JOURNAL OF VIROLOGY, Apr. 1988, p. 1460–1463
0022-538X/88/041460-04$02.00/0

Vol. 62, No. 4

A Viable Mutation in Cauliflower Mosaic Virus, a Retroviruslike Plant Virus, Separates Its Capsid Protein and Polymerase Genes

J. PENSWICK, R. HÜBLER, AND T. HOHN*

Friedrich-Miescher-Institut, P.O. Box 2543, CH-4002 Basel, Switzerland

Received 20 July 1987/Accepted 23 December 1987

A viable strain of cauliflower mosaic virus is described which arose by illegitimate recombination of two lethal parents. In this strain, the normally overlapping open reading frames IV and V, corresponding to the retrovirus *gag* and *pol* genes, are separated by a short intergenic region, suggesting that in this virus and in contrast to retroviruses, fusion of *gag* and *pol* gene products is not obligatory.

Cauliflower mosaic virus (CaMV) is a plant DNA virus that replicates its genome via reverse transcription of an RNA with 180-nucleotide terminal repeats (for reviews, see references 12, 14, and 20 and J. M. Bonneville, T. Hohn, and P. Pfeiffer, *in* E. Domingo, P. Ahlquist, and J. J. Holland, ed., *RNA genetics*, II, in press). A portion of its genome, encompassing open reading frames (ORFs) IV, V, and VI, shows homology in organization, sequence, and function to the retrovirus *gag-pol-env* coding region (7, 20). ORFs IV and V overlap for a short distance, much like the *gag* and *pol* genes of avian and some other retroviruses. In these retroviruses, the *pol* gene is translated as a large *gag/pol* fusion protein produced by a frameshift within the overlapping region (16). Some other retroviruses produce *gag/pol* fusion protein by stop codon suppression (e.g., see reference 30).

Like retroviruses and unlike proper double-stranded DNA viruses, CaMV has a very high recombination rate (2, 6, 10, 15, 19), and as in retroviruses (reference 29, section 7.VI.B.6), it is tempting to correlate this rate to the peculiarities of the retro replication cycle. It would thus be best explained by assuming replicative recombination and copy choice (5), i.e., switches of the nascent DNA strands from one parent template to the other, taking into account that the reverse transcriptase has to perform template switches at certain sites obligatorily (11). Also, deletions (17) and duplications within the genome can be interpreted as being caused by recombination processes, with the template switches occurring at sites of little homology.

For CaMV, an additional type of recombination has been observed as a consequence of mechanical inoculation with DNA linearized by restriction: viral DNA linearized at homologous sites can ligate in plant cells to form heteropolymers (9, 28) from which hybrid viral genomic RNA can be produced. This mechanism even functions with viral subgenomic restriction fragments cloned separately (19).

Here we describe a viable recombinant apparently produced by a combination of both mechanisms. This recombinant has a duplication of the ORF IV-V junction. The mere existence of this recombinant indicates that CaMV follows a strategy for *pol* translation different from that of the avian retroviruses discussed above.

Recombination experiments have been performed by coinoculation of host plants (*Brassica rapa*) with pairs of cloned and (in vitro) mutagenized CaMV genomes after their uncoupling from the bacterial vector at the *Sal*I cloning site. One parent used, Ca422 (Fig. 1), was constructed as a viral vector to accept a payload of 1,000 base pairs (bp) by deletion of as much of its sequence as we knew to be nonessential. Removal of ORFs II and VII in this strain does not affect infectivity by mechanical inoculation, but a third deletion of 147 bp in the untranslated leader region causes an increase of the latency period from the normal 2 weeks to 12 weeks. The reason for this defect is under study. The other parent, Ca169 (Fig. 1), is a lethal linker insertion (frameshift) mutant within the presumptive polymerase region of ORF V. The two mutants originate from different backgrounds (CM4.184 [13] and CaMV.JI [3]) and hence show some restriction enzyme site (Fig. 1) and sequence (see Fig. 2 and 3) polymorphism in addition to the differences caused by the mutations.

