Prunus necrotic ringspot ilarvirus: nucleotide sequence of RNA3 and the relationship to other ilarviruses based on coat protein comparison

Deyin Guo,1 Edgar Maiss,1* Günter Adam2 and Rudolf Casper1

1Institute of Biochemistry and Plant Virology, Federal Biological Research Centre for Agriculture and Forestry, Messeweg 11/12, D-38104 Braunschweig and 2DSM Deutsche Sammlung von Mikroorganismen und Zellkulturen, Abteilung Pflanzenviren, c/o Messeweg 11/12, D-38104 Braunschweig, Germany

The RNA3 of prunus necrotic ringspot ilarivirus (PNRSV) has been cloned and its entire sequence determined. The RNA3 consists of 1943 nucleotides (nt) and possesses two large open reading frames (ORFs) separated by an intergenic region of 74 nt. The 5' proximal ORF is 855 nt in length and codes for a protein of molecular mass 31.4 kDa which has homologies with the putative movement protein of other members of the Bromoviridae. The 3' proximal ORF of 675 nt is the cistron for the coat protein (CP) and has a predicted molecular mass of 24.9 kDa. The sequence of the 3' non-coding region (NCR) of PNRSV RNA3 showed a high degree of similarity with those of tobacco streak virus (TSV), prune dwarf virus (PDV), apple mosaic virus (ApMV) and also alfalfa mosaic virus (AIVM). In addition it contained potential stem-loop structures with interspersed AUGC motifs characteristic for ilar- and alfamoviruses. This conserved primary and secondary structure in all 3' NCRs may be responsible for the interaction with homologous and heterologous CPs and subsequent activation of genome replication. The CP gene of an ApMV isolate (ApMV-G) of 657 nt has also been cloned and sequenced. Although ApMV and PNRSV have a distant serological relationship, the deduced amino acid sequences of their CPs have an identity of only 51.8%. The N termini of PNRSV and ApMV CPs have in common a zinc-finger motif and the potential to form an amphipathic helix.

Introduction

Prunus necrotic ringspot virus (PNRSV) belongs to the genus ilarvirus of the family Bromoviridae, which also includes the genera Alfamovirus, Bromovirus and Cucumovirus (Martelli, 1992). Members of this family have a positive-sense single-stranded RNA genome divided into three species, designated RNA1, 2, and 3 in order of decreasing size. RNA1 and RNA2 are monocistronic and encode non-structural proteins involved in viral RNA synthesis. RNA3 is bicistronic and codes for the putative movement protein, P3a, (Stussi-Garaud et al., 1987; Erny et al., 1992) and the coat protein (CP). P3a is translated directly from RNA3, whereas the CP is translated from a subgenomic mRNA, RNA4, which is also encapsidated. The complete nucleotide sequences of the RNA3 of tobacco streak virus (TSV) (Cornelissen et al., 1984) and prune dwarf virus (PDV) (Bachman et al., 1994) are available and the sequence data indicate that the RNA3 of ilarviruses share a genome organization similar to that of other members of the Bromoviridae.

Two properties are common for ilarviruses as well as alfalfa mosaic virus (AIVM): first is the formation of non-isometric nucleoprotein particles proportional in size to the encapsidated RNAs (Lister & Saksena, 1976), and second is the requirement of CP to initiate viral genome replication in host plants (Bol et al., 1971; Smit et al., 1981). This differentiates them significantly from other genera of the Bromoviridae. Furthermore, the CP of several ilarviruses and AIVM are interchangeable in the process of genome activation, although they have no apparent sequence or serological relationship (Van Vloten-Doting, 1975; Gonsalves & Garnsey, 1975; Gonsalves & Fulton, 1977). These phenomena indicate a common mechanism for the activation of genome replication by the CP. It has been shown for AIVM that genome activation necessarily depends on the specific interaction of the trypsin-cleavable N-terminal part of the CP (Bol et al., 1974; Zuidema et al., 1983b) with the 3'-terminal sequences of viral RNA3 (Houwing & Jaspars, 1978, 1982; Zuidema et al., 1983a; Zuidema & Jaspars, 1984). Recently it was confirmed that the N terminus of AIVM CP is necessary and sufficient for binding to viral RNA and the initiation of genome replication (Baer et al., 1994). It was suggested that the stem–loop structures flanked by the sequence AUGC near the 3' termini of genomic RNAs are responsible for

