Complete Nucleotide Sequence of RNA 3 from Cucumber Mosaic Virus (CMV) Strain O: Comparative Study of Nucleotide Sequences and Amino Acid Sequences among CMV Strains O, Q, D and Y

By TAKAHIKO HAYAKAWA,1* MAKOTO MIZUKAMI,2 MIDORI NAKAJIMA1 and MASAHIKO SUZUKI1

1Plantech Research Institute, c/o MKC Research Center, Kamoshida 1000, Midoriku, Yokohama, Japan 227 and 2Nihon Shokuhin Kako Co. Ltd, Tazima 30, Fuji-sh, Shizuoka, Japan 417

(Accepted 24 October 1988)

SUMMARY

The complete nucleotide sequence (2217 residues) of RNA 3 of cucumber mosaic virus strain O (CMV-O) was determined. Two open reading frames were identified, encoding a 3A protein (279 amino acid residues) in the 5' -proximal region and a coat protein (218 amino acid residues). The amino acid sequence of the coat protein C terminus was determined directly from purified protein, and confirmed the presence of the coat protein open reading frame in CMV-O RNA 3. Comparison of nucleotide sequences and amino acid sequences of CMV strains O, Q, D and Y indicated the close relationship between these strains. A tRNA-like structure could be adopted by the 3' non-coding region, and this resembled a similar structure in CMV-Q in spite of nucleotide substitutions or deletions.

Cucumber mosaic virus (CMV), a multicomponent virus of the cucumovirus group (Kaper & Waterworth, 1981), has a single-stranded plus-sense RNA genome consisting of three RNA species which are designated RNA 1, 2 and 3 in order of decreasing Mr. In addition, there is a fourth RNA, RNA 4, which is subgenomic to RNA 3 and is reported to serve as a monocistronic messenger for in vitro synthesis of the coat protein (Schwinghamer & Symons, 1977). RNA 3 encodes two in vitro translation products, a 3A protein and a coat protein. By comparison with the function of tobacco mosaic virus 30K protein, the 3A protein is deduced to be responsible for cell-to-cell movement (Mesi et al., 1987). Some strains have a small virus-dependent satellite RNA, RNA 5, which has been demonstrated to alter the symptoms of disease (Kaper & Waterworth, 1977).

CMV is one of the most widespread plant viruses; up to now more than 60 strains have been reported. These strains have been studied and characterized according to their symptoms in indicator hosts, their serology and particle morphology (Kaper & Waterworth, 1981). However, there is only one strain (CMV-Q) for which the complete sequences of the four RNA species have been determined (Gould & Symons, 1982; Rezaian et al., 1984, 1985). Recently, the sequence of CMV-Q RNA 3 has been revised by Davies & Symons (1988).

Although a comparative study of nucleotide sequence has been reported for the satellite RNA (Richards et al., 1978; Kaper et al., 1988), such studies on genomic RNAs have not yet been reported. Hidaka et al. (1985) determined a 106 nucleotide sequence at the 5' -terminal region of CMV-Y RNA 4 using a direct RNA sequencing method. Recently, Cuozzo et al. (1988) reported an almost complete sequence for CMV-D RNA 4. It will be of interest to discover whether or not differences in nucleotide sequences relate to each strain's characteristics, such as symptoms of disease or host specificity. However, the partial sequences of the CMV-Y and CMV-D subgenomic RNAs are not sufficient for comparison. Thus, determination of the complete sequences of other CMV RNA genomes in addition to strain Q has been awaited.

We present here the complete nucleotide sequence of RNA 3 of CMV-O, which was isolated...
Fig. 1. Nucleotide sequence of CMV-O RNA 3. The complete nucleotide sequence (2217 residues) of RNA 3 is shown with the open reading frames for the 3A protein and the coat protein located within the upper and lower boxes respectively. The identical sequences seen in the 5' non-coding region of RNAs 3 and 4 are underlined. Dots show the region complementary to the 18S rRNA. The 5' terminus of subgenomic RNA 4 is indicated by ▼.

in Japan, and we compare the nucleotide sequences and amino acid sequences of the RNAs 3 and 4 of CMV strains.

First and second strand cDNAs of CMV-O RNA 3 were synthesized according to the method of Gubler & Hoffman (1983). DNA complementary to the Y-terminal sequence of CMV-O RNA 3 was synthesized on the basis of the results obtained by RNA direct sequencing (Peattie, 1979) and used as the primer. RNA 3 purified by sucrose density gradient centrifugation (Takanami & Fraenkel-Conrat, 1982) was used as the template. The double-stranded cDNA was treated with T4 DNA polymerase to produce blunt termini and inserted into the SmaI site of cloning vector pUC12. Transformation was carried out according to Hanahan (1985).

The nucleotide sequence was determined by a combination of the following methods: (i) dideoxy chain termination sequencing of both strands of the cDNA, (ii) RNA dideoxy chain termination sequencing using reverse transcriptase (Geliebter et al., 1986) and synthetic primers and (iii) 5'-terminal sequencing of RNAs 3 and 4 by two-dimensional homochromatography (Silberklang et al., 1979), which confirmed the 5' terminus of RNA 3 and identified the initiation site of the subgenomic RNA 4. The complete sequence of the 2217 nucleotides of CMV RNA 3 (Mr 0.8 × 10⁶) is presented in Fig. 1.
Short communication

Two open reading frames were predicted from the nucleotide sequence of RNA 3. The coding region for the 3A protein (279 amino acid residues) is preceded by a 5' non-coding region of 122 nucleotide residues and separated from the coat protein cistron (218 amino acid residues) by an intercistronic region of 298 nucleotide residues. A non-coding region at the 3' terminus (300 nucleotide residues) flanks the coat protein cistron. The Mr of the translation products of the 3A protein and the coat protein were deduced to be 30400 and 24300, respectively.

