The nucleotide sequence and gene organization of red clover necrotic mosaic virus RNA-2

S.A. Lommel*, M. Weston-Fina, Z. Xiong and G.P. Lomonossoff

Department of Plant Pathology, Throckmorton Hall, Kansas State University, Manhattan, KS 66506, USA and The John Innes Institute, AFRC Institute of Plant Science Research, Colney Lane, Norwich NR4 7UH, UK

Received November 18, 1987; Revised and Accepted July 27, 1988 Accession no. X08021

ABSTRACT

Red clover necrotic mosaic virus, a member of the dianthovirus group, is characterized by a genome composed of two nonhomologous single-stranded RNAs of approximately 4.0 (RNA-1) and 1.4 kb (RNA-2). The complete nucleotide sequence of the RNA-2 has been determined. RNA-2 is 1448 nucleotides in length with a 5' terminal m7G cap and no 3' terminal poly-A tail or 5' terminal VPg. An open reading frame beginning at the first initiation codon at nucleotide 80 and ending at nucleotide 1030 has been identified which can encode a polypeptide of 35 kDa. RNA-2 directs the synthesis of a 35 kDa polypeptide in vitro. SP6 and T7 transcripts from full length RNA-2 cDNA clones directed the synthesis of a polypeptide with the same electrophoretic mobility as the polypeptide directed from authentic RNA-2. Clones with various 3' terminal deletions both outside and within the 35 kDa open reading frame were transcribed and translated in vitro to define the limits of the 35 kDa open reading frame. A second, small open reading frame capable of encoding a polypeptide of 4.9 kDa was also indicated from the sequence; however, there was no evidence for a protein product of that size. RNA-2 is presumed to be monocistronic and encode a cell-to-cell movement function. A small but significant amino acid sequence homology was observed with the brome mosaic virus RNA-3a polypeptide.

INTRODUCTION

The genome of red clover necrotic mosaic virus (RCNMV), a member of the dianthovirus group (1), consists of two single stranded RNA species of approximately 1.5 X 10^6 Da (RNA-1) and 0.5 X 10^6 Da (RNA-2) (2). Early studies by Morris-Krisnich et al. (3) showed that RCNMV RNAs directed the synthesis of 34 kDa, 36 kDa and 39 kDa polypeptides. The single 39 kDa capsid protein species was programmed by RNA-1 in vitro (3). Pseudorecombination studies between RCNMV and other members of the dianthovirus group (4, 5) confirmed that RNA-1 encodes the capsid protein. Recent translation studies of the RCNMV genome (Xiong and Lommel, submitted) showed that the RNA-1 encodes 50 kDa and 90 kDa polypeptides as well as the previously described products and that RNA-2 encodes a single polypeptide of 35 kDa. The polypeptides synthesized in vitro from the RCNMV genome and their
organization in the genome will be published elsewhere (Xiong and Lommel, in preparation).

Infectivity studies with separated RCNMV RNA-1 and RNA-2 demonstrated that both RNA-1 and RNA-2 were required for systemic infection in plants (2, 5). However, data obtained by Paje-Manalo and Lommel (in press) with protoplasts from the Nicotiana tabacum BY2 suspension cell line (6) and data obtained by Osemann and Buck (7) with cowpea mesophyll protoplasts showed that RNA-1 could replicate independent of RNA-2. Productive infections by RNA-1 occurred in protoplasts as indicated by the increase of viral RNA-1 copies as well as the synthesis of capsid protein in the absence of RNA-2. The requirement of both RNA-1 and RNA-2 for whole plant infection, as well as the fact that RCNMV RNA-2 determines lesion size, and virus systemic spread in cowpea (8) suggests that RCNMV RNA-2 encodes a cell-to-cell movement function. To further study the viral cell-to-cell function, near to full-length clones of RNA-2 were synthesized and the primary structure of the RNA determined. To ensure the fidelity of the cDNA clones and the sequence derived from them, transcripts were made from the clones and assayed in a rabbit reticulocyte lysate in vitro translation system, for the protein product.

