Nucleotide sequence of barley yellow mosaic virus RNA 1: a close evolutionary relationship with potyviruses

Satoshi Kashiwazaki,1* Yuzo Minobe,2 Toshihiro Omura1 and Hiroyuki Hibino1

1National Agriculture Research Center and 2National Institute of Agrobiological Resources, Tsukuba, Ibaraki 305, Japan

The complete nucleotide sequence of barley yellow mosaic virus (BaYMV) RNA 1 was obtained by analysis of overlapping cDNA clones and by direct RNA sequencing. The sequence is 7632 nucleotides in length, excluding a 3' poly(A) tail. The first AUG codon at nucleotide 172 appeared to be the initiator for a single long open reading frame encoding a protein of 2410 amino acids with an M, of 270755. Amino acid sequence comparisons revealed that the BaYMV 270K protein contains three regions upstream of the C-terminal capsid protein which share significant homologies with the cytoplasmic inclusion and two nuclear inclusion proteins of potyviruses thus indicating their similarities in genetic organization. However, the apparent low levels of homology in the corresponding proteins of BaYMV and potyviruses are in contrast with the high conservation among potyviruses. Moreover, our data indicate that BaYMV RNA 1 has no counterpart to the two cistrons located in the 5'-terminal region of the potyvirus genome. Although the data suggest a close evolutionary relationship between BaYMV and potyviruses, the striking differences set BaYMV apart from potyviruses.

Introduction

Barley yellow mosaic virus (BaYMV) has flexuous rod-shaped particles with two modal lengths of 550 and 275 nm (Inouye & Saito, 1975), which contain two 3'-polyadenylated single-stranded RNA species of M, 2·6 × 10⁶ (RNA 1) and 1·5 × 10⁶ (RNA 2) (Usugi et al., 1989; Kashiwazaki et al., 1989a). The nucleotide sequence has been determined for the Y-terminal region of BaYMV RNA 1, where the single capsid protein is encoded (Kashiwazaki et al., 1989a).

It has been proposed that BaYMV, wheat yellow mosaic, wheat spindle streak mosaic, oat mosaic and rice necrosis mosaic viruses should be excluded from the potyvirus group and form the bymovirus group based on their transmissibility by the fungus Polymyxa graminis, bipartite particle morphology and the absence of serological relationships between these viruses and potyviruses (Usugi et al., 1989). This proposal is supported by the finding that the BaYMV capsid protein shares only small blocks of amino acid sequence homology with those of potyviruses in contrast with the extensive matches among the capsid proteins of potyviruses (Kashiwazaki et al., 1989a). The similarities between BaYMV and potyviruses in the position in the genome of the capsid protein gene and in the way in which it is expressed by processing from a larger polyprotein precursor (Kashiwazaki et al., 1989a), as well as the formation in cells infected with either virus type of similar cytoplasmic inclusions (Hibino et al., 1981), suggest a close evolutionary relationship.

In this paper, we report the complete nucleotide sequence of BaYMV RNA 1. From comparisons of the deduced translation products with those of potyviruses and other positive strand RNA viruses, we propose a putative genetic organization for RNA 1 and assess further the taxonomic position of BaYMV.

Methods

**Virus purification and RNA isolation.** The type strain (II-1) of BaYMV (Kashiwazaki et al., 1989b) was propagated in barley by mechanical inoculation and purified as described by Usugi & Saito (1976). Viral RNA was isolated from a purified virus preparation as described previously (Kashiwazaki et al., 1989a).

**cDNA cloning.** Recombinant pBR322 plasmid pBYT10 (Kashiwazaki et al., 1989a) was used for sequence analysis of the 3'-terminal half of BaYMV RNA 1 (Fig. 1). First strand cDNA of the 5'-terminal half was synthesized using avian myeloblastosis virus (AMV) reverse transcriptase (Boehringer) and an oligonucleotide primer P10C (5'-AAATTCTTCAACTCGGG 3') complementary to nucleotides 4305 to 4322 of RNA 1) designed on the basis of sequence data from pBYT10. The second strand was synthesized as described by Gubler & Hoffman (1983). Double-stranded cDNA was cloned into pBR322 and the clones obtained were screened as described previously (Kashiwazaki et al., 1989a). The two selected cDNA clones pBYTS12 and pBYTS352 were used for sequence analysis.
DNA sequencing. cDNA inserts were subcloned into M13mp8/19 or Bluescript II KS/SK. Series of unidirectionally deleted cDNAs were formed by digestion from one end by exonuclease III (Takara Shuzo) and the ssDNAs were prepared for sequencing. Annealing of DNA templates and primers labelled with fluorescent dye (Applied Biosystems) was carried out at 60 °C for 10 min. The dideoxynucleotide chain termination reaction (Sanger et al., 1977) with Taq DNA polymerase (Promega) was done at 70 °C for 10 min according to a commercially provided protocol. The DNA sequence was analysed in an automated DNA sequencer (370A, Applied Biosystems). Some DNA sequences were determined manually by the α-32P labelling method using synthesized oligonucleotide primers. All parts of the cDNAs were sequenced in both orientations.