* Author for correspondence. Fax +49 531 299 3006.
the specific binding of the CP (Koper-Zwarthoff & Bol, 1980; Houwing & Jaspars, 1982; Zuidema et al., 1983a). In addition, it was shown that some mutations in AUGC boxes destroyed the ability of the RNA to bind with CP (Houser-Scott et al., 1994; Reusken et al., 1994). For TSV and AMV, Sehnke et al. (1989) proposed that a zinc-finger motif may be the functional region of CPs for genome activation, but the absence of a zinc-finger motif in the PDV CP suggests that this may not be required in all cases (Bachman et al., 1994).

In order to obtain more information about the genomic organization of ilarviruses and the molecular mechanism of genome activation by homologous and heterologous CP, we have determined the complete nucleotide sequence of the PNRSV RNA3 and of the CP gene of an ApMV isolate. In addition, the comparison of sequences from different isolates as well as related viruses would be helpful to elucidate the evolutionary relationships among ilarviruses and their relationship to other genera of the Bromoviridae.

**Methods**

*Virus isolates and antisera.* The PNRSV isolate (Cat. No.: PV-0096, DSM–Deutsche Sammlung von Mikroorganismen und Zellkulturen) originated from Prunus mahaleb. The ApMV-G isolate was obtained from A. Eppler, Institute of Plant Pathology, Giessen, Germany. Both viruses were propagated on cucumber (*Cucumis sativus*) cv. Riesenschmal.

*Virion purification and RNA isolation.* Both viruses were purified according to Ong & Mink (1989). The viral RNA was extracted from virus particles as described by Mais et al. (1988).

*cDNA cloning and sequencing strategies.* Viral RNA (5 µg) was polyadenylated with *Escherichia coli* poly(A) polymerase (Gibco/BRL) and subsequently used for first-strand cDNA synthesis with oligo(dT) and random hexanucleotide primers and avian myeloblastosis virus (AMV) reverse transcriptase (Promega). Second-strand cDNA was synthesized according to Gubler & Hoffman (1983). Double-stranded cDNA was made blunt-ended with T4 DNA polymerase and inserted in the HincII site of pT7T3–19U (Pharmacia).

Sixteen PNRSV cDNA clones, covering almost the entire RNA3 genome, were selected after Northern blot analysis against viral RNA for nucleotide sequencing. Subclones were generated by restriction digestion. The nucleotide sequences were determined from double-stranded cDNA clones by single-strand sequencing after digestion with *HindIII* and *EcoRI*–*EcoRV* and subsequent amplification by RT–PCR from viral RNA, and cloned in pT7T3–19U for sequencing after digestion with *BanHI* and HindIII.

**Results and Discussion**

*Nucleotide sequence of PNRSV RNA3 and its coding capacity.*

The complete nucleotide sequence of PNRSV RNA3 deduced from cDNA including the ultimate 5' end consisted of 1943 residues (Fig. 1). It is the smallest RNA among the RNAs 3 of the members of the Bromoviridae sequenced so far. The overall base content of the RNA is G, 26.1%; A, 26.8%; T, 26.8% and C, 20.3%. One of the five 3' terminus of the 3' terminus. It is unknown whether this is a polymerase error or represents a heterogeneity of the 3' terminus. Five 5' terminus sequences analysed displayed the same 5' terminal sequence. Three one-base differences were found in the non-translated regions by sequencing overlapping cDNA clones (denoted above the sequence in Fig. 1).

Computer-assisted analysis revealed two large ORFs located on PNRSV RNA3. The 5' proximal ORF (ORF3a) starts at nt 174 and terminates with a UGA triplet at nt 1023–1025, while the 3' proximal ORF (ORF3b) begins at nt 1100 and terminates with the UAG triplet at residues 1772–1774. No other ORFs longer than 200 nt were found on either plus or minus strands. Thus, the overall organization of PNRSV RNA3 is very similar to that of the RNA 3 of PDV (Bachman et al., 1994), TSV (Cornelissen et al., 1984), and AlMV (Barker et
The alignment analysis indicated that there is a sequence similarity in the core region of the 3a proteins of tripartite viruses. A domain of 17 amino acids starting at amino acid 101 of PNRSV was aligned with analogous proteins of other members of Bromoviridae. The deduced amino acid sequence of PNRSV 3a protein was compared as well as amino acid sequence comparison, it was found that the 3a proteins of the Bromoviridae function as movement proteins (Stussi-Garaud et al., 1983), brome mosaic virus (BMV) (Ahlquist et al., 1981) and cucumber mosaic virus (CMV) (Davies & Symons, 1988).