MATERIALS AND METHODS
Virus and RNA Purification

The Australian strain of RCNMV was propagated on Nicotiana clevelandii. Virus was purified from tissue maintained under standard glass house conditions, 2-weeks after mechanical inoculation as described by Lommel (9). RNA was extracted from purified virus preparations by incubation with protease K followed by phenol extraction as described by Carrington and Morris (10).

cDNA Clones

Unfractionated viral RNAs were polyadenylated as described by Carrington and Morris (10). First strand cDNA synthesis was done as described by Gubler and Hoffman (11). Synthesis was performed for 1 h at 42°C in 50 mM Tris, pH 7.5, 75 mM KCl, 10 mM DTT, 3 mM MgCl2, 1 mM each of dATP, dTTP, dCTP, and dGTP, 0.1 μg/ml of BSA, 4 μg of oligo dT12-18, 2.5 units human placental ribonuclease inhibitor, 200 units of cloned M-MLV reverse transcriptase (BRL) per μg of polyadenylated RNA, and [α-32P]dATP. The reaction was terminated by the addition of 5 μl of 0.5 M EDTA, pH 8.0, and extraction with phenol and chloroform. Second strand synthesis was carried out in 20 mM Tris, pH 7.5, 5
mM MgCl2, 10 mM (NH4)2SO4, 100 mM KCl, 0.15 mM β-NAD, 50 μg/ml BSA, 40 μM each of dATP, dTTP, dGTP, and dCTP, 8.5 unit/ml of E. coli RNase H, 250 unit/ml of DNA polymerase I, and 10 unit/ml T4 ligase. Incubations were sequentially for 1 h at 15°C and 1 h at 22°C. EDTA was added to 20 mM final concentration to stop the reaction. The final products were extracted twice with phenol and precipitated with 95% ethanol.

An oligo-dCTP tail was added to the 3' hydroxy groups of the cDNA duplex by incubation in 50 mM potassium cacodylate, pH 7.2, 2 mM CoCl2, 0.2 mM DTT, 0.5 mM dCTP, and 25 units of terminal deoxynucleotidyl transferase for 10 min at 37°C. Equal molar ratios of dC-tailed cDNA and Pat-I linearized dG-tailed pBR322 (BRL) were annealed by heating for 5 min at 65°C followed by incubation for 2 h at 57°C in 10 mM Tris, pH 7.8, 0.1 M NaCl and 1 mM EDTA, and slow-cooled overnight.

Transformation of E. coli (RRL) with annealed plasmids was performed according to Maniatis et al. (12). Transformants recovered from tetracycline plates were analyzed for inserts by the alkaline lysis mini-preparation method (13) followed by restriction of plasmid with Pat-1 and electrophoresis in 1% agarose gels. RNA-2 specific cDNA clones were identified by Southern and Northern blot hybridizations using [32P]-labeled RCNMV cDNA and nick-translated clones as probes, respectively. Three RNA-2 specific cDNA clones, pRCN004, pRCN005, pRCN1085, were selected for sequence and transcription analysis.

Sequence Analysis

cDNA clones were restricted with various restriction endonucleases and fragments subcloned into M13mpl8 and M13mpl9. Sequence was determined by the dideoxy chain terminating method (14, 15) using [α-32P]dATP as the label. Sequencing data was stored and analyzed using "Seqaid" (16) analysis program. Oligodeoxynucleotide primers were used with restriction fragments over 300 bp and to sequence in both directions at the 3' terminus. Primers were synthesized using an Applied Biosystems Model 381A DNA synthesizer.

To determine the 3'-terminal sequence of RNA-2, gel-purified RNA was 3' end labeled using [5'-32P]pCp and RNA ligase. The 3'-terminal nucleotide was identified by complete alkaline hydrolysis of the end-labeled RNA followed by electrophoresis on Whatman 3-mm paper at pH 3.5 (17). Additional sequence information was obtained by partially hydrolyzing the RNA with alkali and fractionating the digests by two-dimensional electrophoresis-homochromatography (18,19).

The 5' terminus of RCNMV RNA-2 was sequenced by primer extension. A 143

8589
bp Sau-3a-Alu-1 restriction fragment representing bases 37-180 was isolated and labeled by [γ-32P]dATP incorporation within the restriction site overhang using Klenow fragment. This fragment was used as the primer for reverse transcription in a dideoxy chain terminating sequence reaction. End-labeled cDNA runoff transcripts were separated on polyacrylamide sequencing gels (20).

