Massive production of small RNAs from a non-coding region of Cauliflower mosaic virus in plant defense and viral counter-defense

Todd Blevins¹,², Rajendran Rajeswaran¹, Michael Aregger¹, Basanta K. Borah¹, Mikhail Schepetilnikov¹, Loïc Baerlocher³, Laurent Farinelli³, Frederick Meins Jr², Thomas Hohn¹ and Mikhail M. Pooggin¹,*

¹Institute of Botany, University of Basel, Schönbeinstrasse 6, 4056 Basel, ²Friedrich Miescher Institute for Biomedical Research, Maulbeerstrasse 66, 4058 Basel and ³Fasteris SA, Ch. du Pont-du-Centenaire 109, 1228 Plan-les-Ouates, Switzerland

Received December 21, 2010; Revised February 14, 2011; Accepted February 15, 2011

ABSTRACT

To successfully infect plants, viruses must counteract small RNA-based host defense responses. During infection of Arabidopsis, Cauliflower mosaic pararetrovirus (CaMV) is transcribed into pregenomic 35S and subgenomic 19S RNAs. The 35S RNA is both reverse transcribed and also used as an mRNA with highly structured 600 nt leader. We found that this leader region is transcribed into long sense- and antisense-RNAs and spawns a massive quantity of 21, 22 and 24 nt viral small RNAs (vsRNAs), comparable to the entire complement of host-encoded small-interfering RNAs and microRNAs. Leader-derived vsRNAs were detected bound to the Argonaute 1 (AGO1) effector protein, unlike vsRNAs from other viral regions. Only negligible amounts of leader-derived vsRNAs were bound to AGO4. Genetic evidence showed that all four Dicer-like (DCL) proteins mediate vsRNA biogenesis, whereas the RNA polymerases Pol IV, Pol V, RDR1, RDR2 and RDR6 are not required for this process. Surprisingly, CaMV titers were not increased in dcl1/2/3/4 quadruple mutants that accumulate only residual amounts of vsRNAs. Ectopic expression of CaMV leader vsRNAs from an attenuated geminivirus led to increased accumulation of this chimeric virus. Thus, massive production of leader-derived vsRNAs does not restrict viral replication but may serve as a decoy diverting the silencing machinery from viral promoter and coding regions.

INTRODUCTION

Multicellular eukaryotes have evolved conserved, small RNA (sRNA)-generating pathways that regulate gene expression and silence transposons (1,2). In plants and some animals, sRNA pathways also have crucial functions in antiviral defense (3,4). Viral small RNA (vsRNA) duplexes processed from longer double-stranded RNAs (dsRNA) feed a cycle of viral RNA degradation, which spawns silencing signals that hinder viral spread (4). As a counter-defense, viruses express silencing-suppressor proteins able to block RNA silencing mechanisms of the host. Demonstrated activities of these suppressors include dsRNA binding to block vsRNA biogenesis, sequestration of vsRNA duplexes and inhibition of Argonaute protein (AGO)-mediated cleavage of viral transcripts (5).

The diverse silencing-suppression strategies reflect redundancy and functional specialization within plant sRNA pathways (4,5). Arabidopsis expresses at least three RNA-dependent RNA polymerases (RDR1, 2 and 6) thought to convert aberrant single-stranded RNA into dsRNA (6) and four Dicer-like (DCL) enzymes that process dsRNA or structured RNA into 21–24 nt sRNAs (1). RDR2 is required for the production of endogenous 24 nt, small-interfering RNA (siRNAs) that direct DNA methylation in plants (7). RDR6 generates...
precursors for 21 nt trans-acting siRNAs that control the timing of organ development (1). During RNA virus infection, RDR1 and RDR6 have functions in the biogenesis or amplification of viral dsRNA (8–11), while DCL4 and DCL2 act to process this dsRNA into 21 nt and 22 nt vsRNAs, respectively (12). Arabidopsis mutants deficient for multiple DCL proteins are hypersusceptible to RNA viruses (12–14). The biogenesis of vsRNAs accompanying DNA virus infection is even more complex. In response to geminivirus infection, three Arabidopsis DCLs produce distinct size-classes of vsRNAs and during infection by the pararetrovirus Cauliflower mosaic virus (CaMV), four DCL enzymes are implicated, including DCL1 (14,15). DCL1 normally functions to excise plant miRNAs from hairpin RNA precursors. The particular mechanism of CaMV replication and structured RNA encoded in its genome may explain the involvement of all four DCLs (14,16).

The CaMV genome is transcribed into a more than genome length, 35S transcript in the nucleus by host RNA polymerase II (17). A shorter, 19S RNA transcript, co-terminal with the 35S RNA, is also produced during CaMV infection. It encodes the multifunctional transactivator/viroplasmin (TAV) protein, which is a weak silencing suppressor (18–20). The 35S RNA functions as both a polycistronic mRNA and as the template for reverse transcription (RT) required for viral genome replication (21). These functions are controlled by the 600 nt leader sequence that folds into an elongated hairpin (22–25). Interestingly, the leader region was previously identified as a rich source of vsRNAs (14,16).

