J. gen. Virol. (1989), **70**, 1871–1875. *Printed in Great Britain*

Key words: *nucleotide sequence/CYMV/potexviruses*

Nucleotide Sequence of the 3′-Terminal Region of Clover Yellow Mosaic Virus RNA

By M. G. ABOUHAIDAR* AND R. LAI

Department of Botany, University of Toronto, 6 Queen's Park Crescent West, Toronto, Ontario, Canada M5S 1A1

(*Accepted 6 March 1989*)

SUMMARY

The nucleotide sequence of the 3′-terminal region of clover yellow mosaic virus RNA determined from cDNA clones contains one major open reading frame (ORF) encoding a protein with an M_r of 28093 (28·1K). A smaller ORF, in frame with the major one, is also present and encodes a 23·5K protein which is similar in size to the capsid protein of the virus and has several sequence similarities with the coat proteins of four other potexviruses. The smaller ORF is probably used for the expression of the coat protein because the capsid subgenomic mRNA probably does not include the AUG of the 28·1K protein. Comparison of the amino acid sequences of the coat proteins of five potexviruses revealed a large number of identical and conservative replacements of amino acid residues.

Clover yellow mosaic potexvirus (CYMV) has a genome composed of a positive sense ssRNA molecule with an M_r of 2×10^6. The 5′ end of the RNA is capped and the 3′ end is terminated with a poly(A) tail (AbouHaidar, 1983). The viral capsid protein self-assembles *in vitro* with the RNA to form infectious particles (Bancroft *et al.*, 1979). *In vitro* translation of genomic RNA results in the production of a large polypeptide with an M_r of between 160000 and 180000, and of the coat protein (M_r 23000) (Bendena & Mackie, 1986). In this paper we report the nucleotide sequence of 1050 nucleotides at the 3′ end of the RNA.

CYMV was purified from broad bean plants (*Vicia faba*) as previously described (Bancroft *et al.*, 1979), viral RNA was extracted as described (AbouHaidar, 1988) and cDNA was synthesized and cloned in pUC18 essentially as reported by AbouHaidar (1988). Two recombinant plasmids (RD48 and CH32) were selected which contained a poly(dA) tail and 0·8 kb and 4·0 kb inserts respectively. Several subclones were also generated by digestion with *Pst*I, *Mlu*I and *Acc*I and cloning blunt-ended fragments in pUC18 (Maniatis *et al.*, 1982). Sequential deletions were made using exonucleases III and VII (Yanisch-Perron *et al.*, 1985) and clones of them were used to determine the nucleotide sequence. Synthetic DNA primers complementary to specific regions of the sequence were also used for sequencing.

Nucleotide sequences were determined by the chain termination method (Sanger *et al.*, 1977) using 7-deaza-dGTP in place of dGTP or with the modified T7 DNA polymerase (Sequenase from USBC) and using dITP in place of dGTP. Nucleic acid and amino acid sequences were analysed using IBI DNA analysis computer programs developed by Pustel, and FASTP (Lipman & Pearson, 1985).

The clones were sequenced in both orientations. The nucleotide sequence of 1050 nucleotides corresponding to the 3′ end region of CYMV RNA, excluding a poly(A) tract of about 50 residues, is presented in Fig. 1. Previous analysis of genomic RNA has indicated the presence of a poly(A) tail of 75 to 100 residues (AbouHaidar, 1983) that shows in this analysis the likely location of these clones at the 3′ terminus.

An open reading frame (ORF) started at the AUG codon at position 140 and ended with a termination codon at position 911, giving rise to a protein containing 257 amino acid residues

