The 119 kDa and 124 kDa polyproteins of arabis mosaic nepovirus (isolate S) are encoded by two distinct RNA2 species

A. M. Loudes, C. Ritzenthaler, M. Pinck, M. A. Serghini† and L. Pinck*

Institut de Biologie Moléculaire des Plantes du CNRS et Université Louis Pasteur, Laboratoire de Virologie, 12 rue du Général Zimmer, 67084 Strasbourg Cedex, France

Arabis mosaic virus (ArMV) is a nepovirus that is serologically distantly related to grapevine fanleaf virus (GFLV). Both ArMV and GFLV induce grapevine degeneration disease. Several ArMV isolates, unlike isolates of GFLV, produce upon in vitro translation of RNA2 a polyprotein (P2) that forms a double band in polyacrylamide–SDS gels. Cloning of full-length copies of RNA2 of an ArMV isolate from grapevine (ArMV-S) revealed that this isolate contained two RNA2s of different length, called RNA2-U and RNA2-L. The two species were not readily separated by electrophoresis of the virion RNA under denaturing gel electrophoresis conditions but could be distinguished by analysis of primer extension and in vitro translation products. The size difference of the two RNA2s is due mostly if not exclusively to differences in their coding regions. The 124 kDa RNA2-U-encoded polyprotein P2' and the 119 kDa RNA2-L-encoded polyprotein P2", which co-migrate, respectively, with the upper and lower polyprotein bands produced by RNA2 of ArMV-S, were more than 95% identical except in their N-terminal domains. In vitro maturation experiments and sequence comparisons indicate that the N-terminal products of P2' and P2" have a molecular mass of 31 kDa and 26 kDa. The genomic organization proposed is similar to that of GFLV RNA2.

Introduction

Arabis mosaic virus (ArMV) and grapevine fanleaf virus (GFLV) are two nepoviruses of the same subgroup with a bipartite plus-sense RNA genome of similar size (Martelli & Taylor, 1989). Each genomic RNA 5' end is linked to a small VPg protein and the 3' untranslated region ends with a poly(A) stretch. ArMV RNA2 of the lilac isolate (ArMV-L) has been partly sequenced (Bertioli et al., 1991) and the genomic RNAs of the F13 isolate of GFLV have been fully characterized (Ritzen-thaler et al., 1991; Serghini et al., 1990). The viruses are serologically distantly related (Dias & Harrison, 1963). This relationship is supported by the great similarity (69%) in amino acid sequence of their coat protein (CP) (Steinkellner et al., 1992). On the other hand, ArMV and GFLV are very different in terms of host spectrum and vector specificity. Whereas GFLV infects only grapevine in nature, ArMV has a very wide natural host range, about 93 species in 28 different families in addition to cultivated plants (Murant, 1981). ArMV is transmitted by Xiphinema diversicaudatum and GFLV by X. index and to a lesser extent by X. italiae (Martelli & Taylor, 1989).

In both viruses, CP results from the cleavage of an Arg–Gly bond in polyprotein P2 (Serghini et al., 1990; Steinkellner et al., 1990; Bertioli et al., 1991). Polyprotein P2 of the GFLV-F13 isolate, when translated in vitro from purified virion RNA (Pinck et al., 1988) or from full-length transcripts of RNA2 (Margis et al., 1993), migrates in SDS–polyacrylamide gels as a single product of molecular mass 122 kDa. This contrasts with the pattern of translation products of ArMV-L (Hellen et al., 1991) and grapevine isolates of ArMV (isolates S and 862, this work), where two products of apparent molecular mass 115 kDa and 105 kDa are synthesized. Since antibodies directed against the CP of ArMV-L immunoprecipitate both of the RNA2-encoded polyproteins it was proposed by Hellen et al. (1991) that ArMV RNA2 resembles the corresponding component of comoviruses in being translated to yield two overlapping polyproteins which differ only in their N-terminal region (Chen & Bruening, 1992). This hypothesis implies that ArMV RNA2 is translated in a manner similar to that in cowpea mosaic virus (CPMV), where the presence of two RNA-M-encoded polyproteins...
of 105 kDa and 95 kDa has been shown to be due to internal initiation of the 95 kDa species (Thomas et al., 1991; Verver et al., 1991). We were interested in determining the origin of the 115 and 105 kDa polyproteins produced by ArMV-S isolate from grapevine (Syrah). We report here the complete sequence of RNA2 from ArMV isolate S, and give evidence for the origin of the P2 polyprotein doublet found in most ArMV isolates. The encoded putative protein domains are compared with their counterparts in GFLV and a genomic organization for ArMV RNA2 is proposed.

