Nucleotide Sequence and Evolutionary Relationships of Cucumber Mosaic Virus (CMV) Strains: CMV RNA 2

By THOMAS M. RIZZO AND PETER PALUKAITIS*

Department of Plant Pathology, Cornell University, Ithaca, New York 14853, U.S.A.

(Accepted 25 April 1988)

SUMMARY

The nucleotide sequence of RNA 2 of the Fny strain (Subgroup I) of cucumber mosaic virus (CMV) was determined and compared at both the nucleic acid and protein level with the previously determined corresponding sequence of RNA 2 of the Q strain (Subgroup 2) of CMV. Fny-CMV RNA II 2 consisted of 3050 nucleotides and contained a single open reading frame (ORF) of 2571 nucleotides, whereas Q-CMV RNA 2 consists of 3035 nucleotides and contains a single ORF of 2517 nucleotides. At the nucleotide level, there was 71% sequence homology between the two RNAs, while at the protein level sequence homology was 73%. Protein homology was greater (89%) in the central third than in either the N-terminal (64%) or the C-terminal (56%) thirds. The secondary structures of the 3' end of the RNAs were very similar, even though the nucleotide sequence homology between the 3'-terminal 180 nucleotides was only 62%. By contrast, there was 80% sequence homology between the 5'-terminal 86 residue, non-translated regions of the two RNAs. The evolutionary relationships and the divergence and retention of specific sequences among the two CMV strains and other plant viruses are discussed.

INTRODUCTION

Cucumber mosaic virus (CMV) is a positive-sense RNA plant virus with a broad host range of over 775 species in 365 genera and 85 families, including both monocotyledonous and dicotyledonous plants (Douine et al., 1979; Kaper & Waterworth, 1981). There are many isolates or strains of this virus and it has a world-wide distribution (Kaper & Waterworth, 1981); strains differ in either pathology or host range.

The genome of CMV consists of three single-stranded RNA species, designated RNAs 1, 2 and 3 in decreasing order of Mr (Peden & Symons, 1973); virions also contain a subgenomic RNA (RNA 4) derived from RNA 3, which is the mRNA used for coat protein synthesis (Schwinghamer & Symons, 1975). On the basis of serology (Devergne & Cardin, 1975) and nucleic acid hybridization (Gonda & Symons, 1978; Piazzolla et al., 1979), the strains of CMV appear to fall into two subgroups. Of 39 strains examined in this and other laboratories (Gonda & Symons, 1978; Piazzolla et al., 1979; F. Garcia-Arenal & P. Palukaitis, unpublished results) by nucleic acid hybridization analysis, 30 belong to one subgroup (Subgroup I) and nine belong to a second (Subgroup II). Complementary DNA–RNA hybridization between the two subgroups has indicated varying degrees (10 to 50%) of nucleotide sequence homology; however, RNAs belonging to the two subgroups can be reassorted to construct viable pseudorecombinants (Rao & Francki, 1982; Edwards et al., 1983). Thus, the two subgroups contain genetically compatible RNAs. The nucleotide sequences of the genomic RNAs of one strain from Subgroup II, Q-CMV, have been determined (Gould & Symons, 1982; Rezaian et al., 1984, 1985). In this paper, we describe the nucleotide sequence of RNA 2 of a Subgroup I strain, Fny-CMV, and compare it with that of Q-CMV RNA 2.
METHODS

Bacterial strains and plasmids. Escherichia coli strain JM101 (Messing, 1979) and pUC18 (Norrander et al., 1983) were used as recipient and vector, respectively, in the construction of cDNA clones of Fny-CMV RNA 2. Recipient E. coli strains JM109 (Yanisch-Perron et al., 1985) or DH5αF' (Liss, 1987) and vectors M13mp18 or M13mp19 (Norrander et al., 1983) were used in the construction of a Bal 31-generated, ordered set of deletions (Poncz et al., 1982) used for sequencing Fny-CMV RNA 2.