We performed a global analysis of vsRNA populations in CaMV-infected Arabidopsis mutants and evaluated the impact of these vsRNAs on virus replication. Surprisingly, dcl-mutants that exhibit only minimal vsRNA accumulation did not show hypersusceptibility to CaMV infection. Massive vsRNA generation from the CaMV leader region, biased association of leader-derived vsRNAs with the AGO1 effector protein and the ability of this 600-bp non-coding sequence to confer enhanced infectivity to a chimeric geminivirus suggest that it plays a role in viral counter-defense.

MATERIALS AND METHODS

Arabidopsis mutant lines and virus infection

The drd1-6, drd2-4, and drd3-7 (26), nrpd2a-1 (27), rdr2-1 (7) and rdr6-15 (28) were described previously. Homozygous rdr2 rdr6 was obtained in the F2 generation of a cross between rdr2-1 and rdr6-15. Homozygous rdr1 rdr2 rdr6 was obtained in the F2 generation of a cross between rdr1-1 (7) and the abovementioned rdr2 rdr6 line. Quadruple mutant del1/2/3/4 material was obtained by crossing either del1-8 (sin1-2) (29) or del1-9 (cafl) (30) to the triple mutant del2/3/4 (14). del1-9 used in the above crosses had been introgressed into Col-0 from Ler/Ws backgrounds by 5× outcrossing. After crossing del1 to del2/3/4, F2 individuals of genotype del1 (+/−) del2 (−/−) del3 (−/−) del4 (−/−) were selected and self-fertilized. About 25% of resulting F3 progeny were quadruple homozygous.

Plant growth and inoculation with CaMV strain CM1841 were done as described previously (14). Infected plants were harvested 1-month post-inoculation, unless otherwise stated. Virus DNA titers were measured by semi-quantitative PCR as described (14) and by real-time quantitative PCR (qPCR) for viral DNA and qRT–PCR for polyadenylated viral RNAs. For qRT–PCR, cDNA was synthesized from 2 μg total RNA using 100 pmol of d(T)16 primer. This mixture was heated to 70°C for 10 min and chilled on ice for 5 min. Then, the following reagents were added: 4 μl of 5× first-strand synthesis buffer [250 mM Tris–HCl (pH 8.3), 375 mM KCl, 15 mM MgCl2, 0.1 M DTT], 2 μl 0.1 M DTT, 1 μl 10 mM dNTP mix and 1 μl (200 U) Superscript III reverse transcriptase (Invitrogen). The reaction was incubated at 50°C for 1 h and stopped by heating at 95°C for 5 min. One-tenth volume of the reaction mix was used for real-time PCR. Real-time PCR analysis with SYBR® Green (Applied Biosystems, UK) was carried out using the Applied Biosystems 7500 system. qPCR and qRT–PCR primers (Supplementary Table S3) specific for the virus (the CaMV leader and the TAV gene or the CaLCuV AC4 gene) and the internal control gene (ACT2 for qRT–PCR or 18S rDNA for qPCR) were designed using Beacon designer 2 software (PREMIER Biosoft International).

Small RNA analysis by Illumina sequencing

Total RNA was isolated from pools of three plants using the Trizol method as described previously (14). Ten micrograms of total RNA for each sample from mock-inoculated and CaMV-infected wild-type or mutant plants were taken for preparation of sRNA libraries following Illumina’s modified protocol for the sRNA library construction kit. The 19–30 nt RNA fraction from total RNA samples was purified on a 15% TBE–Urea acrylamide gel. A 5’-adenylated single-stranded adapter was first ligated to the 3’-end of the RNA using T4 RNA ligase without ATP followed by a second single-stranded adapter ligated at the 5’-end of the RNA using T4 RNA ligase in the presence of ATP. The resulting products were purified on a 10% TBE–Urea acrylamide gel before performing the cDNA synthesis and PCR amplification. The resulting libraries were sequenced on an Illumina Genome Analyzer following the manufacturer’s protocol.

After trimming the adaptor sequences, the datasets of all and unique reads were mapped to the reference genomes of Arabidopsis thaliana Col-0 (NCBI build 8.1) and CaMV (Genbank V00140) using the Mapping and Assembly with Qualities (MAQ) program (version 0.7.1) with a maximum of two mismatches to the reference sequences allowed. Reads mapping to several positions on the references with the same ‘mapping quality’ (i.e. number of mismatches and quality of the bases generating the mismatches) were assigned at random to one of them and attributed a ‘zero’ mapping quality. To account for the circular virus genome, reads that were not mapped on
the references were retrieved and mapped with the MAQ software on the viral junction, built from the last and first 50 bases of the viral sequence. The total reads that were mapped to the reference genomes with zero mismatches constituted up to ~90% in the libraries of a high coverage and quality run (‘Col-0 mock’, ‘Col-0 CaMV’, ‘dcl1/2/6 CaMV’ and ‘dcl1-8 CaMV’) and ~50% in the libraries of a lower coverage and quality run (‘Col-0 CaMV*’ and ‘dcl2/3/4 CaMV**’) and were taken for further analysis. For each reference genome and each sRNA size-class (20–25 nt), we counted once for all reads mapped and once for all reads mapped with a minimum quality of 10 (which implies uniquely mapped reads) the number of matches starting at this position in reverse (last base of the read) and forward (first base of the read) orientations.