One of a few recombinants (viable, short latency) between Ca422 and Ca169 attracted our attention because of an unusual restriction pattern and was therefore studied in more detail. Its DNA was isolated immediately after appearance of systemic symptoms and recloned into the bacterial vector pUC8. Ten independently obtained single clones analyzed showed identical restriction patterns for all restriction markers polymorphic for the original parents (Fig. 1) and the frequent *Cla*I sites, and these corresponded to the bulk of the viral DNA sample from which the clones were derived. All clones were infectious when excised from the vector, yielding systemic symptoms after short latency. The critical portions of one of them, clone Ca534, were sequenced (Fig. 2 and 3), and Ca534 turned out to be a hybrid strain, with the bulk of its sequence originating from Ca169 and a minor portion originating from Ca422. The minor portion included the wild-type allele of ORF V, and the major portion included the wild-type allele of the leader. The sequence polymorphism of the parent strains allowed us to determine the crossover points of the recombinant. The one crossover point is located within a 60-bp sequence around the *Sal*I restriction site (Fig. 2). It cannot be mapped more precisely because of a lack of polymorphic markers within this stretch. However, since the *Sal*I site had been used for cloning and the inoculated DNA species had been excised from the vector plasmids at this site, the crossover point probably reflects heterodimer (polymer) formation of the original inoculum. The other crossover point can be mapped exactly to single nonhomologous nucleotides in the ORF IV-V border region, resulting in a 182-bp duplication (Fig. 3). The two versions of this duplication clearly originate from one parent each, as can be seen from the sequence polymorphism concerning five specific base pairs in this region (marked in Fig. 3).

* Corresponding author.

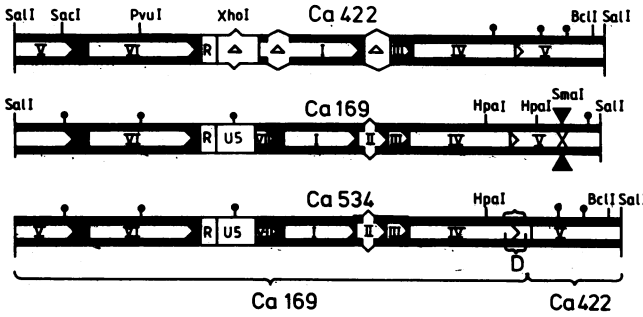

FIG. 1. Physical maps of the CaMV strains described (wild type and derivatives). The maps, in fact circular, are presented in their *Sal*I-linearized DNA form, as obtained from DNA clones and used to inoculate plants. The main seven ORFs (I to VII; ORF V is interrupted by the linearization) are shown, and the regions that correspond to retrovirus, R and U5, are indicated. R in this case is defined as the terminal repeat of the CaMV genomic RNA, and U5 is the sequence between R and the primer-binding site. The *Sal*I sites providing the termini of the inoculating DNA and the sites involved in restriction enzyme polymorphism used to distinguish the strains are shown by the conventional symbols where present and by the symbol ▼ where absent. Ca422 contains three deletion (△) mutations, and Ca169 contains a linker insertion mutation (⧗). Ca534 is the recombinant described in detail. The duplication (D) is indicated, as is the origin of the sequence from either parent (Ca422 and Ca169).

Since the duplication provides a large recombination target, it should disappear fast from the virus DNA population if it were of any disadvantage for viral fitness. This was not the case; after reinoculation of cloned Ca534 DNA to five different plants and waiting for 3 weeks after appearance of