To confirm the length of the coding region of the coat protein we directly determined the sequence of C-terminal amino acid residues. The coat protein was digested with lysinopeptidase (Masaki et al., 1981). The cleaved peptides were separated by HPLC with a Cosmosil 5C8 column (0.4 × 15 cm) (Nakarai Chemical Co.) in a 0 to 60% acetonitrile gradient containing 0.1% trifluoroacetate. Each peak was collected and analysed for its amino acid composition. Since lysinopeptidase specifically cleaves peptide bonds on the carboxyl side of the lysine residue, the fragment representing the C terminus of the protein should not contain a lysine residue. Thus, the fragment containing no lysine was treated with carboxypeptidase Y to determine the amino acid sequence at the C terminus (Hayashi et al., 1973). The C terminus was determined to be Val followed by Pro and Leu (data not shown). The sequence agreed with the amino acid sequence deduced from the nucleotide sequence.

Fig. 2 (a) shows a comparison of the amino acid sequences of the coat protein (a) and the 3A protein (b) among CMVs (strain type written alongside). The dashed lines show the position where both amino acids are identical between CMV-O and -Q or CMV-O and -D. The boxes show the sequences identical with CMV-Y determined by Hidaka et al. (1985).

Fig. 2. Comparison of amino acid sequences of the coat protein (a) and the 3A protein (b) among CMVs (strain type written alongside). The dashed lines show the position where both amino acids are identical between CMV-O and -Q or CMV-O and -D. The boxes show the sequences identical with CMV-Y determined by Hidaka et al. (1985).
Fig. 3. Comparison of secondary folding of the 3' termini between CMV-O and -Q. The tRNA-like structure could be adopted by the 3'-terminal 131 residues (2087 to 2217) of CMV-O RNA 3. The secondary structure of CMV-Q is shown as proposed by Symons (1979). Shaded letters show the nucleotides that are identical in CMV-O and -Q. Arrows indicate the nucleotides changed between CMV-O and -D. For clarity, the base pairing between a and b (shown by lines) has not been drawn.

homology was calculated to be 97% between strains O and D and 75% between strains O and Q, according to the method of Takeishi & Gotho (1982). The identity of the amino acid sequence of the 3A protein between CMV-O and -Q was 83% (Fig. 2b). The difference in homology between these CMVs may reflect the phylogenetic history of each strain. Hidaka et al. (1985) have determined the partial amino acid sequences of CNBr-cleaved fragments of the CMV-Y coat protein. These sequences were also found in the CMV-O coat protein (Fig. 2a).

There are indications that the non-coding region of the 3' terminus are well conserved (Symons, 1979). The function of the so-called tRNA-like structure of the 3' non-coding region has been investigated in detail for brome mosaic virus (BMV) (Bujarski et al., 1986). CMV RNA can also adopt a tRNA-like structure and is known to be aminoacylated by tyrosine (Kohl & Hall, 1974). As shown in Fig. 3, in spite of nucleotide substitution or deletion CMV-O RNA 3 can adopt a highly complex secondary structure which resembles that of CMV-Q (Symons, 1979). The nucleotides in CMV-O and -Q that are identical are shown in the shaded areas. Substitutions and deletions occurred mainly on the arms of a clover-leaf structure, not in single-stranded but in base-paired regions. Nucleotides changed in CMV-D compared with CMV-O are shown by arrows; these substitutions do not affect the base pairing of the tRNA-like structure. We calculated the value of ΔG to be -145.6 kJ and -184.9 kJ for the arms of the clover-leaf structures of CMV-O and CMV-Q, respectively, on the basis of the rules formulated by Tinoco et al. (1973). The similar secondary structure and the low free energy values seen in the 3' non-coding regions of the CMV strains support the idea that the region plays a significant role in the stability of the genomes and viral activities such as replication.

Concerning the 5' non-coding regions, CMV-O shows few sequences homologous to those of CMV-Q, except in the first 14 nucleotide residues; however, both strains have GU-rich sequences (22 and 21 GU dinucleotides for CMV-O and CMV-Q respectively), the meaning of which is not yet clear.
Using an in vitro translation system, Yamaguchi et al. (1982) have demonstrated that BMV RNA 4, which has a sequence complementary to that of 18S rRNA in the leader sequence, can bind to ribosomes to form an initiation complex for protein synthesis much more efficiently than CMV satellite RNA, RNA 5, which does not contain such a sequence. Hidaka et al. (1985) have indicated that the six residues (58 to 63) at the 5′ non-coding region of CMV-Y RNA 3 were complementary to those present in the 3′-terminal region of 18S rRNA. We have also found such a complementary sequence in the 5′ non-coding region of CMV-O RNAs 3 and 4 (Fig. 1, indicated by dots). Furthermore, an identical sequence (UGUGUX₃₂UGAGUCG) has been found at the 5′ non-coding regions of both RNAs 3 and 4, and the sequence complementary to the 18S rRNA has been located within this novel sequence (Fig. 1). The significance of the region in the process of initiation of protein synthesis remains unclear and awaits investigation of the interaction between 18S rRNA and the 5′ non-coding region.

The authors gratefully acknowledge Professor Y. Okada and Dr T. Meshi at Tokyo University for their kind instruction. We also thank Dr K. Hanada at the National Institute of Agricultural Research Center for his kind gift of CMV-Q, and Dr J. Kondo at Mitsubishi Kasei Co for his kind gift of lysinopeptidase and helpful suggestions. This work was supported by a grant from Agrobiological Gene Analysis Research Association (AGARA).

REFERENCES


Short communication


(Received 13 July 1988)