**RNA blot hybridization**

sRNA blot hybridization analysis was performed as in Blevins et al. (14) using short DNA oligonucleotide probes listed in Supplementary Table S3.

**8S RNA mapping by cRT–PCR**

Circularization reverse transcription–PCR (cRT–PCR) was carried out as described previously (31). Briefly, decapping was performed on 10 μg of total RNA using tobacco acid pyrophosphatase (Epigen Technologies, Madison, WI, USA) in the presence of RNase inhibitor RNasin (Promega, Madison, WI, USA). After chloroform extraction, RNA was precipitated with ethanol and circularized using T4 RNA ligase 1 (NEB) in the presence of RNase inhibitor (Roche). Following extraction with chloroform, ligation products were precipitated with ethanol. Circular RNA was reverse transcribed using SuperScript III reverse transcriptase (Invitrogen) with either L94 as or L410_s primer (Supplementary Table S3). The cDNA synthesized with the L410_s primer was PCR amplified using Taq DNA polymerase (NEB) with the L410_s primer together with L13_as or L94_as primers and cDNA synthesized with the L94_as primer was PCR amplified with the L94_as primer together with L410_s or L514_s primers (Supplementary Table S3). PCR products were analyzed by agarose gel electrophoresis (Supplementary Figure S4), excised from the gel, cloned in pGEM-T Easy vector (Promega) and sequenced. In the cRT–PCR experiment performed without pyrophosphatase treatment, 10 μg of total RNA was directly taken for ligation, while other parameters remained the same.

**Immunopurification of AGO proteins and associated sRNAs**

This experiment followed an established protocol (32). Briefly, *Arabidopsis* tissue from mock-infected and CaMV-infected plants was ground in liquid nitrogen and incubated with ice-cold extraction buffer [50 mM Tris–HCl (pH 7.5), 300 mM NaCl, 10 mM MgCl2, 0.5% NP-40, 5 mM dithiothreitol (DTT), complete protease inhibitor (Roche) and 10 μM MG-132 (Sigma)] with agitation for 30 min at 4°C. Insoluble material was removed by centrifugation for 15 min at 15 300 rpm at 4°C. Lysate was pre-cleaned by incubation with 30 μl of protein A-agarose beads (Roche) at 4°C for 1 h. The 1 mL of supernatant was then incubated overnight at 4°C with either 5 μl of normal rabbit serum (NRS, Sigma), or 5 μl (5 μg IgG) of anti-AGO1 or anti-AGO4 sera (Agisera) pre-bound to 30 μl protein A-agarose beads. Immunoprecipitates were washed three times with the extraction buffer. Immunoprecipitated RNA was purified using the Trizol method and analyzed by RNA blot hybridization as described above. Aliquots of precipitated AGO proteins were analyzed by western blot.

**Construction and testing of the CaLCuV-CaMV leader chimeric virus**

The 612-bp CaMV leader fragment was PCR amplified with CamvL_Xho_s and CamvL_Xho_s primers (Supplementary Table S3), trimmed with XhoI (the site imbedded into each primer) and cloned into the XhoI site of the Cabbage leaf curl virus vector pMTCbLCVA.007 (33) in the antisense orientation. The resulting chimeric virus CaLCuV-CaMV was tested for infectivity on *Arabidopsis* using biolistic delivery as described by Blevins et al. (14). As a control, we used the CaLCuV vector carrying a complete, green fluorescent protein (GFP) coding sequence of 714 bp inserted between the CaLCuV vector carrying a complete, green fluorescent protein (GFP) coding sequence of 714 bp inserted between the CaLCuV vector and the CaLCuV-GFP. Accumulation of viral DNA in infected plants was measured by semi-quantitative PCR as described earlier (14) and confirmed by qPCR; for PCR primers see Supplementary Table S3. The vsRNAs derived from the CaMV and CaLCuV sequences were analyzed by RNA blot hybridization as described earlier (14) using DNA oligonucleotide probes (Supplementary Table S3).

**RESULTS AND DISCUSSION**

By ethidium bromide (EtBr) staining of size-separated RNA, we detected massive quantities of 21–24 nt RNAs in CaMV-infected *Arabidopsis*. CaMV-derived sRNAs appeared to be more abundant in infected plants than the complement of host sRNAs detected in the mock-inoculated control (Figure 1A). In contrast, EtBr staining of RNA from *Cabbage leaf curl geminivirus* or *Oilseed rape mosaic tobamovirus* infections yielded scarcely detectable levels of sRNAs (data not shown). Abundant amounts of sRNAs were still detected in CaMV-infected plants deficient in DCL1, which is required for host miRNA biogenesis (dcl1-8, a hypomorphic mutant). They were also detected in infected plants deficient in RDR2, RDR6 or both RDR6 and RDR2 required for host siRNA biogenesis (null mutants rdr2-1, rdr6-15 and double mutant rdr2/6) (Figure 1A). Thus, a large fraction of sRNAs detected by EtBr staining are likely to be vsRNAs of CaMV origin.
The 24-nt sRNA signal was abolished in the null mutant dcl3-1, confirming that the biogenesis of this size-class of viral and host sRNAs is DCL3 dependent (7,14).