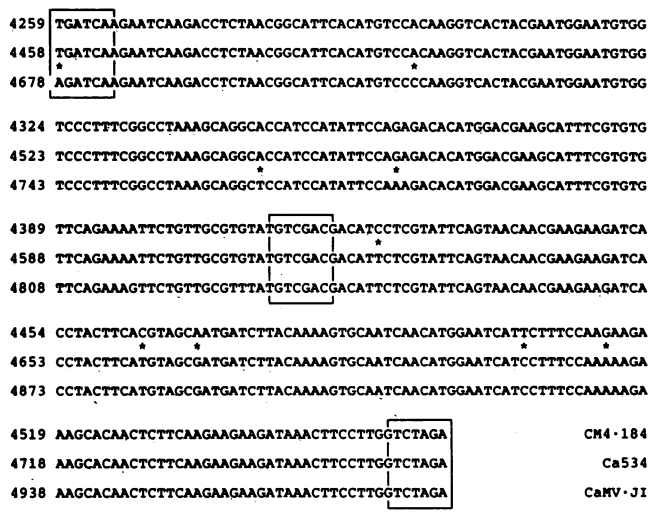

FIG. 2. Nucleotide sequence of a *Bcl*I-*Xba*I restriction fragment of Ca534 in comparison with those of CaMV.CM4-184 and CaMV.JI. *Bcl*I-*Sal*I and *Sal*I-*Xba*I subclones were sequenced. Locations of the restriction sites are boxed (5′ terminus, *Bcl*I; middle, *Sal*I; 3′ terminus, *Xba*I). Nucleotide polymorphisms at four positions upstream from the *Sal*I site and at five positions downstream from it are marked by asterisks and suggest a crossover point around the *Sal*I site. Sequencing (26) was performed in both directions after subcloning of the respective restriction fragments in bacteriophage M13mp18 and M13mp19 (21) with universal and reverse primers (New England BioLabs, Inc.). Reference sequences were those of CM4-184 (6, 13) and CaMV.JI, kindly provided by J. Stanley (J. Stanley and J. Davies, personal communication).

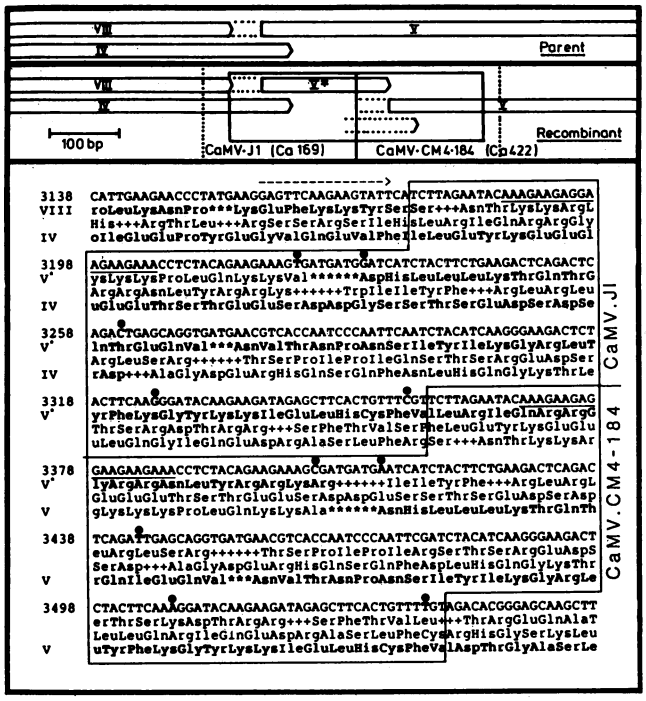

FIG. 3. Locations of ORFs in the region of duplication in Ca534 compared with those in the wild type. The duplicated stretch of 182 bp is boxed. The part between vertical dotted lines is shown in detail below. Codons in all three phases are given, with *** being start codons and +++ indicating stop codons. ORFs VIII (3′-terminal portion), IV (3′-terminal portion), V (5′-terminal portion), and V* (a short ORF with a 5′ terminus similar to that of V but a new 3′ terminus from another reading phase) are indicated. The border between the CaMV.CM4-184- and CaMV.JI-derived sequences is given exactly. Nucleotide polymorphism of the repeated sequences at five positions are marked with ●, and the polypurine stretch mentioned in the text is underlined. Sequencing was performed in both directions as described in the legend to Fig. 2, from a *Hin*dIII fragment (base pairs 2845 to 3554) and a *Hin*dIII-*Pst*I fragment (base pairs 2845 to 3024) by using the universal primer and, in addition, a synthetic primer (arrow). This primer was also used to sequence revertants of the deletion.