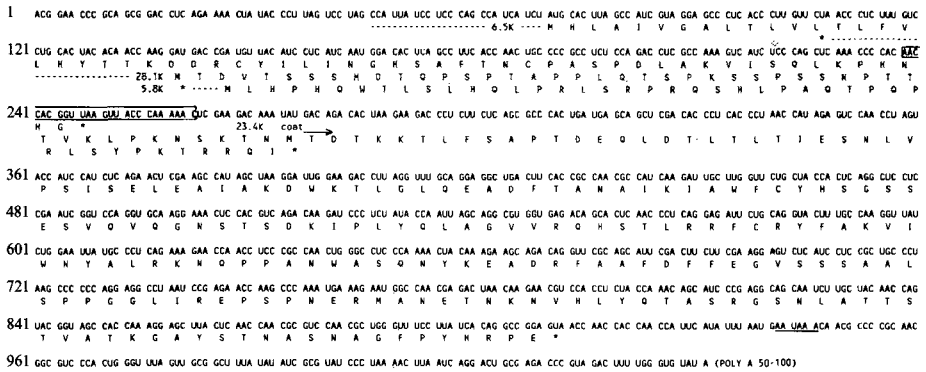

Fig. 1. The nucleotide sequence of the 3′-terminal region of CYMV RNA numbered from the 5′ end. The amino acid sequences of the proteins encoded by the major ORFs are shown below the corresponding region. Termination codons are indicated by asterisks. The putative polyadenylation signal AAUAAA is underlined and the possible cap-site for subgenomic RNA is boxed. The N termini of the encoded proteins are indicated (28·1K, 23·4K coat; 6·5K, 5·8K).

(Fig. 1) with an M_r of 28093 (28·1K). The presence of another AUG at position 275 may potentially initiate a putative protein of 212 amino acid residues with an M_r of 23436 identical to that of native CYMV coat protein (Bendena *et al.*, 1987; Brown & Wood, 1987; M. G. AbouHaidar & R. Lai, unpublished results). The amino acid sequence of this protein is similar to that of other potexvirus coat proteins (see below) which suggests that the 3′ end region of the RNA contains the coat protein-encoding sequences. The coat protein of CYMV was shown to be translated efficiently from a 1 kb subgenomic RNA (Bendena *et al.*, 1987). Furthermore CYMV-infected plants contained at least three 3′ coterminal subgenomic RNAs (approx. 1·0, 1·2 and 2·1 kb) which hybridized with coat protein-encoding sequences. The *in vitro* translation of genomic CYMV RNA resulted in the synthesis of a 31K polypeptide which was shown by Bendena *et al.* (1987) to be immunologically related to the coat protein. This suggests that the two proteins may be translated from two independent, but coterminal 1·0 and 1·2 kb subgenomic RNAs. The 28·1K protein-encoding sequence probably originated at the start codon at position 140 (Fig. 1). We have not characterized this 28·1K protein. The 1 kb subgenomic RNA which is translated into the coat protein and which is found in infected plants (Bendena *et al.*, 1987) probably originates around the AUG at position 275. Indeed, the sequence AACCACGGUUAAGUUACCCAAAAA, 13 nucleotides upstream of the start codon of CYMV putative coat protein (23·1K) is similar to a sequence near the start of regions coding for coat proteins and other proteins of other potexviruses (Zuidema *et al.*, 1989). The sequence ACGGUUAAGUU is identical to that in RNA of potato virus X (PVX) in which the first G residues were shown to be the cap-site for the subgenomic RNA found in PVX-infected plants (Dolja *et al.*, 1987). If the CYMV subgenomic RNA begins at the corresponding position, the AUG at position 275 will be the likely start codon for CYMV coat protein as the upstream AUGs would not be present on this subgenomic RNA. In such a case the coat protein subgenomic mRNA would be about 0·85 to 0·9 kb in size, close to that found in infected plants. Experiments are now under way to map the start of the coat protein subgenomic mRNA. It is also possible that the coat protein is first translated as a 28·1K protein which is then processed to the 23K protein, a strategy which would be unique to CYMV and not used by other potexviruses.

Several ORFs that could give rise to small peptides were found on both the positive and negative sense RNA strands. The largest one is in the putative negative RNA strand and could produce a protein 90 amino acids long with an M_r of 9·5K. A similar, though larger, protein (15K) was found to be encoded by the negative strand of papaya mosaic virus (PMV) RNA (AbouHaidar, 1988). Smaller proteins (6·5K and 5·8K) are encoded by sequences in front of the

Fig. 2. Coat protein amino acid sequences of CYMV and PAMV (Bundin *et al.*, 1986), PVX (Morozov *et al.*, 1983), PMV (AbouHaidar, 1988) and WClMV (Harbison *et al.*, 1988) compared using FASTP. Regions of homology are boxed, direct homologies are indicated by colons, amino acid replacements that occur frequently in evolution are indicated by dots and assumed deletions allowing better alignments are indicated by dashes.

coat protein-encoding sequence starting at positions 70 and 147 and terminating at positions 247 and 275 respectively (Fig. 1). Several smaller ORFs were also found (data not shown) but none was found to overlap that of the coat protein.