To survey genome-wide interactions of CaMV with the plant RNA silencing machinery we sequenced sRNA populations from virus-infected wild-type and mutant plants using Illumina technology (www.illumina.com). The procedure selected for sRNAs with 5'-phosphate and 3'-hydroxyl groups, i.e. DCL activity products. Consistent with earlier reports (34,35), 81% of the 5,495,540 sRNA reads from mock-inoculated seedlings fell into the 20–25 nt range, of which 75% mapped to the Arabidopsis genome with zero mismatches; 24 nt and 21 nt reads were predominant (Figure 2A and Supplementary Table S1). Likewise, 81% of the 1,470,760 reads from CaMV-infected plants fell into the 20–25 nt range, of which 40% mapped perfectly to the Arabidopsis genome (~120 Mb) and 39% mapped perfectly to the CaMV genome (~8 kb) (Figure 1B, Supplementary Table S1). Among 20–25 nt vsRNAs, the 24 nt fraction was largest (47%), while the 21 and 22 nt fractions were second (27%) and third (14%) largest in abundance (Figure 2B), respectively. This size-class distribution agrees with our results for the relative abundance of CaMV-derived sRNAs detected by blot hybridization (14). To summarize the gel analysis and sequencing data are in agreement and show that the vsRNA pool is comparable in size to the entire complement of host miRNAs and siRNAs.

To investigate whether CaMV vsRNAs are redundantly amplified by RDR activities known to mediate sRNA biogenesis (6), we generated a triple null mutant rdr1 rdr2 rdr6 (rdr1/2/6) and sequenced sRNAs from infected rdr1/2/6 plants. In rdr1/2/6, 20–25 nt sRNAs perfectly matching the CaMV genome represented 54% of 748,199 reads, while those perfectly matching the Arabidopsis genome only represented 28% of the reads (Supplementary Table S1 and Figure 1B). The decreased percentage of host sRNAs in this mutant compared with the wild-type resulted from a 10-fold decrease in RDR2-dependent 24 nt sRNAs. In contrast to the profile for host sRNAs, the profile for vsRNAs did not change significantly in rdr1/2/6 relative to wild-type (Figure 2B and Supplementary Table S1). From this, we conclude that at the late stage of infection, CaMV-derived vsRNAs are largely produced independently of RDR1, RDR2 and RDR6. We cannot, however, exclude that CaMV vsRNAs can potentially be amplified by these RDRs at the earlier stages of infection or under certain conditions. In RNA virus infections of Arabidopsis, vsRNAs are amplified by
RDR1 and/or RDR6; this amplification becomes most evident when a viral silencing suppressor gene is inactivated (8–11). The ability of the CaMV TAV protein to interfere with host RDR6-dependent sRNA pathways (19,20) could account for the lack of RDR6-dependent vsRNAs in CaMV-infected plants. Likewise, TAV may also inhibit production of RDR1-dependent vsRNAs. It would be interesting to investigate whether a CaMV mutant deficient for the silencing suppressor activity of TAV can generate RDR-dependent vsRNAs. However, the suppressor activity would first need to be genetically uncoupled
from other TAV functions that are essential for viral replication.

We also sequenced sRNAs from a hypomorphic mutant deficient for DCL1 (dcl1-8). In CaMV-infected dcl1-8, viral reads outnumbered Arabidopsis genome-derived reads due to the reduction in host miRNA accumulation (Figure 1B and Supplementary Table S1). Of 20–25 nt vsRNAs in dcl1-8, the fraction of 21 nt reads (17%) was significantly lower than that in wild-type plants (27%) (Figure 2B). This confirms our earlier finding that DCL1 contributes to the biogenesis of 21 nt vsRNAs (14). The role of DCL1 in vsRNA biogenesis is also supported by the sRNA sequence profiles we obtained for CaMV-infected dcl2 dcl3 dcl4 triple null mutant (dcl2/3/4) plants. The relative levels of 22 and 24 nt vsRNAs were drastically reduced, whereas the relative level of 21 nt vsRNA was increased compared to wild-type plants (Figure 2B and Supplementary Table S1).