symptoms, only about 20% of virus progeny in each of the individual plants had the duplication deleted. This proportion increased to 50% after reinoculation of new plants with sap from an infected leaf from the first inoculation and then remained constant for several additional cycles of reinoculation. This contrasts with the fast removal of inserts at other parts of the CaMV genome when they interfere with viral functions (1, 4, 8, 23; K. Sieg and B. Gronenborn, Abstr. NATO Adv. Course Stud. Inst. Adv. Course, C8, p. 154, 1982), even if only a few homologous base pairs are available as recombination targets. The reverted genome, after reconing, infected plants with a latency period similar to that of the genome with the duplication. This shows that the duplication is just a product of nonhomologous recombination and does not provide a selective advantage. Two of the revertants originating from different plants were analyzed by sequencing. They had the duplication exactly removed, restoring the original ORF IV-V arrangement. In both cases, the copy with the Ca.JI sequence was deleted.

We cannot draw any conclusions about how the second crossover point was created. It could have been by breaking and joining of the probable heterodimer intermediate indi-

cated in Fig. 4 (pathway a) or, more likely, by double crossover by template switching of the replicative machinery between RNA formed from the heterodimer and RNA formed from Ca169, as outlined in Fig. 4 (pathway b). Whatever the pathway, it involved a crossover at an illegitimate position that could have been avoided.

There are many other examples of CaMV genome derivatives that must have arisen by recombination at nonhomologous sites: removal of foreign genes cloned in CaMV (1; Sieg and Gronenborn, abstract); spontaneous deletion of parts of ORFs II (13) and VII (8); duplication in a clone of the CaMV Xinjing strain (25); and accumulation of a population of subgenomic satellites in infected plants (22). This might indicate a lack of precision of replicative template switching in CaMV. Duplications might then be retained in virus populations if parent genomes are of smaller than usual size, as in our case, and packaging constraints are absent. Removal of the same copy of the duplicated sequence in two independent revertants suggests that the long polypurine stretch marked in Fig. 3 facilitates recombination in both creation of the duplication and its subsequent reversion.

Production of *gag/pol* fusion proteins in retroviruses (i) spares a separate mRNA, (ii) controls relative amounts of structural and enzymatic viral proteins, and (iii) might facilitate targeting of the polymerase into the viral capsid. Because of its similarities in genome arrangement, CaMV might have been thought to follow a similar strategy and form a *gag/pol* (ORF IV-V) fusion product. However, in Ca534, ORFs IV and V are separated (Fig. 3), and this separation does not impair viability; latency and yield are similar to those of wild-type infections. An ORF IV-V fusion protein could still be produced in an altered way but then has to use multiple frameshifts (ORF IV to ORF V* and then to ORF V; Fig. 3), and such a hypothetical fusion protein would have to cope with 60 new amino acids added to its center (see Fig. 3 for the ORF situation). The more likely explanation is that CaMV does not depend on *gag/pol* fusion. In contrast to many retroviruses, CaMV ORF V does have an ATG codon in the overlap region; it is located 14 codons upstream from the ORF IV termination codon. We therefore propose a different strategy for CaMV polymerase translation, which is related to the reinitiation (relay race) model of Sieg and Gronenborn (4; Sieg and Gronenborn, abstract) for CaMV ORFs VII, I, II, and III and to the backwards-scanning model of Thomas and Capecchi (27). After translation of ORF IV and release of the product, some of the ribosomes, or at least some of their 40S subunits, remain attached to the RNA and scan backwards for the ATG codon of ORF V to reinitiate translation and produce an unfused ORF V product. Although in this strategy a separate mRNA remains spared and ORF V translation is down regulated with respect to ORF IV translation, incorporation of the resulting polymerase into the capsid, if necessary, would have to be guided by noncovalent molecular interactions. The concentration of all viral proteins within the inclusion bodies might facilitate these interactions (7). Whether this reinitiation is assisted by additional features of the RNA sequence could be studied by in vitro mutagenesis. Alternatively, one could propose either a separate subgenomic mRNA for ORF V (24) or that, in contrast to the general scanning model of Kozak (reviewed in reference 18), ribosomes can enter eucaryotic RNA at internal positions.