The non-coding region of the 3′ terminus of CYMV RNA [upstream of the poly(A) tail] is 138 nucleotides long (Fig. 1). Its nucleotide composition is about 27% for each of C, A and U and only 19% of G. The non-coding region contained a putative polyadenylation signal AAUAAA located at about 110 nucleotides from the poly(A) tail. The organization of the 3′ end region of CYMV RNA described here is similar to that of other potexviruses and other plant RNA viruses. The position of this putative polyadenylation signal is much farther from the beginning of the poly(A) tail than it is in eukaryotic messenger RNAs in which the AAUAAA is usually 10 to 30 nucleotides upstream from poly(A) (Nevins, 1983), or in white clover mosaic virus (WClMV) RNA [13 nucleotides upstream from poly(A) (Harbison *et al.*, 1988)]. However AAUAAA is 124 nucleotides upstream of poly(A) in PMV RNA (AbouHaidar, 1988), but is absent from PVX RNA (Morozov *et al.*, 1983); thus the role of this putative signal in these RNAs is not clear.

The CYMV coat protein has a relatively hydrophilic C-terminal half and a relatively hydrophobic N-terminal half. Two prominent hydrophilic regions (amino acids 110 to 140 and 155 to 190) were found near the C terminus and there is a relatively basic core region (amino acids 93 to 116) containing five arginine and two lysine residues. Several other clusters of acidic and basic regions are also found (e.g. amino acids 129 to 174; Fig. 2). The relative positions of certain arginine and/or lysine residues are conserved in the coat protein sequences of five potexviruses (positions 99, 129, 133, 157 and 194 in CYMV; Fig. 2). Possibly these basic amino acids are involved in binding to the negatively charged phosphates of the RNA.

A comparison of the amino acid sequences of the CYMV putative coat protein with those of four other potexviruses [PMV (AbouHaidar, 1988), PVX (Morozov *et al.*, 1983), potato aucuba mosaic virus (PAMV; Bundin *et al.*, 1986) and WClMV (Harbison *et al.*, 1988)] is given in Fig. 2. There are a number of direct matches of identical amino acids in the core regions between these proteins. There are 26 amino acids identical among the five proteins in a stretch of 67 amino acids starting at position 87. Amino acid sequences of the coat proteins of PVX and WClMV were 48% identical, those of PAMV and WClMV and PMV and CYMV were 45% and 43% identical respectively, whereas those of CYMV and PVX were only 30·5% identical (Fig. 2). When the algorithm of Lipman & Pearson (1985) (which takes into consideration not only the identical amino acids but also those amino acid replacements which occur more frequently in evolution) is used to compare the potexvirus coat protein sequences, the homology becomes very noticeable. Most similarities were in the central and C-terminal regions. All coat proteins, with the exception of that of WClMV, contain one or several proline residues at, or very close to, the C terminus (Fig. 2). The N termini do not share much homology and PAMV and PVX proteins are distinctly longer than the other three. Morozov *et al.* (1987) have previously described four blocks of homology among the sequences of the coat proteins of PVX and four potyviruses. Some of these conserved amino acids are also conserved in the other potexviruses and are indicated by triangles below the sequence in Fig. 2. One of the blocks of homology is found in a region highly conserved in the potexvirus coat proteins (between amino acid residues 133 and 140 of CYMV). These residues are probably involved in structural features which are common to the two viral groups, perhaps to keep similar structural arrangements in the respective virus particles. However, the N termini are quite variable and may lie at the large radius on the surface of the virus particle (Sawyer *et al.*, 1987) as they do in TMV particles (Stubbs *et al.*, 1977; Bloomer *et al.*, 1978). This theory might explain the weak immunological relationships between some potexviruses and some potyviruses.

We would like to thank R. A. Collins and T. L. Sit for helpful discussions. This work was supported in part by a grant to M.G.A. from the Natural Sciences and Engineering Research Council of Canada.

REFERENCES

ABOUHAIDAR, M. G. (1983). The structure of the 5′ and 3′ ends of clover yellow mosaic virus RNA. *Canadian Journal of Microbiology* **29**, 151–156.

