Nucleotide sequence of barley yellow mosaic virus RNA 2

Satoshi Kashiwazaki, 1* Yuzo Minobe 2 and Hiroyuki Hibino 1

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

The sequence of the 3585 nucleotides [excluding the 3' poly(A) tail] of barley yellow mosaic virus (BaYMV) RNA 2 was obtained by analysis of cDNA clones and by direct RNA sequencing. The first initiation codon at nucleotide 155 was followed by a single long open reading frame encoding a protein of 890 amino acids with an Mr of 98 458. Amino acid sequence comparisons indicated that the BaYMV 98K protein contains a region similar to the C-terminal proteinase domain of the potyvirus helper component (HC) protein towards its N terminus, but that it has no sequences that resemble the N-terminal part of the HC protein or other proteins of potyviruses. The data reveal striking differences in genetic organization between BaYMV RNA 2 and the 5'-terminal region of the potyvirus genome, despite a close genetic relationship between BaYMV RNA 1 and the rest of the potyvirus genome.

Barley yellow mosaic virus (BaYMV), the type member of the proposed bymovirus group (Usugi et al., 1989), is transmitted by the fungus Polymyxa graminis and has flexuous rod-shaped particles with two modal lengths of 550 and 275 nm (Inouye & Saito, 1975). The BaYMV genome consists of two single-stranded RNA species of Mr 2.6 x 10^6 (RNA 1) and 1.5 x 10^6 (RNA 2) which are 3'-polyadenylated and probably contain a 5' genome-linked protein (Usugi et al., 1989; Kashiwazaki et al., 1989a, 1990). Since Northern blot analysis did not show cross-hybridization between BaYMV RNA 1 and RNA 2, they are not likely to have an overlapping cistron (Kashiwazaki et al., 1989a).

Nucleotide sequence analysis of BaYMV RNA 1 suggests that it is translated into a single 270K polyprotein from which functional proteins are derived by proteolytic cleavage (Kashiwazaki et al., 1990). Amino acid sequence alignment of the BaYMV 270K polyprotein with the C-terminal three-quarters of the large polyprotein encoded by the monopartite genome of potyviruses indicates close similarities in genetic organization. However, the BaYMV 270K protein contains no counterpart to the two proteins located in the N-terminal region of the potyvirus polyprotein. It was thus of interest to determine whether BaYMV RNA 2 has a region corresponding to the 5'-terminal region of the potyvirus genome.

Here we report the nucleotide sequence of BaYMV RNA 2 and assess further the genetic relationships between BaYMV and potyviruses.

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).

Recombinant pBR322 plasmid pBYT8 selected from the cDNA library of the BaYMV genome (Kashiwazaki et al., 1989a) was used to sequence most (3.4 kb) of BaYMV RNA 2. Another clone, pBYT296, was selected from the library by colony hybridization using the 3'-terminal portion of pBYT8 as a probe and used for analysis of the 3'-terminal sequence (0.2 kb) of BaYMV RNA 2 that is not contained within pBYT8. cDNA inserts were cleaved by suitable restriction endonucleases and subcloned into M13mpl8 or -mpl9. The DNA sequence was determined by the dideoxynucleotide chain termination method (Sanger et al., 1977). All parts of the cDNAs were sequenced in both orientations. Sequence data were analysed using a Hitachi Software Engineering DNASIS system (Version 6.0).

The 5'-terminal sequence of RNA 2 was determined directly on the viral RNA as described previously (Kashiwazaki et al., 1990) using the oligonucleotide primer P8E (5' TTTGGCCAGTCAAGACTG 3'; complementary to nucleotides 193 to 210 of RNA 2). The RNA sequencing provided the 5'-terminal 192 nucleotides, of which the first 51 nucleotides were identical in sequence to those reported for RNA 1 (Kashiwazaki et al., 1990). However, the cDNA insert of pBYT8 included an additional 16-nucleotide sequence 5' CAAAA-CAAAA-ACCGAAC 3' beyond the 5' terminus determined by primer extension on RNA 2. This additional sequence is a direct repeat of nucleotides 10 to 25 from the 5' terminus of RNA 2. In a separate primer extension trial using a different RNA preparation, the reaction was terminated at the same position, suggesting that the
additional portion of pBYT8 could be an artefact introduced during cDNA cloning. We have tentatively excluded the 16 nucleotides from the sequence of BaYMV RNA 2 shown in Fig. 1.