5' m^G Cap and Penultimate Residue Analysis

The identification of the RCNMV RNA-2 m^G cap and the 5' terminal penultimate residue was performed by methylating the penultimate residue with [3H] using the (nucleoside-2'-)-methyltransferase activity associated with infectious vaccinia virus extracts according to the methods described by Moss (21). The vaccinia virus extract was provided by B. Moss, National Institutes of Health. After methylation, RNA-2 was digested with Nuclease P1 and products analyzed by ascending paper chromatography (21) with m^GpppA and m^GpppG standards. Labeled products were eluted from the paper and digested with snake venom pyrophosphatase and bacterial alkaline phosphatase followed by reanalysis by ascending paper chromatography with A and G standards (21). For both the cap and penultimate residue analysis, turnip yellow mosaic virus (TyMV) was coprocessed as a positive control. The TyMV genomic RNA possesses a m^GpppG cap (22) and the partially packaged subgenomic RNA a m^GpppA cap (23).

Translation of RNA-2 and Cloned Transcripts

The close to full-length RNA-2 cDNA insert was isolated from pRCN005 by restriction with Bst-1 followed by electroelution. The insert was subcloned into the Bst-1 site of pGEM3. pGEM3 containing the cDNA clone was linearized by restriction with EcoRI prior to transcription, and a positive polarity transcript was synthesized using SP6 RNA polymerase. A negative polarity transcript was prepared by linearizing the plasmid with Hind-III prior to transcription with T7 RNA polymerase. Two 3' end truncated transcripts were made to confirm the end of the largest open reading frame deduced from the sequence data. The Bst-1-Bgl-2 and Bst-1-Xba-1 fragments from the pRCN005 insert corresponding to bases 15-911 and 15-1029 respectively, were subcloned into pGEM3. A 5' end truncated clone beginning at the Xba-1 site to the 3' terminal Bst-1 site corresponding to bases 1030-1451 was cloned into pGEM3 to determine if the small observed open reading frame could be expressed in vitro. The RNA transcripts were synthesized with 20 unit/µg template of the respective RNA polymerase in 40 mM Tris, pH 8.0, 8 mM MgCl2, 2 mM spermidine
(HCl)₃, 25 mM NaCl, 5 mM DTT, 0.4 mM each of ATP, UTP, GTP, and CTP. The reaction was stopped by adding 1 µl of 0.5 M EDTA after incubation at 37°C for 1 h. The newly synthesized RNAs were precipitated together with their templates.

Purified RCNMV RNA-2 and RNA transcripts synthesized in vitro were used as messengers in the translation studies. The rabbit reticulocyte in vitro translation conditions were those described by Dougherty and Hiebert (24). Incubation was for 60 min at 37°C. Lysates were denatured, electrophoresed in 12.5% SDS-polyacrylamide gels, and fluorographed as previously described.

RESULTS AND DISCUSSION

CDNA Cloning Strategy

RCNMV RNA was presumed to lack a 3' poly(A) tail, because the RNA was not retained on an oligo(dT)-cellulose column, and no virus specific clones were generated when first strand synthesis was primed with oligo(dT). Therefore, enzymatically polyadenylated RCNMV RNA-2 was used as template to direct synthesis of double stranded cDNA which were cloned into Pst-I restricted pBR322 by G-C tailing. The three longest RNA-2 specific clones (pRCN004, pRCN005, and pRCN1085) were chosen for further analysis. RCNMV RNA-2 was originally estimated to be 1.4 kb long (2). The longest clones representing RNA-2 were between 1.4 and 1.5 kb, thus being full or close to full length (Fig. la). Comparative restriction map analysis of the three selected clones illustrated that pRCN005 and pRCN1085 were nearly identical in length except pRCN1085 was slightly longer at the 5' end. pRCN005 and pRCN004 were 5' co-terminal, but pRCN004 was approximately 150 bases shorter at the 3' terminus (Fig. lc). A large number of shorter, 5' terminal co-linear overlapping clones were generated. Considering that oligo(dT) was used to prime the polyadenylated RNAs, more 3' co-terminal clones were expected. RNA molecules with the 3' terminus missing may be preferred substrates for poly(A) addition, possibly because of 3'-terminal secondary structures in full-length RNA.