Methods

Virus, bacteria and plasmids. ArMV isolate S (ArMV-S) originated from Syrah grapevines in France and ArMV isolate 862 (ArMV-862) was initially obtained from Dr Gueorgieva (Bulgaria). Both isolates, kindly provided by Dr B. Walter (INRA, Colmar, France), have been routinely multiplied on Chenopodium quinoa at the INRA station. Isolate ArMV-SF was kindly provided by Dr J. I. Cooper (NERC Institute of Virology & Environmental Microbiology, Oxford, UK). These isolates were also routinely multiplied in C. quinoa. Virus and RNA extractions were as described for GFLV (Pinck et al., 1991). Escherichia coli strain NM522 was transformed by electroporation (Plaf & Youderian, 1990). All constructions were cloned in pBluescript KS(+) (Stratagene) and pUC18 vectors.

Reverse transcription and plasmid amplification conditions. The conditions for reverse transcription and PCR amplification were as previously described in Margis et al. (1991) and Viry et al. (1993).

cDNA2: first strand synthesis and cloning. First strand cDNA synthesis was accomplished by primer extension on total virion RNA primed with oligonucleotide P3495 (5' GAATTCAGCTCGTGATCCGTACAGTCTGTCGACT22) containing the sequence for EcoRI, SacI, KpnI, BgII and SalI sites upstream of the 22 T complementary to the poly(A) sequence. The second strand was synthesized by a one-tube second-strand synthesis following the manufacturer’s protocol (GIBCO-BRL). After SalI digestion, the products were analysed on a 1% agarose gel. The band corresponding approximately to the size expected for RNA2 cDNA was excised and the DNA eluted after freezing in liquid nitrogen (Koenen, 1989). After phenol-chloroform extraction and ethanol precipitation, the DNA was digested with SalI, inserted in SalI-Smal-digested pKS and cloned in NM522 E. coli cells to obtain plasmid pG64.

RNA2 5' end cDNA synthesis and amplification by PCR. Primer P5972 (5' TGTAGCTGGCGCCGGCAGCAAGTAAGAAA, 3' TGTACGTAGCGGCCGCGGGCAGAAAAGTAAG), containing KmII, NotI and BglII sites, and primer P6287 were used. After fill-in and phosphorylation, the PCR product was inserted in Hinfl-linearized dephosphorylated pUC18 and used to transform NM522 E. coli cells. The inserts of nine positive clones were analysed with the appropriate restriction enzymes and six clones, including clones p60 and p81, were selected and analysed by in vitro translation for their capacity to produce full-size polyprotein P2.

Analysis of in vitro translation products. The coupled transcription-translation (TXT) protocol (Promega) was used for bacteriophage T7 RNA polymerase transcription and rabbit reticulocyte lysate translation of the clones containing DNA copies of RNA2. Virion RNA or transcripts were translated in reticulocyte lysates (Promega) as described in Margis et al. (1991). The translation products were analysed on 10% polyacrylamide gels according to Laemmli (1970).

DNA sequencing. A series of tested plasmid templates were prepared from plasmids pG64, p60 and p81 digested using the exonuclease III method (Henikoff, 1984). Plasmid DNA was sequenced using the chain termination method (Sanger et al., 1977). Sequencing was carried out using Sequenase II (US Biochemical) as recommended by the manufacturer. Sequences were assembled and compared with the GAP, BESTFIT and LINEUP programs (Devereux et al., 1984).

