Volume 11 Number 9 1983 Nucleic Acids Research

Complete nucleotide sequence of alfalfa mosaic virus RNA3

R.F.Barker, N.P.Jarvis, D.V.Thompson, L.S.Loesch-Fries and T.C.Hall

Agrigenetics Advanced Research Laboratory, 5649 East Buckeye Road, Madison, WI 53716, USA

Received 14 February 1983; Revised and Accepted 31 March 1983

ABSTRACT

A full-length cDNA clone of alfalfa mosaic virus (AMV) RNA3 was prepared and sequenced. The 2,037 base sequence contains two open reading frames of 903 and 666 nucleotides that code for a 32,400 dalton protein (32.4K protein) and the 24,380 dalton coat protein, respectively. A 5'-noncoding sequence of 240 bases preceeding the 32.4K protein contains homologous regions that may have a function in its translation. The intercistronic junction is 49 bases long, the last 36 bases representing the 5'-end of the subgenomic RNA4. The remaining 179 bases comprise the 3'-terminal noncoding sequence.

INTRODUCTION

Alfalfa mosaic virus (AMV) has a divided genome with RNAs of approximately 1.04×10^6 (RNA1), 0.73×10^6 (RNA2), and 0.68×10^6 (RNA3) daltons (1). In addition to these RNAs, a subgenomic RNA (RNA4) derived from RNA3 is encapsidated into virions (2). RNA4 is the messenger for coat protein; a sequence of 881 nucleotides has been previously reported for this RNA (3). All RNAs are believed to be monocistronic except for RNA3 which is dicistronic. *In vitro*, only one protein, thought to have a regulatory function *in vivo* (4), is translated from the 5'-cistron of RNA3. RNA4 apparently does not replicate autonomously (5) and as yet it is not understood how it is derived from the 3'-end of RNA3. We now report the complete sequence of RNA3, which should be of help in elucidating features involved in replication.

MATERIALS AND METHODS

Materials

Ultra pure urea, bovine serum albumin (BSA) and *E. coli* poly(A) polymerase were obtained from BRL (Gaithersburg, Maryland), polyacrylamide from BDH (Poole, England), calf intestinal alkaline phosphatase from Boehringer (Mannheim, W. Germany), polynucleotide kinase and terminal transferase from P.L. Biochemicals, Inc. (Milwaukee, Wisconsin), and [γ-^{32}P]ATP, [^{3}H]ATP and

[α-^{32}P]dCTP from New England Nuclear (Boston, Massachusetts). RNasin and the restriction enzymes AvaI, BamHI, HincII and PstI were from Biotech (Madison, Wisconsin). ClaI and RsaI were purchased from New England Biolabs (Beverly, Massachusetts). Avian myeloblastosis virus reverse transcriptase was from Life Sciences, Inc. (St. Petersburg, Florida).

Isolation of RNA

AMV (strain 425) was purified from infected Nicotiana tabacum L. 'Xanthinc' as previously described (6). RNA, prepared by phenol extraction of the virus, was enriched for RNAs 1,2, and 3 by three cycles of linear-log sucrose gradient centrifugation (7).

cDNA Cloning

AMV RNA was polyadenylated using a modification of the procedure described by Sippel (8). A reaction of 70µl contained 10 mM $MgCl_2$, 2.5 mM $MnCl_2$, 250 mM NaCl, 28 µg BSA, 4 µg viral RNA, 56 nmol ATP, 1.4 units poly(A) polymerase and 2.3 nmol [^{3}H]ATP (20 Ci/mmol) in 50 mM Tris-HCl, pH 7.9. This procedure allowed an average of 80 adenosine residues to be added to the 3'-end of RNA3 in 30 minutes with minimal RNA degradation. Reverse transcription was primed with oligo(dT)$_{12-18}$. Double stranded DNA was synthesized and tailed with oligo(dC) by the method described by Land et al. (9) which avoids the use of S1 nuclease. This was annealed to PstI-cut, oligo(dG)-tailed pBR322 and transformed into E. coli HB101(10). Tet^r-Amp^s colonies were transferred to nitrocellulose (11) and screened by using an RNA3 probe prepared as follows: AMV RNA3 was excised from a denaturing agarose gel, randomly nicked under alkaline conditions, and 5'-labeled with [^{32}P] using polynucleotide kinase. Colonies that hybridized strongly were selected, plasmid DNA was isolated (12) and restriction maps of the clones were determined.

DNA Sequencing

The sequencing strategy for the cDNA clone of RNA3 selected for sequencing (pAMV170) is outlined in Fig. 1. The cDNA clone was cut with a suitable restriction enzyme, treated with alkaline phosphatase, and 5'-labeled using polynucleotide kinase. The double end-labeled fragments were then recut with suitable restriction enzymes to produce single end-labeled fragments which were then separated on and eluted from a polyacrylamide gel (procedures 4, 5a, 7 and 9 of Maxam & Gilbert) (13). The DNA sequencing reactions were then carried out as described (13) with the following modifications. The limiting G + A reaction as described by Cooke et al. (14) was used. The incubation time for the G reaction was reduced to 30 seconds at 20°C and the times for the G + A, C + T and C reactions were reduced to three minutes at 20°C. These

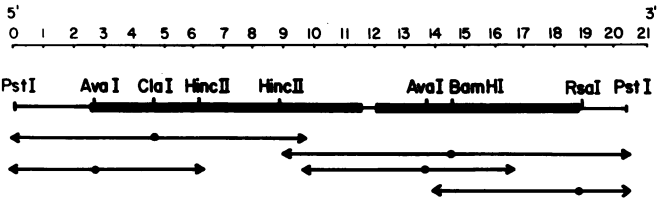

Fig. 1. Physical map and sequencing strategy for cDNA clone pAMV170. The scale is in hundreds of nucleotides. The darker lines indicate the positions of the 32.4K protein and the coat protein messenger RNA. The dots indicate the position of the 5'-end labeling and the arrows show the direction and length of sequencing.