Complementary DNA cloning. Fny-CMV was propagated and isolated, and the viral RNA was extracted and purified as previously described for other CMV strains (Palukaitis & Zaitlin, 1984). Complementary DNA was prepared to total CMV RNA, by the procedure of Gubler & Hoffman (1983) using a decamer primer (5'-TGGTCTCCTT-3') complementary to the 3'-terminal 10 nucleotides of all four CMV RNAs. The sequence of the 3' end was determined on B-CMV RNA 4; B-CMV and Fny-CMV belong to the same subgroup of CMV and have extensive sequence homology (P. Palukaitis, unpublished results).

Subsequently, this cDNA was blunt end-ligated to Smal-linearized pUC18; the ligation mixture was used to transform competent E. coli strain JM101. The resultant white colonies were blotted onto nitrocellulose and probed with randomly primed 32P-labelled cDNA made to total Fny-CMV RNA, and transformants carrying CMV-specific cDNA clones were detected.

DNA purification and manipulation. Alkaline lysis (Maniatis et al., 1982) was used in the large scale isolation of plasmid DNA. DNA fragments were separated by electrophoresis and extracted from low melting temperature agarose (Maniatis et al., 1982). Escherichia coli was transformed as described by Messing (1983). A modification of the procedure of Birnboim & Doly (1979) was used to screen recombinant clones rapidly.

Nucleic acid hybridizations. Agarose gel electrophoresis of RNA after denaturation with formamide/formaldehyde, transfer of RNA to nitrocellulose and hybridization of immobilized RNA to 32P-labelled DNA were as previously described (Palukaitis, 1984). Agarose gel electrophoresis of restriction enzyme-cleaved plasmid DNA, Southern transfer of DNA to nitrocellulose and hybridization were as described by Maniatis et al. (1982).

Northern blots were probed with 32P-labelled plasmid DNA prepared by the 'oligolabelling' method of Feinberg & Vogelstein (1983). Southern blots were probed with randomly primed 32P-labelled cDNA made to total Fny-CMV RNA, and transformants carrying CMV-specific cDNA clones were detected.

Enzymes and chemicals. Restriction endonucleases were purchased from Bethesda Research Laboratories, Boehringer Mannheim, New England Biolabs and United States Biochemicals. T4 DNA ligase, Klenow fragment and polynucleotide kinase were obtained from United States Biochemicals. Bal 31 was obtained from Boehringer Mannheim, and T4 RNA ligase was purchased from Bethesda Research Laboratories. All enzymes were used as recommended by their manufacturers. Amersham and New England Nuclear supplied [α-32P]dATP (400 to 800 Ci/mmol) and [γ-32P]ATP (3000 Ci/mmol). All other chemicals were of analytical grade.

RESULTS AND DISCUSSION

Identification of an RNA 2-specific cDNA clone

Virtually all of the Fny-CMV RNA 2 sequence was obtained using a near full-length RNA 2-specific cDNA clone. After initial screening of the cDNA library for RNA 3- and RNA 4-specific clones by colony hybridization with randomly primed 32P-labelled cDNA prepared to gel-purified RNA 4, the remaining clones were screened by analysing insert size. The largest plasmid (pFny200) contained an insert of length 3.0 kb. A Northern blot of total CMV RNA was probed with 32P-labelled pFny200, which specifically hybridized to RNA 2 (results not shown). The length of the RNA 2-specific cDNA insert in pFny200 was that expected for a full-length clone.
Strategy for sequencing Fny-CMV RNA 2

Initially, pFny200 was cleaved with KpnI and one resultant cDNA fragment was subcloned into pUC18 to form pFny202, while the other was circularized to form pFny201 (Fig. 1). Both pFny202 and pFny201 could be linearized by cleavage at a unique restriction enzyme site at either vector/cDNA insert junction. These sites are within the pUC18 polylinker regions which flank both ends of each cDNA insert. Such linearization could not be done with pFny200.

The linearized plasmids were treated with nuclease Bal 31 for various lengths of time, and the resultant truncated fragments were cleaved at the remaining vector/cDNA insert junction. This mixture of vector and cDNA fragments was subcloned into either M13mp18 or M13mp19 such that the ends of the deleted regions were adjacent to the M13 primer site. Subclones containing cDNA were identified by restriction enzyme analysis of M13 replicative forms and/or hybridization of M13 replicative forms to randomly primed 32P-labelled cDNA made to total Fny-CMV RNA (results not shown).