Sequencing of RNA-2 CDNA Clones

Three independent clones were sequenced to reduce the possibility of selection of a cloning artifact or a nonrepresentative RNA as template (Fig. lc). The sequence strategy is detailed in Fig. 1. Sequence was derived from M13 subclones except for the 3' terminus where no conveniently spaced restriction sites were present, and in the 3' to 5' direction at the 3'
Fig. 1. Map of cDNA clones, location of the 35 kDa open reading frame and sequence strategy for RCNMV RNA-2. (A) Schematic of RCNMV RNA-2 illustrating the length of RCNMV RNA-2 in bases beginning at the 5' end and the location of restriction sites used for M13 subcloning: A=Alu-I; B=Hsp-2; H=Hae-3; S=Sau-3A; T=Tag-1; and X=Xba-1. (B) Solid arrows indicate 5' terminal sequence determined directly by sequencing RNA with a Sau-3A-Alu-I primer. Open arrow indicates 3' end sequence determined by "wandering spot" analysis of RCNMV RNA-2. (C) RCNMV RNA-2 cDNA inserts used for dideoxy sequence analysis. The clones represent final nucleotide positions: pRCN005, 15-1451; pRCN1085, 7-1448; and pRCN004, 15-1278. Solid arrows indicate location, direction, and length of single stranded DNA templates obtained by subcloning into M13 and dashed arrows indicate sequenced single-stranded DNA templates derived from oligodeoxynucleotide primers. (D) Position and length of sequence coding for the 35 kDa open reading frame.

terminus where the homopolymer tracts disrupted the ability to clearly read sequencing gels. In these two cases oligonucleotides from residues 1266-1281 and 1432-1448 were synthesized and used as primers for dideoxy sequencing reactions (Fig. 1c).

Analysis of the sequences of the clones revealed that pRCN1085, pRCN005, and pRCN004 contained inserts of 1442, 1437, and 1264 nucleotides, respectively. The pRCN005 and pRCN004 inserts are 5' co-terminal and start at position 15 of the final derived sequence. pRCN1085 has an additional eight bases at the 5' end but has three nucleotides less than pRCN005 at its 3' end. In view of the known size of RCNMV RNA-2, it was clear that the
Fig. 2. Complete nucleotide sequence of HCNMV RNA-2 and derived putative amino acid sequence of the 35 kDa polypeptide. Amino acids encoded by the major open reading frame are given in one letter code below the first base of each codon in the frame. Numbers to the right of each line identify nucleotide and amino acid residue (Bold) of the last residue on that line. Underlined sequence identifies a second open reading frame with a putative polypeptide of 4.9 kDa.

three sequenced clones covered most of the RNA. However, it was still necessary to determine the sequences at both ends of the RNA directly.

Direct RNA Sequencing

To determine the sequence at the 3' end of RNA-2, gel-purified RNA was end labeled using [5'-32P]pCp and RNA ligase. Complete alkaline hydrolysis
revealed that RNA-2 ends with a "C". The sequence adjacent to this residue was determined by "wandering spot" analysis of partially alkali-degraded pCp-labeled RNA-2 (Fig. 3). The major pattern of spots corresponds to the sequence 5' AGGCAC(10)C_pCp_3', a sequence identical to that found at the 3' end of pRQN1085 but lacking the 3' terminal TCC found in pRQN005. This suggests that there is some heterogeneity at the 3' end in the RNA population, with the sequence found in pRQN1085 being the major species. This suggestion is corroborated by the fact that other, less intense, spots are visible on the autoradiogram shown in Fig. 3. Further evidence that the major RNA species in the population ends AGGCACA(10)C_pCp rather than AGGCACA(10)C_pCpU(10)C was provided by M-value analysis (17) of pCp-labeled RNA-2 which has been treated with alkaline phosphatase and partially degraded with nuclease P1. The results (not shown) were consistent with a sequence of 4 C residues at the 3' end and not with the sequence CUCC. Whether all species in the population are equally infectious must await the construction of full-length clones.