The sequence obtained is 3585 nucleotides long, excluding the 3' poly(A) tail. It contains a single long open reading frame (ORF) of 2670 nucleotides in one of the reading frames of the positive strand (virion polarity). The first AUG codon at nucleotide 155 in this frame appears to be the initiator for the long ORF, since the A in position −3 and the C in position +2 fit the consensus sequence for translation initiation in plants (AACAAUGGCCC) proposed by Lütke et al. (1987). The predicted translation product contains 890 amino acids with a calculated M₀ of 98458 (98K). Other reading frames of the positive and negative strands...
pairing conserved in the 5' non-coding regions of BaYMV RNA 1 and RNA 2.

The 5' non-coding region of BaYMV RNA 2 is 761 nucleotides long upstream of the poly(A) tail and is much longer than that of RNA 1 (231 nucleotides) (Kashiwazaki et al., 1990). Like BaYMV RNA 1 (Kashiwazaki et al., 1990), the 3' non-coding region of BaYMV RNA 2 contains a potential polyadenylation signal (UAUGU; Zaret & Sherman, 1982) 115 nucleotides upstream of the 3' poly(A) tail but does not have a second polyadenylation signal (AAUAAA; Nevins, 1983). The two RNAs of BaYMV share a small block of conserved amino acids around the signal (UAUGUUC).

The 5' non-coding region of BaYMV RNA 2 consists of 154 nucleotides and is similar in size to that of RNA 1 (171 nucleotides) (Kashiwazaki et al., 1990). Alignment of the 5' non-coding regions of the BaYMV RNAs displays considerable similarity (Fig. 2a); of the 154 nucleotides upstream of the first AUG codon found in RNA 2, only 20 nucleotides in scattered positions differ. However, no extensive similarities were detected between the 3' non-coding regions of BaYMV RNA 1 and RNA 2. This contrasts with the significant homologies found for the 3' non-coding regions of the different RNA components of several other positive-strand RNA viruses (Davies et al., 1979; Lomonossoff & Shanks, 1983; Rezaian et al., 1985; Cornelissen et al., 1986; Bouzoubaa et al., 1987; Dodd & Robinson, 1987; Greif et al., 1988; Xiong & Lommel, 1989). Although some stretches of potential base-pairing are generated by computer analysis in the non-coding regions of the BaYMV RNAs, only a potential pairing of the sequences in the 5' non-coding regions at positions 26 to 32 and 135 to 141 appears to be conserved in the two RNAs (Fig. 2b). The data may indicate that the primary sequence and secondary structure conserved in the 5' non-coding regions of the BaYMV RNAs play an important role in a process which involves only the 5' terminus, such as encapsidation or translation, or in which both termini participate but in different ways, such as replication.

Computer analysis shows that the N-terminal region of the BaYMV 98K protein has significant amino acid homologies with the helper component proteinase (HC-Pro) active domains (Carrington et al., 1989a) from four potyviruses, potato virus Y (PVY; Robaglia et al., 1989), plum pox virus (PPV; Maiss et al., 1989), tobacco etch virus (TEV; Allison et al., 1986) and tobacco vein mottling virus (TVMV; Domier et al., 1986) (Fig. 3). The sequence GYCyg found in the BaYMV 98K protein at amino acid positions 141 to 144 is identical to the sequence surrounding the catalytic C residue of the HC-Pro of potyviruses (Oh & Carrington, 1989). The H residue which is also essential for the HC-Pro activity (Oh & Carrington, 1989) is present in the BaYMV 98K protein at position 215. The HC-Pro catalyses autolytic cleavage between a G-G dipeptide at its C terminus (Carrington et al., 1989b). In the corresponding position of the BaYMV 98K protein, a G-S dipeptide is present (at position 255) and is preceded by a V residue as is the G-G dipeptide at the C terminus of the potyvirus HC-Pro (Oh & Carrington, 1989). No significant homologies were detected between the BaYMV 98K protein and the potyvirus polyprotein except in these proteinase domains.

BaYMV RNA 1 encodes a putative proteinase which corresponds to the potyvirus NIa proteinase (Dougherty & Carrington, 1988) and which may catalyse cleavage at Q-A dipeptides thought to be involved in processing of the 270K polyprotein (Kashiwazaki et al., 1990). It is unlikely that this putative proteinase plays a role in processing of the 98K protein encoded by RNA 2, because there is no Q-X dipeptide that is surrounded by the conserved sequence found at the possible Q-A cleavage sites in the 270K protein (Kashiwazaki et al., 1990).

