1997) Transgenic Arabidopsis lines expressing gene VI from cauliflower mosaic virus variants exhibit a range of symptom-like phenotypes and accumulate inclusion bodies. Mol Plant Microbe Interact 10: 1094–1101