Fig. 3. Autoradiogram of "wandering spot" analysis of the [32p]pCp-labeled partial alkali digested RCMY RNA-2. Fractionation was by electrophoresis on cellulose acetate at pH 3.5 in the first dimension (1) and homochromatography in the second dimension (2).
The 5' terminal sequence was determined directly by runoff sequencing using reverse transcriptase, viral RNA-2 as template, and a restriction fragment representing bases 37-180 (Sau-3a-Alu-1) as primer (Fig. 1b). From these data it is clear that none of the clones represented the complete 5' termini. pRCN1085 was closest to representing the complete 5' terminus, missing only 5'-AAAACCT-3'.

Both [\(^{125}\)I] Bolton-Hunter (25) and Na \([^{125}\)I] Iodogen (26) methods were employed to label VPg proteins. Polyacrylamide gel electrophoresis of labeled samples with cowpea mosaic virus (27) as positive and tobacco mosaic virus (28) as negative controls indicated RCNMV RNA-2 either does not possess a 5' terminal VPg or has one without a free lysine or tyrosine residue (data not shown).

RCNMV RNA-2 was \([^{3}\)H] labeled with vaccinia virus nucleoside-2'-methyltransferase. This enzyme catalyzes the transfer of the \([^{3}\)H]-methyl group from S-adenosyl[methyl-\(^{3}\)H]methionine to the penultimate purine of either m\(^{7}\)GpppA or m\(^{7}\)GpppG (24). After nuclease P1 digestion of RNA-2, nearly

![Figure 4](image-url)

**Fig. 4.** Identification of the RCNMV RNA-2 5' terminal cap structure and penultimate purine by paper chromatography of \([^{3}\)H] methyl-labeled (21) RCNMV RNA-2 digestion products. A. Labeled RCNMV RNA-2 and TyMV RNA was digested with nuclease P1 and subjected to ascending paper chromatography in isobutyric acid/0.5 M NH\(_4\)OH. 10/6, v/v on Whatman No. 1 paper. Location of the capped standards are marked by solid line. B. Labeled material was eluted from the paper, digested with snake venom pyrophosphatase and bacterial alkaline phosphatase and subjected to paper chromatography in isopropanol/H\(_2\)O/ conc. NH\(_4\)OH, 7/2/1 on Whatman No. 1 paper. The location of the methylated purine residue standards are marked by solid lines.
all the labeled material co-chromatographed with the m⁷GpppA standard (Fig. 4a). Further digestion of the cap dinucleotide with snake venom pyrophosphatase and bacterial alkaline phosphatase indicated that only A was labeled (Fig. 4b). This data establishes that RCNMV RNA-2 contains a m⁷G cap and also that the penultimate purine is an "A". The penultimate residue analysis is in agreement with the 5' terminal "A" determined by the runoff sequence. Interestingly, RCNMV RNA-2 having a cap linked to an "A" residue (m⁷GpppA) is unlike most plant virus RNAs which contain a m⁷GpppG 5' terminal cap (29). Only the RNA-2 of tobacco rattle virus (30) and the turnip yellow mosaic virus cape id protein subgenomic messenger RNA (23) are known to have A-linked 5' terminal cap structures.

Open Reading Frame Analysis

Computer translation of the complete RNA sequence and its complement in all three reading frames revealed the presence of a single large open reading frame, initiating at the first AUG codon at base number 80 and terminating at an UAG codon at nucleotide position 1031 (Fig. 1d and 2). This open reading frame is capable of encoding a polypeptide of 317 amino acids residues, with a calculated molecular weight of 34,652 Da (hereafter referred to as the 35 kDa protein). Within this open reading frame are several in frame AUG codons (beginning at positions: 309, 312, 377, 428, and 688) and all generating nested open reading frames with the same stop codon at position 1031. The first initiation codon conforms well with the eukaryotic messenger RNA consensus sequence (31, 32). A purine residue is present at nucleotide position -3, and a G is present at nucleotide position +4. This is the only AUG codon near the 5' terminus that correlates with the Kozak sequence. For this reason it is likely that this is the true initiation codon for the observed 35 kDa polypeptide.

The next largest open reading frame begins at position 1224 and ends at 1358, with a potential to encode a polypeptide of 4,860 Da (Fig. 2). This open reading frame does not overlap the 35 kDa sequence and is shifted one frame. No polypeptides have been observed in vitro for this open reading frame. If this is a functional open reading frame, the question arises, how is it translated in vivo? No subgenomic RNA has been detected by Northern blot hybridization or dsRNA analysis (9) of the size (400 nucleotides) that could account for the expression for this open reading frame. Given the evidence presented, it is unlikely that this open reading frame is functional.
Fig. 5. Model of the possible base pairing the 3' terminal 96 nucleotides of RCNMV RNA-2. The structure was computed from the algorithm developed by Zuker and Stiegler (33). These same two stem-loop structures were generated when anywhere from 90 to 300 of the 3' terminal nucleotides were analyzed.