Our data suggest that BaYMV RNA 2 is translated into a 98K polyprotein precursor from which a putative 28K proteinase may auto-excite by cleavage at the C terminus of the G-S dipeptide to generate another 70K protein. It is likely that this 70K protein, for which no counterpart was found in potyviruses, has a function specific for BaYMV. Potential N-glycosylation sites (NXS/T), which have been reported for several viral proteins (Elleman et al., 1983; Asamizu et al., 1985;
Short communication

The close similarities in organization and amino acid sequences between the BaYMV 270K protein and the C-terminal three-quarters of the potyvirus polyprotein suggest that these viruses have a common ancestry (Kashiwazaki et al., 1990). The presence in the BaYMV 98K protein of the putative second proteinase domain homologous to the potyvirus HC-Pro active domain further supports this suggestion. Nevertheless, the map proposed for the BaYMV bipartite genome does not represent a mere split analogue to the monopartite genome of potyviruses; it reveals striking differences in genetic organization between BaYMV RNA 2 and the 5'-terminal region of the potyvirus genome, despite a close genetic relationship between BaYMV RNA 1 and the rest of the potyvirus genome.

Fig. 3. Alignment of the N-terminal region of the BaYMV 98K protein with the HC-Pro proteolytically active domains from four potyviruses. Amino acids identical to the BaYMV sequence are boxed. Catalytic C and H residues are indicated by asterisks. The position of possible or known proteolytic cleavage sites is indicated by an arrow.

Fig. 4. Possible map proposed for the BaYMV genome by analogy with the PVY genome (Robaglia et al., 1989). A map for BaYMV RNA 1 has been reported (Kashiwazaki et al., 1990). ORFs are represented as open bars, VPgs as black boxes and the poly(A) tails as A. Specific amino acid motifs and known or possible cleavage sites are indicated.

References


BaYMV 130 VQFIAPDFPAHOTCVLSPIPLSFRPFTPEKMTET orbits FLEQL-PDLGAPFLAIA YTMELFAP

PVY 613 GDSDMETIATKOGCTVINFPAGMILNISERDEFKTRKDYCMYPKLQYPTINMODLATTACMQFYY

PPV 639 AKGEMATIAKACITCVINFPAGMLINNEDMAYEFXTYRDITVPMYQTPMQLVWDTACHFLAVYY

TEV 636 LNEEXMIIAKOGCICPINFALLNMKADEAKFHPRTDITVPMQPMQVATACLYLSL

TVMV 586 EISHLMTIAKOGTCICPINFALMNVEAIKQFMPTDITVPMQVATACLYLSL

BaYMV 195 TPVLQAPFDPAHOCVLSPIPLSFRPFTPEKMTET orbits FLEQL-PDLGAPFLAIA YTMELFAP

PVY 679 PDMEDALEPRILYDTDDTCTMQWDFPSQCTGGYRILKASYKHTLDFDELSDKEHYRTYG

PPV 705 PETRMAELPRILYDNNAKIFWNSFGLSGTQWMYLIMANIQILSFASDTILSNMITYLTG

TEV 702 PDYRALPFLPLYDTDPKTMWVDLSYSGTTRTQYEMWNMTTSQIEFVHGSEGSEXMTYLTG

TVMV 652 PLAASAEELPLLHYDAQKITIHYWSGSLNTCYHILANTHSQLXFASEHDEMPWAQYKLY

Fig. 3. Alignment of the N-terminal region of the BaYMV 98K protein with the HC-Pro proteolytically active domains from four potyviruses. Amino acids identical to the BaYMV sequence are boxed. Catalytic C and H residues are indicated by asterisks. The position of possible or known proteolytic cleavage sites is indicated by an arrow.

Robaglia et al., 1989), are present in the 70K protein at positions 360, 587 and 661; their functional significance remains to be studied. The BaYMV 98K protein lacks a region comparable to the N-terminal region of the potyvirus HC protein which is implicated in aphid transmission (Domier et al., 1987). Assuming that the HC-Pro has a two-domain structure (Oh & Carrington, 1989), only the domain which is responsible for proteolytic activity is shared by BaYMV. This seems to be in agreement with the fact that BaYMV is not transmitted by aphids (Usugi et al., 1989).

Based on the present results and the data reported previously (Kashiwazaki et al., 1989b, 1990), we propose a genetic map for the BaYMV genome (Fig. 4). Similar results have been obtained for RNA 2 of a German isolate of BaYMV during parallel work by Davidson et al. (1991).


*Short communication* 999

(Received 8 October 1990; Accepted 20 December 1990)