Results and Discussion

Construction of a full-length copy of RNA2

In a first attempt to obtain cDNA clones of ArMV-S using primer P3495 complementary to the 3' terminal poly(A), a nearly full-length copy of RNA2 (p64) was obtained. To our surprise T7 RNA polymerase transcripts of pG64, when translated in reticulocyte lysates, produced only a single polyprotein which comigrated with the 115 kDa product of ArMV-S; no 105 kDa product could be detected (not shown). Upon further analysis, pG64 was revealed to be an incomplete clone of RNA2: only the 3' part of the 5' untranslated leader sequence of RNA2, 30 nt upstream of the ATG initiation codon, was present. The missing part was separately cloned after reverse transcription using primer P5972, poly(A) tailing and PCR amplification as detailed in Methods and Fig. 1. Its sequence allowed us to determine

![Fig. 1. Schematic diagram showing the different clones derived from ArMV-S RNA2. The coding region of RNA2 is represented as hollow rectangle with the VPg at the 5' end indicated by a circle. The primers used for the different constructs are represented by filled boxes. The clones obtained in the first cloning procedure are represented in the upper part. The two full-length copies of RNA2 obtained by direct PCR amplification are represented in the lower part. 'E' indicates the position of the EcoRI site in p81.](image-url)
ArMV RNA2 sequence and in vitro translation

Fig. 2. For legend see p. 903.
Fig. 2. Complete nucleotide sequences of ArMV-S RNA2-U (a) and RNA2-L (b) corresponding to clone p60 and clone p81, respectively, and deduced amino acid sequence of the large ORF. The double bar, 16 nt long, under both RNA sequences indicates the region used to prime specifically on each RNA2 form. (/) indicates the R/G cleavage between MP and CP and (//) the presumed C/A cleavage between the N-terminal domain and the MP; the dipeptides cleaved are underlined.

Fig. 3. Comparison of genome organization and proteolytic processing of the polyproteins P2' and P2" encoded by RNA2-U and RNA2-L of ArMV-S, respectively. The presumed cleavage sites and their positions within the polyprotein are indicated. The sizes of the untranslated regions are given in nucleotides; those of the different coding domains are in number of amino acids and are represented by rectangles with different shading. The percentage nucleotide or amino acid identities, determined by BESTFIT in the noncoding and in the coding regions, respectively, are indicated for each region.

the 5' end of RNA2 of ArMV-S and thus to design primer P6287, identical to the first 22 nt of RNA2. Using oligonucleotides P4180 and P6287 (Fig. 1), six full-length cDNA clones of RNA2 were obtained after RT-PCR amplification and screening of the colonies with a probe corresponding to a fragment of the P38 coding region of GFLV RNA2. Restriction analysis of the inserts from these positive clones revealed that three out of six clones contained an EcoRI site absent from clone pG64 and from the 5' end sequence. Upon in vitro translation of T7 transcripts from these six clones, the ones without the EcoRI site (e.g. p60) directed the synthesis of the 115 kDa product whereas the remaining clones (e.g. p81) encoded a polyprotein that comigrated with the 105 kDa polyprotein of ArMV-S (data not shown except for clones p60 and p81 in Fig. 5). The sequence of one of each set of clones (p60 and p81) was determined in order to locate the origin of this heterogeneity.

Sequence of the DNA copies of RNA2

The RNA sequence deduced from clone p60 (3852 nt) and p81 (3711 nt) and the corresponding encoded polyprotein is shown in Fig. 2. The BESTFIT program was used to compare nucleotide or amino acid sequences. Both RNAs have noncoding regions of identical size. The 260 nt 5' noncoding sequence displayed 85% nucleotide sequence identity and the 198 nt 3' end noncoding sequence was perfectly identical (Fig. 3). As expected from the procedure used to obtain p60 and p81 with primer P6287, designed on the basis of the sequence obtained from primer extension of P5972 on virion RNA, the 22 nt at the 5' end of the two RNA2s are identical. However, the possibility of small nucleotide differences in this region which do not disrupt base-pairing with primer P6287 cannot be ruled out. Analysis of the coding capacity of each RNA showed that each...
codes for one large polyprotein and that no other polypeptide of significant size was encoded in any reading frame. These polyproteins differ by 47 residues, polyprotein P2' derived from clone p60 contained 1131 aa (molecular mass = 124304 = 124 K) and polyprotein P2" derived from clone p81 contained 1084 aa (molecular mass = 118996 = 119 K). Comparison of their amino acid sequences revealed the presence of possible Cys/Ala and Arg/Gly proteolytic cleavage sites. These putative cleavage sites aligned readily with the sequence of the previously determined sites for GFLV as shown in Fig. 3. Comparison of the three domains thus delimited indicated 98 % amino acid identity for the coat protein (CP) domain, a domain which also displays very high homology with CP domains of other isolates of the same subgroup of nepoviruses and for which the Arg/Gly cleavage has already been reported (Bertioli et al., 1991; Steinkellner et al., 1992). The presence of a 'VQV' triplet, presumed to play a part in transmission specificity of strawberry latent ringspot virus by X. diversicaudatum (Kreiah et al., 1994), was noted near the N terminus of the CP in ArMV RNA2-L (aa 587-589). Such a triplet is not fully conserved in ArMV-S since it is changed into VPV in the same position in the RNA2-U-encoded CP. Mutagenesis and nematode transmission experiments will be needed to support the possible involvement of the 'VQV' motif in vector specificity.