The distribution of 20–25 nt vsRNA reads across the 8031-bp CaMV genome was plotted at the single-nucleotide scale (Figure 2C, Supplementary data SD1, Supplementary Figures S1 and S2). These maps show a biased distribution in terms of strand orientation and position. First, 62% of reads from infected wild-type plants matched the viral transcript polarity (i.e. sense polarity). The sense-strand bias was also observed in the mutant lines (Supplementary Table S1 and Supplementary Figure S1). Second, both in wild-type and mutant plants, a majority of viral reads, 65–82%, aligned to the ~600-bp leader region (Figure 2C, Supplementary Figure S1 and Supplementary Table S1). Leader-specific reads are distributed unevenly: several 50–80 nt ‘hot spots’ give rise to most vsRNAs of either polarity. We found no correlation of these hot spots with strong and regular secondary structures compatible with DCL binding and processing. If unique vsRNA sequences are considered (Supplementary Table S2), they densely tile across the entire leader region in both polarities (Supplementary Figure S2). Within hot spots, almost every nucleotide position marks the 5’-end of 21, 22 and 24 nt species, each represented with up to several thousand reads (Supplementary Figure S2 and Supplementary data SD1). Similar hot spots were found in independent samples from wild-type plants and in samples from rdr1/2/6 and dcl2/3/4 plants. In dcl1-8, the hot spots were shifted slightly (Supplementary Figures S1 and S2 and Supplementary data SD1). Thus, in the absence of a fully functional DCL1, other DCLs process vsRNA precursors with somewhat altered sequence specificity.

We validated the deep-sequencing data by RNA blot hybridization using 36-ntArabidopsis oligonucleotide probes specific to CaMV (Supplementary Table S3) and confirmed a reproducible pattern of vsRNA hot spots covering the leader region in both polarities (Supplementary Figure S3). Both approaches reveal the hot spots in the same zones within the leader region. In contrast, only very low levels of vsRNAs could be detected with probes specific to the 35S promoter region upstream of the leader (Supplementary Figure S3), which also agrees with our deep-sequencing data (Supplementary data SD1). Genetic requirements for vsRNA biogenesis were found to be similar throughout the leader region (Supplementary Figure S3). Together, our blot hybridization and sequencing data indicate that each of the four DCLs processes a leader-length dsRNA precursor or, alternatively, multiple overlapping dsRNAs derived from the entire leader. This dsRNA processing appears to be initiated internally and sporadically, but with some preference for the hot spots. It is also possible that differential vsRNA stability contributes to hot spot formation. No phasing could be discerned that would resemble processing of trans-acting siRNAs from RDR6-dependent dsRNA (28).

The robust systemic infection and viral DNA accumulation that we observed in spite of massive vsRNA production was surprising. In the case of RNA viruses, strong suppressor proteins such as a potyviral HC-Pro are known to inactivate vsRNAs (5,11). However, this is not sufficient to explain our observations for CaMV, which does not encode a strong suppressor protein. The CaMV TAV protein is a weak silencing suppressor that perturbs RDR6-dependent pathways associated with sRNA amplification (18–20). Furthermore, CaMV infection does not suppress transgene silencing triggered by RDR6-independent dsRNA (20). To test whether massively produced CaMV vsRNAs have an antiviral function, we generated two dcl1 dcl2 dcl3 dcl4 (dcl1/2/3/4) quadruple mutants each carrying a hypomorphic allele of DCL1 (dcl1-8 or dcl1-9) and null alleles of DCL2, DCL3 and DCLA. RNA blot hybridization showed that, compared to triple mutant dcl2/3/4, both quadruple mutants accumulated only residual 21 nt vsRNAs (Figure 3A and Supplementary Figure S3). The truncated DCL1 protein encoded by dcl1-9 mutant alleles can process miRNA precursors, but less efficiently and precisely than wild-type DCL1 (36). This leads to destabilization of most miRNAs, including miR173 but excluding miR168 (37): we used these two miRNAs as controls (Figure 3A).

RNA viruses have been shown to accumulate at higher levels in dcl1/2/3/4 plants (13). In contrast, titers of CaMV DNA and polyadenylated 35S and 19S transcripts were not affected substantially in dcl1/2/3/4 or dcl2/3/4 as measured by real-time qPCR and qRT–PCR, respectively (Figure 3B), and by other methods (Supplementary Figure S4). This extends our finding that CaMV DNA accumulation is not dramatically affected in any of the four single dcl mutants (14), and is in contrast to Moisiard and Voinnet (16) who observed a strong increase in accumulation of CaMV 35S RNA and coat protein in dcl2/3/4 compared to wild-type plants. The fact that distinct strains of CaMV were studied—CM1841 in our case and JI in (16)—may account for this discrepancy. In different CaMV strains, the TAV protein is known to be a determinant of host range and symptom severity. TAV function in suppressing RNA silencing-based host defenses may also be strain dependent.