Direct RNA sequencing. The 5'-terminal region of RNA 1 was analysed by the RNA sequencing method described by Geliebter (1987) with some modifications. Approximately 1 μg of a mixture of RNA 1 and RNA 2, and 10 ng of 32P-labelled primer PSA (5’ AAAGAGCCTAGGAGACAC 3’ complementary to nucleotides 277 to 295 of RNA 1, including a SacI site which is underlined) were annealed by heating at 80 °C for 3 min and then incubated at 46 °C for 30 min. Reagents for the sequence reaction were added in the following order; 2 μl of the RNA/primer mixture, 3 μl of reaction buffer (24 mM-Tris–HCl pH 8.3, 16 mM-MgCl2, 8 mM-DTT, 0.4 mM-dATP, 0.4 mM-dCTP, 0.4 mM-dGTP, 0.8 mM-dTTP and 4 mM-sodium pyrophosphate) containing 6 units of avian myeloblastosis virus reverse transcriptase and 1 μg of T4 gene 32 protein (Boehringer), and 1 μl of either 1 mM-ddATP, 1 mM-ddCTP, 1 mM-ddGTP, 2 mM-ddTTP or distilled water (for a primer extension reaction). The reaction mixture was incubated at 50 °C for 50 min. In experiments to sequence the most 5'-terminal portion of the RNA, 5 units of terminal deoxynucleotidyl transferase (BRL) were added to each termination reaction and the mixture was incubated at 37 °C for an additional 10 min. The proteins contained in the reaction mixture were digested by addition of 0.1 g of Proteinase K and incubation at 65 °C for 20 min. After addition of 8 μl of stop buffer (95% formamide, 0.1% bromophenol blue, 0.1% xylene cyanol FF, 1.5 mM-EDTA) the mixture was heated at 100 °C for 3 min and was loaded onto a sequencing gel.

cDNA cloning of the 5' terminus. The cDNA of the most 5'-terminal region of RNA 1 synthesized by primer extension was subjected to the polymerase chain termination reaction (PCR; Saiki et al., 1988) directed by two primers, PSA (described above) and PSB (5’ TTTAAAAACAAAACAAAAACAAAACCCAAG 3’ corresponding to nucleotides 1 to 24 of RNA 1 and TTT added to construct a Dral site which is underlined). The amplified cDNA was digested with Dral and SacI in its terminal portions and was cloned into Bluescript II SK/KS. The five clones with cDNA inserts of the expected size were selected and used for sequence analysis.

Computer analysis. Nucleotide and amino acid sequence data were analysed using a Hitachi Software Engineering DNASIS system (Version 6.0).

Results

Sequence determination of BaYMV RNA 1

Sequence analysis of BaYMV RNA 1 was made using the cDNA inserts from an overlapping set of three recombinant pBR322 plasmids, pBYT10, pBYTS12 and pBYTS352 (Fig. 1).

The sequence of the 5'-terminal region of RNA 1 that is not contained within pBYTS352 was analysed directly on the viral RNA using a specific oligonucleotide primer and reverse transcriptase. The primer extension provided an additional 199 nucleotides beyond the 5' terminus of pBYTS352. The most 5'-terminal nucleotides of the RNA were determined by additional extension with terminal deoxynucleotidyl transferase after the dideoxynucleotide chain termination reaction. As Geliebter (1987) demonstrated, the premature termination of a cDNA transcript, presumably due to the structure of the template RNA, resulted in the appearance of crossbands (bands at one position in all four lanes of the sequencing gel) which prevented sequence determination. Although the addition in the reaction mixture of the T4 gene 32 protein, which binds to single-stranded nucleic acid (Kelly et al., 1976), reduced the incidence of crossbands to some degree, we could not determine nucleotides at several positions in the sequence.