ABOUHAIDAR, M. G. (1988). Nucleotide sequence of the capsid protein gene and 3′ non-coding region of papaya mosaic virus RNA. *Journal of General Virology* **69**, 219–226.

BANCROFT, J. B., ABOUHAIDAR, M. G. & ERICKSON, J. W. (1979). Assembly of clover yellow mosaic virus and its protein. *Virology* **98**, 121–130.

BENDENA, W. G. & MACKIE, G. A. (1986). Translational strategies in potexviruses; products encoded by clover yellow mosaic virus, foxtail mosaic virus, and viola mottle virus RNAs *in vitro*. *Virology* **153**, 220–229.

BENDENA, W. G., BANCROFT, J. G. & MACKIE, G. A. (1987). Molecular cloning of clover yellow mosaic virus RNA. Identification of coat protein sequences *in vivo* and *in vitro*. *Virology* **157**, 276–284.

BLOOMER, A. C., CHAMPNESS, J. N., BRICOGNE, G., STADEN, R. & KLUG, A. (1978). Protein disc of tobacco mosaic virus at 2·8 Å resolution showing the interactions within and between subunits. *Nature, London* **276**, 362–368.

BROWN, L. M. & WOOD, K. R. (1987). Translation of clover yellow mosaic virus RNA in pea mesophyll protoplasts and rabbit reticulocyte lysate. *Journal of General Virology* **68**, 1773–1778.

BUNDIN, V. S., VISHNYAKOVA, O. A., ZAKHARYEV, V. M., MOROZOV, S. Y., ATABEKOV, J. G. & SKRYABIN, K. G. (1986). Comparative studies of potexvirus genomes: homology between the primary structure of coat protein genes. *Doklady Akademii Nauk SSSR* **290**, 728–733.

DOLJA, V. V., GRAMA, D. P., MOROZOV, S. Y. & ATABEKOV, J. G. (1987). Potato virus X-related single- and double-stranded RNAs: characterization and identification of terminal structures. *FEBS Letters* **214**, 208–312.

HARBISON, S. A., FORSTER, R. L. S., GUILFORD, P. J. & GARDNER, R. C. (1988). Organization and interviral homologies of the coat protein gene of white clover mosaic virus. *Virology* **162**, 459–465.

LIPMAN, D. J. & PEARSON, W. R. (1985). Rapid and sensitive protein similarity searches. *Science* **227**, 1435–1441.

MANIATIS, T., FRITSCH, E. F. & SAMBROOK, J. (1982). *Molecular Cloning: A Laboratory Manual*. New York: Cold Spring Harbor Laboratory.

MOROZOV, S. YU., ZAKHARIEV, V. M., CHERNOV, B. K., PRASOLOV, V. S., KOZLOV, YU. V., ATABEKOV, J. G. & SKRYABIN, K. G. (1983). Analysis of primary structure and localisation of the coat protein gene in genomic RNA of potato virus X. *Doklady Akademii Nauk SSSR* **271**, 211–215.

MOROZOV, S. YU., LUKASHEVA, B. K., CHERNOV, B. K., SKRYABIN, K. G. & ATABEKOV, J. G. (1987). Nucleotide sequence of the open reading frames adjacent to the coat protein cistron in potato virus X genome. *FEBS Letters* **213**, 438–442.

NEVINS, J. R. (1983). The pathway of eucaryotic mRNA formation. *Annual Review of Biochemistry* **52**, 441–466.

SANGER, F., NICKLEN, S. & COULSON, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences, U.S.A.* **74**, 5463–5467.

SAWYER, L., TOLLIN, P. & WILSON, H. R. (1987). A comparison between the predicted secondary structures of potato virus X and papaya mosaic virus coat proteins. *Journal of General Virology* **68**, 1229–1232.

STUBBS, G., WARREN, S. & HOLMES, K. (1977). Structure of RNA and RNA binding site in tobacco mosaic virus from 4 Å map calculated from X-ray fibre diagrams. *Nature, London* **267**, 216–221.

YANISCH-PERRON, C., VIEIRA, J. & MESSING, J. (1985). Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene* **33**, 103–119.

ZUIDEMA, D., LINTHORST, H. J. M., HUISMAN, M. J., ASJES, C. J. & BOL, J. F. (1989). Nucleotide sequence of narcissus mosaic virus RNA. *Journal of General Virology* **70**, 267–276.

(*Received 22 September 1988*)