Stability and function of Arabidopsis sRNAs depend on their incorporation into an RNA-induced silencing complex (RISC) through binding to AGO proteins that sort sRNAs based on their 5’-nt and size (38–40). Thus, a majority of 21 nt miRNAs have 5’ uridine (5’U)-termini and associate with AGO1, whereas a large fraction of
24 nt siRNAs have 5’ adenosine (5’A)-termini and associate with AGO4. In line with this, in our data sets, we found that 21 nt host sRNAs predominantly have 5’U (78% of the total 21 nt reads; Supplementary Table S1), whereas 24 nt host sRNAs predominantly have 5’A (60% of the total 24 nt reads). In contrast, 29% of 21 nt viral sRNAs have 5’U-termini and 41% have 5’A-termini (Supplementary Table S1), which would predict their association with AGO1 and AGO2, respectively (37). We previously observed an up-regulation of AGO2 and AGO7 but not other AGO genes in CaMV-infected plants (20). Because AGO7 interacts specifically with miR390 (39), increased availability of AGO2 protein is more likely to account for preferential accumulation of 5’A vsRNAs. AGO2 binds vsRNAs derived from Cucumber mosaic virus, which is an RNA virus (40), but no functions of AGO2 in antiviral defense or silencing of host genes are known.

In principle, CaMV derived 24 nt vsRNAs with 5’A-termini (35% of the total 24 nt reads; Supplementary Table S1) could associate with AGO4 to mediate viral DNA methylation. In addition to AGO4, the host siRNA-directed de novo methylation pathway depends on the RNA polymerases Pol IV and Pol V and the chromatin remodeling factor DRD1 (26,41). Our earlier studies showed that deficiencies in AGO4 and Pol IV do not affect CaMV DNA accumulation and only slightly affect vsRNA accumulation (14); we confirmed this result here using ddr2-4, a different loss-of-function mutant for Pol IV (Supplementary Figure S5). Similarly, we did not find drastic changes in CaMV DNA and vsRNA accumulation caused by deficiencies in Pol V or DRD1 (ddr3-7 and ddr1-6, respectively) (Supplementary Figure S5). These findings and the fact that CaMV DNA is actively transcribed throughout the infection cycle (42) strongly suggest that the de novo methylation pathway is not a major player in CaMV silencing. In line with this, we found very little vsRNAs derived from the 35S promoter region in CaMV-infected plants (Supplementary Figure S3 and our Illumina data). Moreover, only negligible amounts of leader-derived vsRNAs and no vsRNAs from promoter and coding regions were detected bound to the AGO4 protein in our immunoprecipitation (IP) assay with native antibodies for AGO4 (Supplementary Figure S5).

The massive production of vsRNAs of both polarities from the 35S RNA leader region suggests that sense and antisense transcripts covering just this region could be their precursors. Such sense RNA was described earlier as a viral 8S RNA of unknown function (17). By using cRT–PCR (31), we isolated multiple species of 8S RNA from CaMV-infected wild-type or dcl2/3/4 Arabidopsis plants (Figure 2D and Supplementary Figure S6). The 8S RNA starts at either of two consecutive AC-motifs, the first of which was reported as the common start site of 35S and 8S RNAs (17), but ends heterogeneously close to the primer binding site of RT located at the end of the leader sequence (Figure 2D). At this position, the virion CaMV DNA has a discontinuity of the minus (antisense) strand marking the start and the end of RT (21). Abrupt termination of 35S promoter-directed transcription close
to the discontinuity of the unsealed DNA template in the nucleus might lead to accumulation of sense 8S RNA. Alternatively, sense 8S RNA that folds into an extended hairpin structure (see below) may be processed from 35S RNA by ‘Drosha’-like activity of DCL1 (36) as suggested previously (16).

Using cRT–PCR with reverse primers, we found multiple antisense copies of 8S RNA. These antisense transcripts start around the positions of the 3'-ends of sense 8S RNAs and end precisely at the sense 8S RNA start site, essentially mirroring sense 8S RNAs (Figure 2D and Supplementary Figure S6). We propose that sense and antisense 8S RNAs form dsRNAs that are then processed by DCLs to generate vsRNAs. Both sense and antisense 8S RNAs lack a poly(A) tail and are sensitive to pyrophosphatase (Supplementary Figure S6), suggesting that they are either capped or 5'-triphosphorylated. The heterogeneity of the 5'-termini and the lack of adequately positioned promoter consensus sequences in antisense orientation argue against production of antisense 8S RNA on a CaMV DNA template. On the other hand, our mapping data are consistent with a model, in which 8S RNA sense transcripts are converted into dsRNAs by an RNA-directed RNA polymerase activity. This polymerase would drop off at the 5'-end of the template without adding extra non-template nucleotides. Our genetic evidence excludes RDR1, RDR2 and RDR6 as well as Pol IV and Pol V as this enzyme, because wild-type amounts of CaMV leader-derived vsRNAs of both polarities accumulated in rdr1/2/6 (see above) as well as in the loss-of-function mutants for Pol IV [(14), (Supplementary Figure S5; drd2-7 and nrpd2a-1)] and Pol V (Supplementary Figure S5; drd3-7 and nrpd2a-1). Formally, we cannot exclude a possibility that three additional genes of the Arabidopsis RDR family, for which no function was reported so far (6), may be involved. Another, more likely, candidate would be Pol II. Pol II is known to replicate animal hepatitis delta virus and plant RNA viroids, circular single-stranded RNAs that form extensive secondary structures (43). The CaMV leader sequence forms such a viroid-like structure (23). Furthermore, Lipardi and Paterson (44) have reported that the largest subunit of Drosophila Pol II has an RDR activity and participates in RNA silencing and transposon suppression.