The 5'-proximal ORF3a, 855 nt, encoded a poly-peptide of 283 amino acids with a predicted molecular mass of 31-4 kDa. Based upon biochemical and mutation analysis as well as amino acid sequence comparison, it has been proposed that the 3a proteins of the Bromoviridae function as movement proteins (Stussi-Garaud et al., 1987; Erny et al., 1992; Mushegian & Koonin, 1993). The deduced amino acid sequence of PNRSV 3a protein was aligned with analogous proteins of other members of the Bromoviridae. The alignment analysis indicated that there is a sequence similarity in the core region of the 3a proteins of tripartite viruses. A domain of 17 amino acids starting at amino acid 101 of PNRSV was aligned with analogous proteins of other members of the Bromoviridae. The intercistronic region of PNRSV RNA3 consisted of 74 nt and separates the two ORFs. Like that of PDV (40 %), when compared with the total nucleotide composition. It is shorter than that of A1MV (240-345 nt), TSV (210) and PDV (259), but longer than that of BMV (91) and CMV (94). A sequence of 27-30 nt and 15 nucleotides is repeated 2-4 times in the 5' NCR of A1MV (Barker et al., 1983; Langereis et al., 1994), respectively. Three repeats of 7-10 nucleotides were also found in the 5' NCR of A1MV, TSV and PDV, it is markedly enriched in U nucleotides is repeated 2-4 times in the 5' NCR of A1MV, TSV and PDV, it is markedly enriched in U nucleotides is repeated 2-4 times in the 5' NCR of A1MV, TSV and PDV, it is markedly enriched in U nucleotides is repeated 2-4 times in the 5' NCR of A1MV, TSV and PDV, it is markedly enriched in U nucleotides is repeated 2-4 times in the 5' NCR of A1MV, TSV and PDV, it is markedly enriched in U nucleotides is repeated 2-4 times in the 5' NCR of A1MV, TSV and PDV, it is markedly enriched in U nucleotides is repeated 2-4 times in the 5' NCR of A1MV, TSV and PDV, it is markedly enriched in U nucleotides is repeated 2-4 times in the 5' NCR of A1MV, TSV and PDV, it is markedly enriched in U nucleotides is repeated 2-4 times in the 5' NCR of A1MV, TSV and PDV, it is markedly enriched in U nucleotides is repeated 2-4 times in the 5' NCR of A1MV, TSV and PDV, it is markedly enriched in U nucleotides is repeated 2-4 times in the 5' NCR of A1MV, TSV and PDV, it is markedly enriched in U 1983; Van der Kuyl et al., 1991). The exact start point of the subgenomic mRNA4 is not shown.

The 5'-non-coding sequence consisted of 173 nt. Similar to A1MV, TSV and PDV, it is markedly enriched in U (40 %), when compared with the total nucleotide composition. It is shorter than that of A1MV (240-345 nt), TSV (210) and PDV (259), but longer than that of BMV (91) and CMV (94). A sequence of 27-30 nt and 15 nucleotides is repeated 2-4 times in the 5' NCR of A1MV (Barker et al., 1983; Langereis et al., 1986) and PDV (Bachman et al., 1994), respectively. Three repeats of 7-10 nucleotides were also found in the 5' NCR of PNRSV RNA3 (Fig. 1), but the repeats had no significant similarity with those of PDV and A1MV. The function of these repeats is not clear.