The central protein is expected to result from proteolytic cleavage of the polyprotein at the Cys/Ala dipeptide linkage at the N terminus and at Arg/Gly at the C terminus of this domain (see Fig. 3). This domain is a putative movement protein domain homologous to the P38 domain of GFLV for which the movement function has been demonstrated (C. Ritzenthaler and others, unpublished). In this domain the percentage amino acid identity with GFLV is 88 % and 86 % for P38 from P2' and P2", respectively. Comparison with the partial sequence of ArMV-L (2406 nt) (Bertioli et al., 1991) showed a nucleotide similarity of 91 %. The sequenced region encompasses 232 residues of the C-terminal part of the putative movement protein in which 27 aa changes occur of which 20 are conservative substitutions. The most important changes are clustered in the region 496–502 (ArMV-S P2' numbering).

The N-terminal domains, beside the 71 % amino acid identity observed (Fig. 3), differ mainly from P38 and CP in the size of the encoded proteins, P31 for P2' and P26 for P2". The difference in size between P31 and P26 of ArMV and the equivalent P28 domain of GFLV is attributed mainly to deletions within the regions extending from nt 475–540 and 601–636 in RNA2-U. For CPMV (van Bokhoven et al., 1993) as well as for GFLV (unpublished results), the N-terminal domain is believed to be involved in RNA2 replication. Analysis of the primary nucleotide structure of this N-terminal region revealed a 16 nt sequence strictly conserved in the two RNA species but at different distances from the 5' end. This allowed the design of primer P8450, which is identical to nt 742–758 in p60 and to nt 601–617 in p81 (underlined twice in Fig. 2a, b), and was therefore expected to generate run-off products of specific length by primer extension on denatured DNA of SalI-linearized plasmid DNA of clones p60 and p81 and primer extension was with 5 U of DNA polymerase in the presence of [32P]dCTP. The final products were incubated for 3 min at 90 °C in 0·3 M-KOH, ethanol precipitated and analysed on a 4% polyacrylamide sequencing gel. (b) Separation of total virion RNA under denaturing conditions in a 1% agarose-formaldehyde gel stained with ethidium bromide. Lanes 1–3 were loaded with 0·5 μg of ArMV-S, ArMV-SF and ArMV-862 virion RNA. The detail of the gel shows that the double RNA2 band appears only with isolates S and 862.

Fig. 4. Evidences for the presence of two RNA2s in the ArMV-S isolate. (a) Primer extension on virion RNA and on DNA of plasmids p81 and p60. Primer extension products obtained after priming on virion RNA from isolate ArMV-S (lane 1) and ArMV-862 (lane 3) and DNA of plasmid p60 (lane 2) and p81 (lane 4) are shown. Primer extension was with 32P-labelled P8450 annealed to 0·5 μg of virion RNA of isolates S and 862 for 5 min at 70 °C and reverse transcribed with Superscript. In parallel 1 μl of 12·5 μM-primer P8450 was annealed with 2 μg SalI-linearized denatured plasmid DNA of clones p60 and p81 and primer extension was with 5 U of DNA polymerase in the presence of [32P]dCTP. The final products were incubated for 3 min at 90 °C in 0·3 M-KOH, ethanol precipitated and analysed on a 4% polyacrylamide sequencing gel. (b) Separation of total virion RNA under denaturing conditions in a 1% agarose-formaldehyde gel stained with ethidium bromide. Lanes 1–3 were loaded with 0·5 μg of ArMV-SF, ArMV-S and ArMV-862 virion RNA. The detail of the gel shows that the double RNA2 band appears only with isolates S and 862.
expected size (lanes 2 and 4). In contrast, two cDNAs were obtained from virion RNA of ArMV-S or ArMV-862 (Fig. 4a, lanes 1 and 3), each of which comigrated with the respective run-off products from clones p60 and p81. In view of the nearly identical staining intensity of these bands, the two forms of RNA2 are estimated to be present in equal amounts in both isolates.