5' and 3' Noncoding Region Secondary Structure

The 5' noncoding leader sequence of RCNMV RNA-2 is 79 nucleotides, contains no start signals, and is within the size range observed for RNA genomes of other spherical plant viruses (29). This region has a high AT content (67%) with A representing 44% of the nucleotides. Secondary structure analysis (33) revealed little stable secondary structure in the 5' terminal noncoding region; this is typical for plant viruses.

Assuming no other functional open reading frames are present within RNA-2, a 3' terminal noncoding sequence of 417 nucleotides exists. Compared to most other single stranded RNA plant virus genomes, this is a long sequence. Generally, the 3' noncoding region is between 200 to 300 nucleotides in length (29). However, the unrelated tobacco rattle virus RNA-2 (CAM strain) is flanked by a 3' end leader sequence of 554 nucleotides (34).

As is the case for brome mosaic virus (35, 36), the 3' noncoding regions of viral RNAs presumably serve as replicase recognition sites and generally possess complex secondary structures. The 3'-terminus of RCNMV RNA-2 is not
Fig. 6. Map of transcripts from a RCNMV RNA-2 cDNA clone and subclones and fluorograph of rabbit reticulocyte in vitro translation of RCNMV RNA-2 and cDNA clone and transcripts. A. Diagram of RCNMV RNA-2 and restriction sites. B. Location of the 35 kDa open reading frame as deduced from the nucleotide sequence. Numbers above clone and transcripts indicate the nucleotide position of the termini. C. Map of the transcripts constructed from the RNA-2 cDNA clone, pRCN005. D. Fluorograph of RNA samples translated in vitro in the presence of [35S]-methionine and fractionated by electrophoresis in 12.5% SDS-polyacrylamide gels. Lane 1. Protein standard; myosin (H-chain) (200 kDa), phosphorylase b (97.4 kDa), bovine serum albumin (68 kDa), ovalbumin (43 kDa), α-chymotrypsinogen (25.7 kDa), β-lactoglobulin (18.4 kDa), and lysozyme (14.3 kDa). Lane 2. Electrophoretically isolated RCNMV RNA-2. Lane 3. SP6 positive polarity transcript from pRCN005. Lane 4. pGPX1029 transcript. Lane 5. pGPB911 transcript. Lane 6. pGXP1451 transcript. Lane 7. pRCN005 T7 negative polarity transcript. Lane 8. No exogenous RNA.
polyadenylated and does not appear to contain a typical tRNA-like structure. However, using the RNA folding algorithm of Zuker and Stiegler (33), two potential stem-loop structures between residues 1420 and 1444 (ΔG=−18.7 kcal/mol) and the 1355-1411 (ΔG=−21.8 kcal/mol) (Fig. 5) were generated. The stem loop structures were observed in all models formulated in 20 nucleotide increments between the 3' end, 90 to 300 nucleotides. With many of the models generated, the four 3' terminal C residues were paired with various groups of G residues depending on the number of nucleotides involved in the folding. The analysis suggests that these structures may exist and the 3' terminal C residues may be base paired as well. The presence of stable secondary structures was also implicated by the high G-C content and a number of G-C-rich regions within the 3'-terminal sequence. The stable secondary structure may have contributed to the high number of clones not representing the 3' terminus.

Transcription and Translation Studies

A series of positive polarity transcripts from pRCN005 and specific deletions from pRCN005 were generated and translated in vitro to determine the fidelity of the clone and verify the location of the open reading frame deduced from the sequence. The open reading frame for the 35 kDa polypeptide ends at base 1030. A subclone of pRCN005, pGPX1029 was constructed by deletion of the 3' end at the Xba-1 site (nucleotide 1029), which overlaps the putative stop codon (Fig. 6c). In order to generate a transcript that possessed the complete terminal codon, the clone was linearized with BamHI prior to the SP6 transcription reaction. A second deletion subclone (pGPB911) was constructed that terminated at the Bgl-2 site within the 35 kDa open reading frame. A third transcript was constructed representing nucleotides 1030-1451 (pGXP1029) to address the possibility of in vitro expression of the second small open reading frame deduced from the sequence data.