Five recombinant Bluescript II plasmids, BYTP1 to 5, carrying cDNA of the most 5'-terminal region of RNA 1 were analysed to complete the sequence (Fig. 2). Two of them contained a single base replacement of U to C at position 139 or C to U at position 215 which may be attributed to heterogeneity of the native RNA or misincorporation during repeated polymerase reactions as exemplified by Saiki et al. (1988).

Primary structure of RNA 1

BaYMV RNA 1 consists of 7632 nucleotides excluding the 3' poly(A) tail (Fig. 2) and has a calculated Mr of 2451141, which is in agreement with the Mr of 2.6 x 10^6 estimated by electrophoresis for a poly(A)-tailed molecule (Usugi et al., 1989). Computer analysis revealed a single long open reading frame (ORF) of 7230 nucleotides in one of the reading frames of the positive strand (virion polarity). The first AUG codon at nucleotide 172 in this frame appears to be the initiator for the long ORF, since C in position -2 and G in position +1 would fit...
with the consensus sequence for translation initiation in plants (AACAAUGGCC) proposed by Lüttke et al. (1987) and also with that in animals (CCA/GCCAUGG) proposed by Kozak (1984). The predicted translation product of this ORF contains 2410 amino acids with a calculated Mr of 270755 (270K protein). Other reading frames of the positive and negative strands contain numerous stop codons and therefore are not likely to code for proteins of significant size.

Comparisons of the BaYMV 270K protein with other viral proteins

The 270K protein encoded by RNA 1 is thought to be a polyprotein that contains some functional proteins besides the capsid protein, which was located in the C-terminal region (Kashiwazaki et al., 1989a). To identify and locate BaYMV RNA 1-encoded proteins, we searched for amino acid sequence homologies between the BaYMV 270K protein and the polyproteins of potyviruses, comoviruses, nepoviruses and picornaviruses.

Computer analysis using the homology plot program revealed significant homologies in the alignment of the 270K protein with the C-terminal three-fourths of the potato virus Y (PVY) polyprotein which contains the 38K protein, the cytoplasmic inclusion protein (CI), the genome-linked protein (VPg), the two nuclear inclusion proteins (NIa and NIb) and the capsid protein (Robaglia et al., 1989) (Fig. 3). The matrix graph displays small blocks of amino acid matches between the BaYMV and PVY capsid proteins as was demonstrated previously (Kashiwazaki et al., 1989a). In contrast, the BaYMV 270K protein contains regions of extended matches with the CI or NIb proteins of PVY. A few matches were also observed between a region of the BaYMV 270K protein and the PVY NIa protein. Similar distribution patterns of amino acid matches were observed when the BaYMV 270K protein was compared with the polyproteins of three other potyviruses: tobacco etch (TEV; Allison et al., 1986), tobacco vein mottling (TVMV; Domier et al., 1986) and plum pox viruses (PPV; Maiss et al., 1989).

The BaYMV 270K protein as well as the four potyivirus polyproteins displayed few extended matches with the polyproteins encoded by cowpea mosaic virus B component RNA (Lomonossoff & Shanks, 1983), tobacco black ring virus RNA 1 (Greif et al., 1988), poliovirus (Kitamura et al., 1981) or other picornavirus RNAs (Carroll et al., 1984; Palmenberg et al., 1984).

Using the amino acid homology search program, the CI, NIa, NIb and capsid proteins of four potyviruses could easily be aligned with the similar regions of the BaYMV 270K protein. The percentage of identical amino acids calculated for the alignments of the four regions of BaYMV and potyviruses (Table 1) agrees with the extent of the similarity displayed in the dot matrix (Fig. 3). Amino acid sequences in the polyproteins of BaYMV and potyviruses are more conserved in the NIb and CI regions than in the NIa and capsid protein regions. However, the similarities in the four regions of BaYMV and potyviruses were limited unlike the high levels observed in comparisons among potyviruses (Table 1).

A search for the amino acid motifs implicated in specific functions of the non-structural proteins of potyviruses, comoviruses, nepoviruses and picornaviruses showed that the three motifs shared by these viruses are also present in the same order within the BaYMV 270K protein. The two hydrophobic stretches GXXGXGKS (where X may be any amino acid residue) and D(E/D) have been proposed as a consensus motif around phosphate-binding pockets in many NTP-binding proteins (Gorbatenya et al., 1988; Gorbatenya & Koonin, 1989). These stretches have been found in the CI proteins of potyviruses and related proteins of comoviruses, nepoviruses and picornaviruses (Gorba-
Fig. 2. Nucleotide sequence of BaYMV RNA I. The predicted amino acid sequence of the single long ORF is shown below the nucleotide sequence. The possible cleavage site involved in the capsid protein maturation is shown by an arrow. Analysis of different cDNA clones showed nucleotide replacements at positions 139 (U to C) and 215 (C to U).
The triad of H, D/E and C/S residues conserved among cysteine proteases of picornaviruses, comoviruses, nepoviruses and potyviruses are suggested to form a charge transfer system analogous to that of chymotrypsin-like serine proteases (Gorbalenya et al., 1989a; Dougherty et al., 1989b). The BaYMV 270K protein also contains this triad at positions 1404, 1440 and 1507 in the region which is aligned with the potyvirus NIa proteinases (Fig. 4b).