The presence of two distinct RNA2s was confirmed by careful examination of virion RNA from different ArMV isolates after prolonged migration in denaturing 1% agarose-formaldehyde slab gels. Provided that 0.3–0.5 μg RNA were loaded in 5 × 1.5 mm slots, the resolution of such gel was sufficient to allow detection of two very closely spaced RNA2 bands in samples of ArMV-S and 862 isolates (Fig. 4b). This mobility difference can account for the 141 nt difference between clones p60 and p81. These two forms of RNA2 will hereafter be referred to as RNA2-U and RNA2-L, respectively, to designate the upper and lower migrating forms of RNA2 of ArMV-S. Virion RNA from ArMV-SF isolate yielded a single RNA2 band under these conditions (Fig. 4b).

Processing of polyproteins P2' and P2" by the proteinase of GFLV

The presence of almost identical sequences at the presumed cleavage sites in polyproteins P2' and P2" of ArMV and P2 of GFLV suggested that ArMV polyprotein could be processed upon addition of GFLV proteinase in vitro. In order to demonstrate this possibility, the in vitro translation products of ArMV-S virion RNA and of the full-length transcripts of p60 (tr60) and p81 (tr81) were processed with the GFLV proteinase and their maturation products compared with those obtained from GFLV RNA2 transcripts (Fig. 5). The polyproteins produced from ArMV-S are referred to as P2' and P2". Polyprotein P2', which was translated from tr60 and presumably from RNA2-U, is equivalent in size to P2 of GFLV. Similarly, P2" is produced from tr81 and presumably from RNA2-L of ArMV-S. The estimated size of the two major maturation products of polyproteins P2' and P2" processed with proteinase VP7 (Margis et al., 1993) were 94 kDa from ArMV, tr60 and tr81 (indicated P94Ar in Fig. 5), 31 kDa from ArMV and tr60 (P31 in Fig. 5, lanes 4 and 6) and 26 kDa from tr81 (P26 in Fig. 5, lane 8). Assuming that the cleavages were similar to those of P2 of GFLV, the protein P94Ar presumably corresponds to the P94 maturation intermediate of P2 including the movement protein (MP) and CP (labelled P94G in Fig. 5). Similarly, the P31 and P26 species would correspond to the N-terminal proteins of P2' and P2", homologous to P28 in P2 of GFLV. We have already shown that this P28 protein is unstable and thus not visible under the experimental conditions used (Margis et al., 1993); similar instability may also apply to its P26 counterpart in ArMV. In our experiments, no CP was detected upon maturation of ArMV polyproteins. The absence of cleavage in vitro of the Arg/Gly site between P38 and CP may be attributed to the small sequence changes in the amino acids upstream of the dipeptide in comparison to those of GFLV (Fig. 3). These changes could induce conformational modifications within the maturation site and therefore hinder the GFLV proteinase action at this site.

All these results argue strongly against the initial hypothesis of internal initiation as the origin of the P2 doublet in ArMV-S and clearly indicate that two species of RNA2 are present in the ArMV-S isolate. This raises the question of the origin of these two RNA2 species. Are they due to the mixing of two distinct isolates, which would imply that two RNA1s would also be present in all ArMV isolates producing two polyproteins upon RNA2 translations, each of them replicating its own RNA2, or is there a unique RNA1 able to support replication of both RNA2s? This question remains to be resolved. The translation pattern and the RNA analysis of the ArMV-SF isolate (Hellen et al., 1991; Liu et al.,