<table>
<thead>
<tr>
<th>Virus</th>
<th>Protein</th>
<th>Residue</th>
</tr>
</thead>
<tbody>
<tr>
<td>RCNMV</td>
<td>35 kDa</td>
<td>141-195</td>
</tr>
<tr>
<td>BMV</td>
<td>3a prot.</td>
<td>181-234</td>
</tr>
</tbody>
</table>

Fig. 7. Alignment of the amino acid sequence of the RCNMV RNA-2 35 kDa polypeptide and the BMV 3a polypeptide (38). Gaps (−) were introduced to affect best alignment. Two asterisks indicate residues are conserved. Single asterisk indicates residues are members of one of the following groups (46): acidic and polar (D,E,N,Q), basic (K,R), hydrophobic (A,C,F,H,I,L,M,V,W,Y), polar (T,S), or strong turn formers (D,G,N,P).
The transcript corresponding to pRCN005 directed the synthesis of a major polypeptide of 35 kDa with the same electrophoretic mobility as the major product synthesized from RCNMV RNA-2 (Fig. 6b). The negative polarity transcript did not direct the synthesis of a discernible polypeptide. The transcript from pGXP1029, with the 417 nucleotides deletion of the 3' terminus, also produced a 35 kDa polypeptide, confirming the location of the large open reading frame. As predicted from the sequence, the transcript from pGXP911 yielded a 30 kDa product. The physical data corroborates the location of the open reading frame deduced from the sequence. The transcript representing bases 1032-1451 which contains the 4.9 kDa open reading frame did not produce a polypeptide upon translation in vitro, suggesting but not definitively establishing that it is not functional in vivo.

**Homology Studies**

Using the FASTP (37) program and the GeneBank (NIH), European molecular biology, and National Biomedical Research Foundation databases, the primary structure of the deduced amino acid sequence of the 35 kDa polypeptide was compared and aligned with all available sequences. No striking homology was observed. The greatest sequence alignment was with the bromo mosaic virus (BMV) RNA 3a protein. There is 23.6% identity in a 54 amino acid overlap between residues 141-195 of the RCNMV 35 kDa and 181-234 of the BMV 3a polypeptides (38) (Fig. 7).

The alfalfa mosaic virus (AlMV) 32 kDa P3 polypeptide encoded by RNA3 has been implicated in potentiating cell-to-cell movement (39, 40). The homology between BMV, AlMV, cucumber mosaic virus, and tobacco mosaic virus (TMV) (41, 42) as well as the similar genome organization of AlMV and BMV, both members of the Tricoroviridae (43), suggest that the RNA 3a protein of BMV is involved in virus cell-to-cell movement. Therefore, it is interesting to note the homology between the putative cell-to-cell movement proteins of RCNMV and BMV. Among plant virus proteins speculated to be involved in cell-to-cell movement, little sequence conservation exists as noted by Meshi et al. (44). There is significant homology between the TMV 30 kDa and tobacco rattle virus 29 kDa polypeptides, but no homology between these and the BMV 3a or AlMV 32 kDa proteins (45). Unlike replicase proteins, cell-to-cell movement proteins appear to possess a similar function, but do not possess obvious conserved domains between distantly related viruses (44).

With the primary structure of RCNMV RNA-2 determined, we hope to now clarify the function of its product. The split genomic system of the dianthoviruses will enable us to critically address the relationship between cell-to-cell movement and functional host range.
ACKNOWLEDGMENTS

This work was supported in part by grants from the USDA (85-CRCR-1-1798 and 86-CRCR-1-2253) and personal awards from The McKnight Foundation and The CIBA Foundation to S.J.V. The authors would like to thank Dr. L. Calvert for aid with the computer graphics and sequence analysis, Dr. B. Moss for the gift of the vaccinia virus extract, and Dr. S. Muthukrishnan for help with the cap analysis. Contribution no. 88-172-J from the Kansas Agricultural Experiment Station, Kansas State University, Manhattan.

*To whom correspondence should be addressed

REFERENCES