*Fig. 1. The complete nucleotide sequence of PNRSV RNA3 and deduced amino acid sequence of the two predicted ORFs. Three one-base divergences are denoted above the sequence in lowercase letters. The deduced amino acid sequences of the two ORFs are given in one-letter code underneath. The arrowhead at nt 1061 denotes the assumed start side of the subgenomic mRNA4. In the 5' NCR one-base divergences are single and double underlined, respectively. In the 3' NCR the AUGCC boxes are shaded.*

RNA3 of prunus necrotic ringspot ilarvirus

Analysis of the non-coding regions (NCR) of PNRSV RNA3

The 5'-non-coding sequence consisted of 173 nt. Similar to A1MV, TSV and PDV, it is U-enriched (40%), when compared with the total nucleotide composition. It is shorter than that of A1MV (240-345 nt), TSV (210) and PDV (259), but longer than that of BMV (91) and CMV (94). A sequence of 27-30 nt and 15 nucleotides is repeated 2-4 times in the 5' NCR of A1MV (Barker et al., 1983; Langereis et al., 1986) and PDV (Bachman et al., 1994), respectively. Three repeats of 7-10 nucleotides were also found in the 5' NCR of PNRSV RNA3 (Fig. 1), but the repeats had no significant similarity with those of PDV and A1MV. The function of these repeats is not clear.

The intercistronic region of PNRSV RNA3 consisted of 74 nt and separates the two ORFs. Like that of PDV and A1MV, it is U-enriched (36-4 %), whereas that of TSV is G-rich. For viruses in the Bromoviridae, this region contains the recognition site for the enzyme that produces the subgenomic mRNA4 (French & Alhquist, 1988; Van der Kuyl et al., 1991). The exact start point of RNA4 is known for A1MV (Barker et al., 1983), BMV
Fig. 2. Folding of 3′ NCR from RNA3 of PNRSV. (a) Prediction of the secondary structure formed by the 3′ NCR of PNRSV. The folding was calculated on the nucleotide sequence determined in this paper starting at nucleotide position 1895 until 1939. (b) Prediction of the secondary structure formed by the 3′ NCR of PNRSV after mutation of the second AUGC to AGGC. The folding was calculated on the nucleotide numbers as shown in (a).

(Ahlquist et al., 1981), CMV (Davies & Symons, 1988), cowpea chlorotic mottle bromovirus (CCMV) (Allison et al., 1989) and peanut stunt cucumovirus (PStV) (Karasawa et al., 1991). All start sites of these RNAs have in common that the first two nucleotides are GU, these are mostly preceded by a C or in case of CCMV an A residue.

The start site of the PNRSV subgenomic RNA4 was mapped by primer extension. The major run-off product of the RNA corresponded to the base C at position 1060 and a second minor band occurred for the downstream G residue (Fig. 1) when compared with the adjacent sequencing ladder made with the same primer (data not shown). If the major product reflects the correct 5′ start site, the 5′-leader sequence of PNRSV RNA4 should be 40 nt long. However, the ultimate 5′ base of all capped eukaryotic messengers so far studied as well as of the above-mentioned subgenomic RNAs of Bromoviridae is a purine rather than a pyrimidine (Banerjee, 1980). Therefore, we cannot exclude the possibility that the minor run-off product, representing the downstream residue G (Fig. 1), is the genuine start. This discrepancy could be due to co-transcription of the 5′-terminal cap m7G by the reverse transcriptase, as it has been observed for beet yellows closterovirus subgenomic mRNAs where the same method of ‘cool primer extension’ also led to a double band (Agranovsky et al., 1994).

The 3′ NCR of PNRSV RNA3 is 169 nucleotides long. When comparing it with the 3′ NCRs of AIMV, TSV, PDV and ApMV, the terminal sequence of 18–23 nucleotides was conserved and had the potential to form a stem–loop structure flanked by AUGC boxes (Fig. 2a). The folding in Fig. 2(a) was obtained using the DNASH program on the PNRSV 3′ NCR from position 1895 to 1939. It consisted of two stem–loop structures, where the first AUGC preceded the first stem and the second AUGC formed a small non-paired bulge at its base.

Similar structures were obtained with all available RNA3 3′ NCR ilarvirus sequences and also with an ilarvirus RNA2 3 NCR (Ge & Scott, 1994). In case of AIMV and TSV it has been shown that these stem–loop structures, flanked by AUGC boxes represent the binding sites with a high affinity for CP (Koper-Zwarthoff & Bol, 1980; Zuidema & Jaspars, 1984; Reusken et al., 1994). In addition it has been pointed out that the AUGC box 3, starting from the 3′-end of AIMV, is essential both for binding to CP and infectivity (Reusken et al., 1994). When we omitted the last 17 nt from the AIMV RNA3 and used the following 53 nt sequence for folding, a structure similar to that shown in Fig. 2(a) was obtained. In this structure the essential AUGC box 3 also forms the bulge at the base of the second stem. Reusken et al. (1994) have demonstrated that a mutation in this box from AUGC to AGGC completely abolishes CP binding. When we used such mutated sequences for folding, in all cases the modification shown in Fig. 2(b) for PNRSV occurred. It seems tempting therefore, to assume that this common structural feature in the 3′ NCR of ilarvirus...
and AIMV RNAs may be the prerequisite for specific interaction with a homologous or heterologous coat protein leading to genome activation. The consensus in the 3' NCR of AIMV and ilarviruses also suggests a close evolutionary relationship among these viruses.