The two characteristic stretches (T/S)-GXXTXXXN(T/S) and GDD are thought to form the core of RNA-dependent RNA polymerases of positive strand RNA viruses (Kamer & Argos, 1984; Domier et al., 1987). The GDD stretch (at positions 1965 to 1967) in the BaYMV 270K protein is preceded by an N residue as it is in the NIb polymerases of four potyviruses (Fig. 4c). The QPSTVVDNT sequence (at position 1925 to 1933) in the BaYMV 270K protein corresponds to the sequence in the former stretch found for the four potyviruses. The substitution of QR for (T/S)G in the BaYMV stretch is striking, particularly because G is conserved in the stretches so far found for positive strand RNA viruses. Some exceptions have been reported for the first or last two residues (Kamer & Argos, 1984; Domier et al., 1987; Gorbalenya et al., 1989b).

Non-coding region

The long ORF found in BaYMV RNA 1 is preceded by a 5' non-coding region of 171 nucleotides and is followed by a 3' non-coding region of 231 nucleotides. The base compositions are 28.3% A, 23.5% C, 22.3% G, 25.9% U for the overall sequence of RNA 1, 32.8% A, 31.6% C, 11.7% G, 24.0% U for the 5' non-coding region and 24.7% A, 24.7% C, 19.9% G, 30.7% U for the 3' non-coding region. As for several other plant viruses (Gallie et al., 1987), the 5' leader sequence of BaYMV RNA 1 contains relatively few G residues.

Alignment of the most 5'-terminal sequence of BaYMV RNA 1 with those of the four potyvirus RNAs revealed a highly conserved block of 12 nucleotides rich in adenine (Fig. 5). In the following 12 nucleotides (positions 14 to 25), however, BaYMV RNA 1 shares only five nucleotides with the four potyvirus RNAs, in contrast with the strict conservation observed among potyviruses (Maiss et al., 1989; Riechmann et al., 1989). Moreover, other blocks of homology reported for the 5' non-coding regions of potyviruses (Lain et al., 1989; Turpen, 1989) are not present in the BaYMV sequence.

The 3' non-coding region of BaYMV RNA 1 displayed no significant sequence similarities with those of the four potyvirus RNAs; there is little sequence conservation among the 3' non-coding regions of different potyviruses (Frenkel et al., 1989; Lain et al., 1989; Quemada et al., 1990).

Discussion

The present study indicates that BaYMV RNA 1 encodes a single large polyprotein which contains the capsid protein (Kashiwazaki et al., 1989a) and at least three non-structural proteins closely related to the cytoplasmic CI protein, NIa proteinase and NIb polymerase of potyviruses. These three non-structural proteins in addition to VPg form a cluster which is encoded by a similarly ordered gene module conserved in potyviruses, comoviruses, nepoviruses and picornaviruses (Goldbach, 1986; Domier et al., 1987; Greif et al., 1988), and which has been implicated in RNA replication (Goldbach & van Kammen, 1985; Takegami et al., 1983; Takeda et al., 1986). The positive strand RNA genomes of these viruses are translated into single large polyprotein precursors from which mature functional proteins are derived by proteolytic cleavage (see reviews by Wellink & van Kammen, 1988; Kräusslich &
Nucleotide sequence of BaYMV RNA 1

(a) BaYMV 2787
PVY 1234
PPV 1246
TEV 1246
TVMV 1190

(b) BaYMV 619
PVY 1287
PPV 1259
TEV 1282
TVMV 1242

(c) BaYMV 854
PVY 1952
PPV 2063
TEV 2069
TVMV 2015

Fig. 4. Alignment of three regions of the BaYMV 270K protein with corresponding regions from the PVY, PPV, TEV and TVMV CI proteins (a), NiA proteins (b) and NiB proteins (c). Amino acids identical to the BaYMV sequence are boxed. Amino acids of specific motifs are indicated by asterisks (see text).