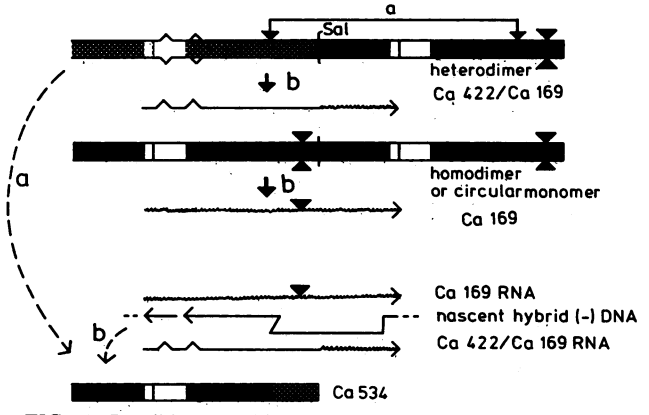

FIG. 4. Possible recombination events leading to Ca534. A heterodimer of Ca422 (left) and Ca169 (right) (as in Fig. 1, but drawn with less detail) is formed during inoculation. Recombinant 534 may have been formed from this directly by breaking and joining at nonhomologous sites (pathway a). Alternatively, the heterodimer could have been transcribed, yielding a hybrid RNA. Recombination between this and Ca169 RNA by template switching of the reverse transcription machinery may have yielded Ca534 (pathway b).

We gratefully acknowledge the generous gifts of mutant CaMV clones by L. Dixon and B. Gronenborn and oligonucleotides by J. Jiricny and W. Zürcher. Privileged communication with J. Davies, J. Stanley, and M. Schultze and discussions with and critical readings of the manuscript by J. Fütterer, M. Schultze, K. Wiebauer, and K. Ballmer are greatly appreciated.

LITERATURE CITED

1. **Brisson, N., J. Paszkowski, J. R. Penswick, B. Gronenborn, I. Potrykus, and T. Hohn.** 1984. Expression of a bacterial gene in plants by using a viral vector. Nature (London) **310**:511–514.
2. **Choe, S., U. Melcher, K. Richards, G. Lebeurier, and R. C. Essenberg.** 1985. Recombination between mutant CaMV DNAs. Plant Mol. Biol. **5**:281–289.
3. **Delseny, M., and R. Hull.** 1983. Isolation and characterization of faithful and altered clones of the genomes of cauliflower mosaic virus isolates Cabb B-JI, CM4-184 and Bari I. Plasmid **9**:31–41.
4. **Dixon, L., and T. Hohn.** 1984. Initiation of translation of the cauliflower mosaic virus genome from a polycistronic mRNA: evidence from deletion mutagenesis. EMBO J. **3**:2731–2736.
5. **Dixon, L. K., and T. Hohn.** 1985. Cloning and manipulating cauliflower mosaic virus, p. 247–275. *In* Y. Becker (ed.), Developments in molecular virology, vol. 5. Martinus Nijhoff Publishing, Boston.
6. **Dixon, L. K., T. Nyffenegger, G. Delley, J. Martinez-Izquierdo, and T. Hohn.** 1986. Evidence for replicative recombination in cauliflower mosaic virus. Virology **150**:463–468.
7. **Fuetterer, J., and T. Hohn.** 1987. Involvement of nucleocapsids in reverse transcription—a general phenomenon? Trends Biochem. Sci. **12**:92–95.
8. **Gardner, R. C.** 1983. Plant viral vectors: CaMV as an experimental tool. Basic Life Sci. **26**:121–142.
9. **Geldreich, A., G. Lebeurier, and L. Hirth.** 1986. In vivo dimerization of CaMV DNA can explain recombination. Gene **48**:277–286.
10. **Grimsley, N., T. Hohn, and B. Hohn.** 1986. Recombination in a plant virus: template-switching in CaMV. EMBO J. **5**:641–646.
11. **Groner, B., and N. E. Hynes.** 1982. Long terminal repeats provide regulatory signals at the ends of retroviral genes. Trends Biochem. Sci. **1**:400–403.
12. **Hohn, T., B. Hohn, and P. Pfeiffer.** 1985. Reverse transcription in a plant virus. Trends Biochem. Sci. **10**:205–209.
13. **Howarth, A. J., R. C. Gardner, J. Messing, and R. J. Shepherd.** 1981. Nucleotide sequence of naturally occurring deletion mutants of cauliflower mosaic virus. Virology **112**:678–685.
14. **Howell, S. H.** 1985. The molecular biology of plant DNA