Structural features of PNRSV and ApMV CPs

CP genes of PNRSV and ApMV were cloned and sequenced (Figs 1 and 3, respectively). The CP gene of ApMV-G was 657 nt in length and coded for a protein of 218 amino acids with a predicted molecular mass of 24.6 kDa.

For ilarviruses and AIMV, the CP not only encapsidates the virions, but is also involved in genome replication. Examination of deduced PNRSV and ApMV coat protein sequences revealed that their N termini were rich in basic amino acids and have the potential to form an amphipathic helix with the majority of the charged residues aligned on one side as suggested by Argos (1981) for secondary protein structures interacting with RNA. In addition and similar to TSV the N termini of PNRSV and ApMV CPs have the sequences CRICNTHAGGCRSCKKCH and CKYCGHTHPGACVNCKWCH, respectively. These represent zinc-finger motifs CZ2,4CZ2,15XZ2,4X (where X can be either His or Cys and Z any amino acid) which have been found in many nucleic acid binding proteins (Berg, 1986) and have been proposed by Sehnke et al. (1989) to be responsible for nucleic acid–CP interaction for TSV. In analogy, the N termini of PNRSV and ApMV coat proteins may play a role in binding of genomic RNA during encapsidation and activation of genome replication, as proposed for TSV and AIMV CP N termini (Zuidema et al., 1983b; Sehnke et al., 1989). However, the absence of the zinc-finger motif in the CP of PDV suggests that this motif is not a common structural feature of all ilarviruses and, therefore, not a general requirement for CP activation of replication (Bachman et al., 1994).

Relationship of ilarviruses based on sequence comparison

PNRSV and ApMV belong to the same subgroup of ilarviruses and are serologically distantly related (Fulton, 1970). Owing to an unexpected homology of our PNRSV CP sequence with a recently published ApMV sequence (Sánchez-Navarro & Pallás, 1994) we also decided to
sequence the CP ORF of an ApMV isolate from Germany (ApMV-G). This sequence was compared with that of other ilarviruses. ApMV-G CP had 88.1% identity with that of ApMV-I. A four-residue deletion was found in ApMV-G CP when compared with that of ApMV-I. At the nucleic acid level, the CP genes of the two isolates shared 87.3% identity and three one-base insertions in ApMV-G CP gene led to a frameshift of 12 amino acids (Fig. 3), in comparison with that of ApMV-I. The amino acid sequence of ApMV-PV32 published recently (Sánchez-Navarro & Pallás, 1994) had only 50.9% and 49.3% identity with that of ApMV-G and ApMV-I, respectively, but 91.6% identity with that of PNRSV. In analogy to the guidelines for the taxonomy of potyviruses (Shukla & Ward, 1989) with respect to sequence divergence between different strains and viruses, it is proposed that ApMV-PV32 should be regarded as a strain of PNRSV.

PNRSV and ApMV CPs were also compared with those of other members of the Bromoviridae. No significant homology between CP sequences from members of the genera Bromovirus and Cucumovirus was found by multiple alignments, however, a similarity became evident in the C-terminal part of the ilarviruses and AlMV CP sequences. The results of the multiple alignment were used to generate a dendrogram, illustrating the tentative phylogenetic relationships in the family Bromoviridae (Fig. 4). This dendrogram emphasizes the high similarities found for members of the same subgroup of ilarviruses, whereas a larger divergence occurs between members of different subgroups. The close relationship between ilarviruses and AlMV also becomes evident, whereas the remaining genera of the family are clearly separated.

D. Guo gratefully acknowledges financial support by Hanns-Seidel Foundation and the Chinese National Education Commission.

References

RNA3 of prunus necrotic ringspot ilarvirus


(Received 27 September 1994; Accepted 10 January 1995)