The DNA sequence of the original full-length cDNA insert of pFny200 was determined by sequencing into insert cDNA in subclones of progressively smaller size. Both strands of the insert were sequenced, since pFny202 and pFny201 had been linearized at both vector/cDNA insert junctions prior to digestion with nuclease Bal 31.

The ordered set of deletions is illustrated in Fig. 1. Since the construction of pFny202 and pFny201 may have resulted in the loss of a small undetected KpnI fragment because of a second KpnI site being very close to the one shown in Fig. 1, a 0.27 kb NruI fragment containing the KpnI restriction site(s) was subcloned in both orientations into M13mp18 to form pFny200.1J and pFny200.2J. Subsequent sequencing of this NruI fragment revealed only one KpnI restriction site.

The 5'-terminal 33 nucleotides of Fny-CMV RNA 2 were determined by direct RNA sequencing. This revealed that the cDNA insert of pFny200 lacked the 5'-terminal 19 nucleotides of RNA 2. Also, four 3'-terminal nucleotides of the primer sequence were lost from pFny200 during the cloning procedure. The complete nucleotide sequence of Fny-CMV RNA 2 is presented in Fig. 2.

Fny-CMV RNA 2 encodes one long open reading frame

The only long open reading frame (ORF) of Fny-CMV RNA 2 begins at the first AUG codon at nucleotides 87 to 89 and contains 2571 nucleotides, encoding a 96720 Mr protein (857 amino acids; Fig. 2). By contrast, the translation product of Q-CMV RNA 2 has a predicted Mr of 94333 (Rezaian et al., 1984). In Fny-CMV RNA 2, the next largest ORF on the positive strand starts at residue 2419 and is 330 nucleotides long; Q-CMV RNA 2 has a corresponding 300 nucleotide ORF beginning at residue 2410 (Rezaian et al., 1984). The longest ORF on the negative strand of Fny-CMV RNA 2 is only 174 nucleotides long; no counterpart ORF is present in Q-CMV RNA 2.

Comparison of Fny-CMV RNA 2 and Q-CMV RNA 2 nucleotide sequences and translation products

The alignment of the RNA 2 sequences of Fny-CMV and Q-CMV is shown in Fig. 3. Translation products of the two RNAs are aligned in Fig. 4. The nucleotide sequences and corresponding translation products have overall homologies of 71% and 73%, respectively. Unmatched nucleotides within the coding regions are localized in six specific RNA segments (Fig. 3). These segments are frameshifted with respect to one another, resulting in corresponding protein segments with little or no amino acid sequence homology (Fig. 4).

Examination of the alignment of the translation products reveals that their central regions are more similar than are either their amino termini or carboxy termini. This is due in part to the absence of frameshifted segments in the central regions. Translation product homologies and corresponding coding region homologies between the amino-terminal, central and carboxy-terminal regions are listed in Table 1.

In the amino termini and carboxy termini, the degree of homology in the coding regions of the nucleotide sequences is higher than that of the corresponding translation products, while the
converse is true for the central region homologies. This is due largely to the following reasons. (i) The two terminal regions contain six frameshifted segments which encode amino acid sequences with relatively little or no (0 to 38%) homology. However, each frameshifted segment exhibits a higher degree of homology at the nucleotide sequence level [Fig. 5a shows that part of frameshifted segment D which contained the highest level (86%) of nucleotide sequence homology and no protein sequence homology; other segments contained lower levels (22 to 79%) of nucleotide sequence homology]. (ii) In the central regions, the degenerate genetic code results in amino acid sequences having a higher degree of homology than their respective coding sequences (an example of this is shown in Fig. 5b), such that 201 out of the 360 aligned codons have one or two base changes; however, 163 of these 201 heterologous codons still encode the same amino acid.

Non-coding 3' terminus of Fny-CMV RNA 2

The non-coding 3'-terminal region of Fny-CMV RNA 2 is 393 nucleotides long (including the stop codon). This compares with 426 nucleotides in Q-CMV RNA 2. Ahlquist et al. (1981) have shown that RNAs 1, 2 and 3 of brome mosaic bromovirus (BMV) and related viruses, such as Q-CMV, have strong sequence similarities within and between viruses at their 3' ends. These authors proposed that these 3' termini can adopt two specific secondary structures on the basis of results of mapping studies using S1 nuclease, the positions of band compressions in sequencing.