viruses. Crit. Rev. Plant Sci. **2**:287–316.
15. **Howell, S. H., L. L. Walker, and R. M. Walden.** 1981. Rescue of in vitro generated mutants of cloned cauliflower mosaic virus. Nature (London) **293**:483–486.
16. **Jacks, T., and H. E. Varmus.** 1985. Expression of the Rous sarcoma *pol* gene by ribosomal frameshifting. Science **230**:1237–1242.
17. **Jolly, D. J., R. C. Willis, and T. Friedmann.** 1986. Variable stability of a selectable provirus after retroviral vector gene transfer into human cells. Mol. Cell. Biol. **6**:1141–1147.
18. **Kozak, M.** 1986. Regulation of protein synthesis in virus infected animal cells. Adv. Virus Res. **31**:229–292.
19. **Lebeurier, G., L. Hirth, B. Hohn, and T. Hohn.** 1982. In vivo recombination of cauliflower mosaic virus DNA. Proc. Natl. Acad. Sci. USA **79**:2932–2936.
20. **Mason, W. S., J. M. Taylor, and R. Hull.** 1987. Retroid virus genome replication. Adv. Virus Res. **32**:35–96.
21. **Norrander, J., T. Kempe, and J. Messing.** 1983. Improved M13 vectors using oligonucleotide-directed mutagenesis. Gene **26**: 101–106.
22. **Olszewski, N., and T. Guilfoyle.** 1983. Nuclei purified from cauliflower mosaic virus-infected turnip leaves contain subgenomic, covalently closed circular cauliflower mosaic virus DNAs. Nucleic Acids Res. **11**:8901–8914.
23. **Pietrzak, M., and T. Hohn.** 1985. Replication of the cauliflower mosaic virus: studies on the putative plus strand primer binding site. Gene **33**:169–179.
24. **Plant, A. L., S. N. Covey, and D. Grierson.** 1985. Detection of a subgenomic mRNA for gene V, the putative reverse transcriptase gene of cauliflower mosaic virus. Nucleic Acids Res. **13**: 8305–8321.
25. **Rongxiang, F., W. Xiaojun, B. Ming, T. Yingchuan, C. Faxing, and M. Keqiang.** 1985. Complete nucleotide sequence of CaMV (Xinjing isolate) genomic DNA. Chin. J. Virol. **1**:247–256.
26. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
27. **Thomas, K. R., and M. R. Capecchi.** 1986. Introduction of homologous DNA sequences into mammalian cells induces mutations in the cognate gene. Nature (London) **324**:34–38.
28. **Walden, R. M., and S. H. Howell.** 1982. Intergenomic recombination events among pairs of defective cauliflower mosaic virus genomes in plants. J. Mol. Appl. Genet. **1**:447–456.
29. **Weiss, R., N. Teich, H. Varmus, and J. Coffin.** 1984. RNA tumor viruses. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
30. **Yoshinaka, Y., I. Katoh, T. D. Copeland, and S. Oroszlan.** 1985. MuLV protease is encoded by the *gag/pol* gene and is synthesized through suppression of an amber termination codon. Proc. Natl. Acad. Sci. USA **82**:1618–1622.