Despite the hindrance to translation initiation on viral 35S RNA due to the presence of a highly structured leader, the approximate size and secondary structure, but not the primary sequence of the leader region are highly conserved in diverse plant pararetroviruses (45). Furthermore, most mutations and second site reversions in the CaMV leader primary sequence have little effect on virus replication (25,46,47). These considerations and the fact that a complex shunt mechanism has evolved to allow scanning ribosomes to bypass the leader (22,25,45) suggest that there is a selective advantage for a long, highly structured leader.

One possibility is that the complexity and length of the leader are required for the function of polyadenylation, packaging and splicing signals that it contains; but, this seems unlikely because these signals require much less space and structure in other DNA viruses. Another possibility is that the vsRNAs produced from the leader region play a role in viral counter-defense. Moisissiar and Voinnet (16) have proposed that the CaMV leader forms a miRNA precursor-like structure and some leader-derived vsRNAs target host genes in a sequence-specific manner to the advantage of the virus. Our transcriptome profiling of CaMV-infected plants and of transgenic plants expressing CaMV TAV protein (20) revealed only modest downregulation of some predicted target genes, which appears to depend on CaMV TAV expression rather than vsRNAs per se (Supplementary Figure S7). The fact that the leader primary sequence is not evolutionary conserved also argues against a sequence-specific role for leader-derived vsRNAs in viral counter-defense.

More consistent with our data is a hypothesis that the 8S duplex RNA derived from the leader region acts as a silencing decoy to engage all four DCLs in massive production of vsRNAs. These highly abundant vsRNAs could prevent less abundant vsRNAs produced from other CaMV genome regions from being assembled into AGO–RISCs. Our finding that CaMV titers are not affected in the dcl1/2/3/4 mutants, in which both the host defense and the viral leader-based counter-defense are undermined, indicates that sporadic production of vsRNAs outside of the leader region does not contribute substantially to restricting virus replication. Thus, in the wild-type plants, these less abundant vsRNAs do not appear to form sufficient amounts of active AGO–RISCs to silence viral transcripts. To test this hypothesis, we analyzed vsRNAs associated with AGO1, the major AGO effector protein implicated in antiviral defense (48–51). IP of AGO1 from CaMV-infected plants with a specific native antibody (49) yielded a sample highly enriched with leader-derived vsRNAs, whereas vsRNAs derived from other CaMV genome regions were not detected (Figure 4).

AGO–RISCs charged with leader-derived vsRNAs of antisense polarity could potentially target 35S RNA. However, the leader sequence of 35S RNA forms a stable secondary structure required for both translation (25,47) and packaging/RT (24) and this structure would prevent RISC binding. In line with this, our previous studies indicate that stability and integrity of the leader secondary structure correlates with CaMV infectivity (25,46,47). Studies of viroids and viral satellites and defective-interfering RNAs have shown that although vsRNAs can be produced from such precursors (presumably during replication when they spawn dsRNAs), their secondary structures largely prevent them from being targeted by the resulting vsRNAs (10,52–55).

Massively produced vsRNAs could potentially compete with host miRNAs and siRNAs for free AGO proteins, interfering with endogenous RISC function. But our AGO-IP experiments do not support this idea: binding of miR173 to AGO1 and the AtREP2 siRNAs to AGO4 was not perturbed by CaMV infection (Figure 4 and Supplementary Figure S5). Accordingly, our previous study indicates that CaMV infection and CaMV TAV expression do not interfere with miR173-directed cleavage of
target transcripts (20). Moreover, our transcriptome profiling did not reveal a general upregulation of miRNA-targeted transcripts upon CaMV infection, except for those silenced by RDR6-dependent secondary siRNAs; the latter effect is attributable to the silencing suppressor activity of CaMV TAV protein (20). Thus AGO-RISCs programmed by host miRNAs and siRNAs do not appear to be perturbed by massive production of vsRNAs. Therefore, leader-derived 8S dsRNA/vsRNAs cannot be considered as a general decoy that interferes with all RNA silencing pathways. Instead, massive production of vsRNAs from the leader region may divert the silencing machinery from targeting the most vulnerable CaMV regions, i.e. viral promoters and unstructured coding regions of 19S and 35S transcripts.