Fig. 1. Ordered set of overlapping deletions used for sequencing pFny200, a cDNA clone of Fny-CMV RNA 2. Abbreviations: B, BamHI; K, KpnI; N, NruI; R, EcoRI. The ends of the cDNA insert corresponding to the 5' and 3' termini of RNA 2 are indicated.
Fig. 2. Nucleotide sequence of Fny-CMV RNA 2 and the encoded amino acid sequence of the large ORF. The initiation and stop codons are underlined.
Fig. 4. Alignment of the translation products of Q-CMV RNA 2 (top line) and Fny-CMV RNA 2 (bottom line). The translation products are divided into three regions: (a) amino-terminal, (b) central and (c) carboxy-terminal (see Table 1). Identical amino acids are indicated by vertical lines; chemically similar amino acids [as described by Dayhoff et al. (1972)] are indicated by circles between lines; gaps are denoted by dashes. Amino acid sequences encoded by frameshifted segments A to F (see Fig. 3) are boxed.

Fig. 3. Alignment of the nucleotide sequences of Q-CMV RNA 2 (upper line of each pair) and Fny-CMV RNA 2 (lower line of each pair). Identical nucleotides are indicated by vertical lines; gaps are denoted as dashes; the initiation and stop codons are underlined; frameshifted segments within coding regions are boxed and labelled A to F.
Fig. 5. The nature of variation between protein and nucleotide percentage sequence homologies. The top and bottom lines in both (a) and (b) are Q-CMV RNA 2 and Fny-CMV RNA 2 sequences, respectively. Numbers refer to either nucleotide (Fig. 3) or amino acid (Fig. 4) positions. Identical nucleotides are indicated by vertical lines; the dash denotes a nucleotide gap; codons are indicated by horizontal lines with the corresponding amino acid shown above or below these lines. (a) Part of frameshifted segment D (Fig. 3 and 4) with 86% nucleotide sequence homology and 0% amino acid sequence homology. (b) Part of the central region (Fig. 4b) with 62% nucleotide sequence homology and 100% amino acid sequence homology.

Table 1. Sequence homologies between nucleotide coding regions and encoded proteins of Q-CMV RNA 2 and Fny-CMV RNA 2*

<table>
<thead>
<tr>
<th>Region</th>
<th>Amino terminus</th>
<th>Central</th>
<th>Carboxy terminus</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nucleotide position</td>
<td>Q-CMV RNA 2</td>
<td>93–950</td>
<td>951–2030</td>
</tr>
<tr>
<td></td>
<td>Fny-CMV RNA 2</td>
<td>87–953</td>
<td>954–2033</td>
</tr>
<tr>
<td>Amino acid position</td>
<td>Q-CMV RNA 2</td>
<td>1–286</td>
<td>287–646</td>
</tr>
<tr>
<td></td>
<td>Fny-CMV RNA 2</td>
<td>1–289</td>
<td>290–649</td>
</tr>
<tr>
<td>Homology (%)</td>
<td>Nucleotide coding region</td>
<td>68</td>
<td>77</td>
</tr>
<tr>
<td></td>
<td>Protein sequence</td>
<td>64</td>
<td>89</td>
</tr>
</tbody>
</table>

* Homologies were calculated from three specific segments of the nucleotide coding regions and the encoded proteins shown in Fig. 3 and 4.

tomato aspermy cucumovirus this stem is even longer, and in broad bean mottle bromovirus RNA it is absent (Joshi et al., 1983). Thus, while variation in stem D is not unexpected, it was necessary to rule out the possibility of a cloning artefact in the 3' end structure of Fny-CMV RNA 2, since reverse transcriptase has been reported to make errors when synthesizing cDNA from an RNA template (Battula & Loeb, 1974). Thus, two additional Fny-CMV RNA 2-specific cDNA clones (pFny250 and pFny251) from the Fny-CMV cDNA clone bank were identified by restriction enzyme analysis (results not shown). Since the inserts of pFny200, pFny250 and pFny251 were of different lengths, they were not sibling cDNA clones. The nucleotide sequence of a StuI–EcoRI cDNA fragment from pFny250 and pFny251 representing 0.4 kb of the 3' terminus of RNA 2 was determined, and revealed that the 3'-terminal sequences of all three cDNA clones were identical except in the primer region, i.e. in pFny250 and pFny251 two and seven 3'-terminal nucleotides, respectively, of the primer sequence were missing. Therefore, it is highly unlikely that the 3'-terminal sequence of Fny-CMV RNA 2 reported here is artefactual.