Wimmer, 1988). The assignment in the BaYMV 270K protein of a homologue to the cysteine proteinase involved in polyprotein processing of these viruses suggests that this mode of genome expression is also employed in BaYMV RNA 1.

The determination of the sequence of the N terminus of the BaYMV capsid protein has suggested that this protein is produced by cleavage at a QA dipeptide (at position 2113) (Kashiwazaki et al., 1989a), which agrees with the cleavage at a QX dipeptide catalysed by the
potyvirus NIa proteinase (Carrington et al., 1988; Dougherty et al., 1989a; Garcia et al., 1989). Detailed analyses of the cleavage specificities of the potyvirus NIa proteinase have revealed a clear consensus for the sequence flanking the cleavage sites (Carrington et al., 1988; Carrington & Dougherty, 1988; Dougherty & Parks, 1989; Dougherty et al., 1989a). Fig. 6(a) shows the positions in the BaYMV 270K protein of the three types of dipeptide (QA, QG and QS) which are known to be cleavage sites for the NIa proteinase (Dougherty et al., 1989a). Comparisons of the sequences surrounding these dipeptides with the sequence found at the verified junction between the putative polymerase and capsid protein of BaYMV (site E) provide the four other possible cleavage sites shown in Fig. 6(b). By analogy with the potyvirus NIa proteinase, the conserved residues in front of the possible cleavage sites, D in position -6, I in position -4 and L in position -2, are assumed to serve as recognition signals for the cleavage events in the BaYMV 270K protein.

As shown in Fig. 6(c), the relative position of the five possible cleavage sites within the BaYMV 270K protein resembles that of the sites cleaved by the NIa proteinase in the C-terminal three-fourths of the PVV polyprotein (Robaglia et al., 1989). The QA at position 1586 (site D in Fig. 6(c)) is likely to be the cleavage site between the putative proteinase and polymerase of BaYMV. For potyviruses, a 6K protein generated by cleavage at the two sites located in the interval between the CI and NIa proteins is thought to be VPg (Carrington et al., 1988), although this has not been proved directly by protein sequencing. However, we could not locate an analogous protein in the corresponding region of the BaYMV 270K protein where we found only one possible site QA at position 1053 (site C). It is also not certain whether the N-terminal region of the BaYMV 270K protein corresponds to the potyvirus polyprotein.

Our data indicate that the BaYMV 270K protein, in proportion to its size, lacks a region corresponding to the N-terminal one-fourth of the potyvirus polyprotein which contains the aphid-transmission helper component and another putative protein (Dougherty & Carrington, 1988). Because northern blot analysis did not show cross-hybridization between BaYMV RNA 1 and RNA 2 (Kashiwazaki et al., 1989a), RNA 2 may contain some cistrons which are absent from RNA 1 and its size is great enough to encode the two proteins. Sequence analysis will clarify genetic relationships between BaYMV RNA 2 and the 5'-terminal part of the potyvirus genome.

The close similarities between BaYMV and potyviruses in the genetic organization and amino acid sequences of the three non-structural and capsid proteins suggest that these viruses have a common ancestry. The finding of a conserved block in the most 5'-terminal nucleotide sequences of their genomic RNAs further supports this suggestion. However, the apparent low levels of homology in the four corresponding proteins of BaYMV and potyviruses in contrast with the high conservation among potyviruses set BaYMV apart from potyviruses. Moreover, it is evident that BaYMV RNA 1 has no counterpart to the two cistrons located at the 5' side of the potyvirus genome. Thus, the evidence presented here further supports the proposal to establish the bymovirus group (Usugi et al., 1989).
The presence of the RNA replication gene module in BaYMV RNA 1 demonstrates that BaYMV is placed in the proposed supergroup of picorna-like plant viruses which includes comoviruses, nepoviruses and potyviruses (Goldbach, 1986; Goldbach & Wellink, 1988). BaYMV and potyviruses have a filamentous capsid architecture and their capsid protein genes are located immediately downstream of the replication gene module. In contrast, comoviruses, nepoviruses and picornaviruses have spherical particles. The capsid proteins of picornaviruses are located far upstream of the replication gene module in their monopartite genomes (Goldbach & Wellink, 1988). BaYMV RNA 2 is required for a better understanding of the proposed supergroup of picorna-like plant viruses. Sequence information on BaYMV RNA 2 is required for a better understanding of the gene organization of the bipartite genome of bymoviruses and its evolutionary relationships with other taxonomic groups.

References


(Received 11 June 1990; Accepted 18 September 1990)