The CaMV leader-based cis-elements and structures are essential for viral replication and therefore the leader sequence cannot be deleted from or replaced by a heterologous leader in CaMV genome without drastically affecting CaMV infectivity (25,46,47). To provide further evidence for our decoy hypothesis, we therefore expressed leader-derived vsRNAs ectopically from a heterologous DNA virus. We made use of an attenuated geminivirus slightly increases viral DNA accumulation. (A) Viral DNA titers in Arabidopsis plants infected with a CaLCuV vector carrying the CaMV leader sequence (CaLCuV-CamvL) versus the vector control virus (CaLCuV-vec) or the vector carrying a non-viral insert (CaLCuV-GFP) were measured by semi-quantitative PCR on serial dilutions (5-fold each) of total DNA isolated from pools of three to four infected or mock-inoculated plants. 18S ribosomal DNA amplification is an internal control. The titers were also confirmed by qPCR. (B) vsRNAs were analyzed by RNA blot hybridization using total RNA from pools of three to four virus-infected (CaMV, CaLCuV-vec, CaLCuV-CamvL and CaLCuV-GFP) or mock-inoculated plants. The membrane was successively hybridized with DNA oligonucleotide probes for the CaMV leader region and CaLCuV AC4 gene. Met-tRNA is shown as a loading control. Note that two independent clones of the CaLCuV-CamvL construct were analyzed.
CaMV-infected plants (Figure 5B). VsRNAs from the CaLCuV genome also accumulated, albeit at lower levels (Figure 5B). The accumulation of the CaLCuV genome-specific vsRNAs from the CaLCuV::CaMV leader chimeric virus was slightly higher than for the case of the CaLCuV empty vector and CaLCuV::GFP control virus, which correlates with their respective viral DNA titers. Thus, despite the production of CaMV leader-derived vsRNAs in addition to CaLCuV-specific vsRNAs, replication of the chimeric virus was not reduced. These data are consistent with our decoy hypothesis.

CONCLUSIONS

Here, we demonstrate that CaMV infection is associated with massive production of RDR1-/RDR2-/RDR6-independent vsRNAs from the leader region, whereas other regions of the viral genome spawn only low amounts of vsRNAs. The quantity of leader-derived vsRNAs is comparable to the entire complement of endogenous plant miRNAs and siRNAs. Illumina sequencing of vsRNAs revealed several hot spots matching both polarities of the leader region. Likely precursors of these vsRNAs are long double-stranded molecules arising from Pol II transcription of the entire leader region into sense and antisense 8S RNAs, which we detected by cRT–PCR. Genetic evidence indicates that all four DCLs process vsRNAs from their precursors. This processing occurs internally with some preference for hot spot sequences, but differential vsRNA stability may also account for hot spot formation.

CaMV titers were not elevated in *dcl*-quadruple mutants, despite vsRNA levels being greatly diminished. Thus, massive vsRNA production in wild-type plants does not appear to restrict viral replication. Furthermore, ectopic expression of CaMV leader vsRNAs from an attenuated geminivirus did not render the chimeric virus more prone to plant antiviral defenses. These results raise the possibility that in addition to expressing the silencing suppressor protein TAV that interferes with RDR-dependent amplification of siRNAs (18–20), CaMV deploys counter-defense based on decoy 8S dsRNA generated from the non-coding leader region (Figure 6). Massive production of primary vsRNAs from 8S dsRNA might divert silencing effectors from targeting CaMV promoters and unstructured transcripts outside of the leader region. AGO1 is a major effector of antiviral silencing in
plants, which explains why RNA viruses express suppressor proteins that target AGO1 for degradation, reduce AGO1 protein translation, or disrupt its RNA cleavage activity (5,48–51). Unlike RNA viruses, CaMV infection did not perturb AGO1 accumulation, its association with endogenous miRNAs or activity. The fact that leader-derived vsRNAs were associated with AGO1, whereas vsRNAs from other regions of CaMV genome were not detected in the AGO1 complexes, lends further support to our decoy model of CaMV counter-defense (Figure 6).

It will be interesting to explore whether other plant viruses use decoy RNAs to evade silencing. Studies on human adenovirus found that highly structured viral RNAs suppress the interferon-mediated antiviral defense whereas vsRNAs from other regions of CaMV genome do not inhibit antiviral mechanism. It will be interesting to explore whether other plant viruses suppress the interferon-mediated antiviral defense may be common among plant and animal viruses.

SUPPLEMENTARY DATA
Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS
We thank Nachelli Malpica and Ekaterina Gubaeva for technical assistance, Marjory Matzke for seeds of drd1-6, drd2-4 and drd3-7 and Craig Pikaard for supporting part of T.B.’s experiments; and the Novartis Research Foundation for supporting work done at the Friedrich Miescher Institute. We are grateful to Thomas Boller, Andres Wiemken and Christian Körner for hosting the group at the Botanical Institute. T.B. and M.M.P. designed research, T.B., R.R., M.A., M.S. and B.K.B. performed research. L.B., L.F., T.B., T.H. and M.M.P. analyzed data, T.B., F.M., T.H. and M.M.P. wrote the article.

FUNDING
Swiss National Science Foundation grants (31003A_122469 to T.H. and M.M.P., 31003A_127514 to M.M.P.); European Molecular Biology Organization (EMBO short term fellowship ASTF 169.00-2010 to M.S.); European Commission (a Marie Curie fellowship PIIF-237493-SUPRA to R.R.). Funding for open access charge: Friedrich Miescher Institute.

Conflict of interest statement. None declared.

REFERENCES
11 Nucleic Acids Research, 2011