Non-coding 5' terminus of Fny-CMV RNA 2

The non-coding 5' terminal region of Fny-CMV RNA 2 contains 86 nucleotides, 69 of which are homologous to the 92 nucleotide, 5' non-coding region of Q-CMV RNA 2 (Fig. 3). Nucleotides at positions 8, 9, 20, 44, 48, 64, 65, 70 and 82 of Fny-CMV RNA 2 differ from the aligned sequences in Q-CMV RNA 2, but are identical to the corresponding nucleotides in Q-CMV RNA 1 (compare with Q-CMV RNA 2; Rezaian et al., 1985). This suggests that certain sequences are interchangeable between the 5' non-coding regions of CMV RNAs 1 and 2.
CMV RNA 2 sequence

(a) G F 3'

(b) A CG U C U G A G

(c) H G F 3'

(d) A U G C G

Fig. 6. Secondary structures of the 3' termini of Q-CMV RNA 2 and Fny-CMV RNA 2. Comparable stem and loop structures are labelled A to H. Configurational representations of the 3' ends of Q-CMV RNA 2 (a and b) and Fny-CMV RNA 2 (c and d) are shown; (a) and (c) predominate in the absence of magnesium ions, while in the presence of magnesium ions, (b) and (d) are formed by the base pairing of the lined sequences (Rietveld et al., 1983). Nucleotides comprising stem and loop structures F, G and H are not shown in (b) and (d). Structure (a) is adapted from Ahlquist et al., 1981 with the permission of the authors and the publishers (copyright held by Cell Press), and structure (b) is taken from Joshi et al., 1983 with the permission of the authors and the publishers (copyright held by IRL Press).

The 5'-terminal nucleotide sequences conserved between Q-CMV RNAs 1 and 2 that are complementary to sequences of the satellite RNA of CMV (Rezaian et al., 1985) are also conserved in Fny-CMV RNA 2. However, whereas the satellite RNA was unable to anneal to and form a complex in vitro with Q-CMV RNA 2 (Rezaian & Symons, 1986), various satellite RNAs of CMV can form such complexes in vitro with RNA 2 of Fny-CMV (Garcia-Arenal & Palukaitis, 1987). The implications of such interactions in terms of changes in either the level of replication of CMV RNAs, the gene expression of these RNAs, or the pathology associated with the presence of satellite RNAs remain obscure.

Limited variation in the central region

Table 1 shows that the central regions of the Q-CMV and Fny-CMV RNA 2 translation products have approximately 89% sequence homology. However, of the 38 mismatched amino acids in this region, 20 reflect conservative changes (Fig. 4b), and 12 of the remaining 18 amino
acid replacements also occur at the same position of the analogous proteins encoded by alfalfa mosaic virus, BMV and/or tobacco mosaic virus. Moreover, three of the remaining six amino acid replacements represent conservative changes with regard to one or more of the other plant viruses. Therefore, amino acid changes within the highly conserved central region do not occur at random, suggesting that a highly conserved function is associated with this domain and, thus, only certain amino acid substitutions are tolerated. The conserved amino acid sequence Gly-Asp-Asp associated with viral replicase proteins (Kamer & Argos, 1984) is located at positions 609 to 611 of the central region of the Fny-CMV RNA 2 translation product. These data suggest that within a virus group a large number of amino acid sequences must remain conserved, presumably to foster interactions with other virus-encoded and host proteins involved in the replication process. It will be of interest to determine whether such conserved central regions also exist in the Q-CMV and Fny°CMV RNA 1 translation products.

The authors thank Dr J. Owen for providing the Fny-CMV cDNA clone bank described here. This work was supported by grant no. 86-CRCR-1-1983 from the USDA CGO.

REFERENCES


(Received 27 January 1988)