modifications resulted in increased resolution in the sequence ladders, especially for fragments longer than 400 bases. The electrophoretic system used here was a modification of the system described by Garoff and Ansorge (15). Ultra thin gels of 0.2mm thickness were chemically bonded to one of the glass support plates. Electrophoresis was carried out at a constant temperature of 50°C maintained by a thermostating plate. The modified sequencing system used 104 cm × 22 cm gels at three polyacrylamide concentrations (4, 6 and 16%). By applying each 5'-labeled fragment to all three concentrations of gel, we were able to sequence routinely an average of 500 bases per fragment. Computer analyses of the DNA and protein sequences were performed using computer programs made available by Drs. O. Smithies and F. Blattner (University of Wisconsin, Madison).

RESULTS AND DISCUSSION

cDNA Cloning

Using the cDNA cloning conditions of Land et al. (9), we were able to obtain a high proportion of long cDNA copies of the AMV RNA genome. Of 171 recombinant colonies screened, eight hybridized strongly with AMV RNA3. These clones were found to have plasmids containing various overlapping portions of a single restriction map. The clone containing the longest insert (pAMV170) was selected for sequencing. It evidently represents a full-length copy because it terminates in sequences previously determined for the 5'- and 3'-ends of RNA3 (16,17).

Sequencing

General Features of AMV RNA3. The complete nucleotide sequence of AMV RNA3 obtained from the cDNA clone pAMV170 and the deduced amino acid sequences are shown in Fig. 2. At the 5'-end is a noncoding sequence of 240 nucleo-

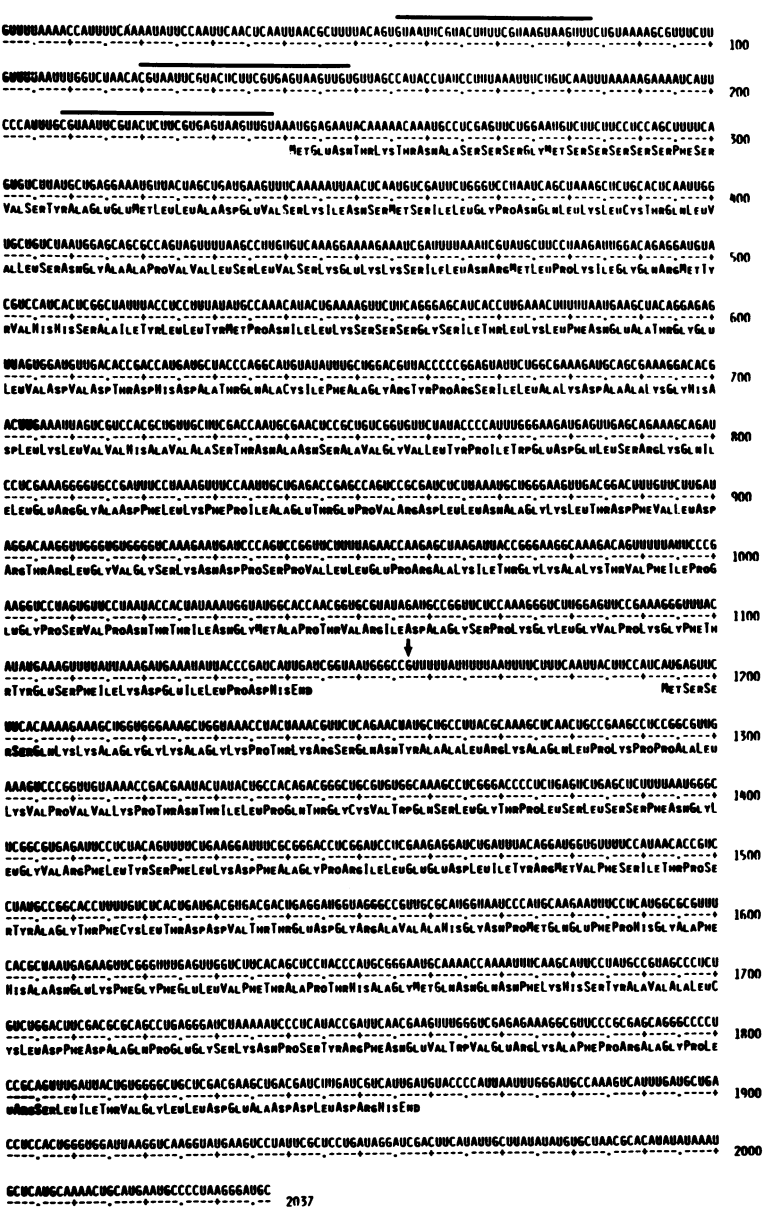

Fig. 2. The complete nucleotide sequence of AMV RNA3 cDNA clone pAMV170. The amino acid sequence of the 32.4K protein and the coat protein is shown beneath the nucleotide sequence. The arrow at position 1157 indicates the 5'-end of RNA4 in the intercistronic junction. The heavy lines indicate homologous repeat sequences in the 5'-terminal noncoding region.

tides. The terminal 5'-end, with three base substitutions, corresponds to the sequence of 100 bases reported by Koper-Zwarthoff et al. (16). The coding region closest to the 5'-end is 903 nucleotides and codes for a protein of 32,400 daltons, referred to as the 32.4K protein. The second coding region of 666 nucleotides codes for a 24,380 dalton protein, which is the coat protein. There is a short intercistronic junction of 49 nucleotides. The remaining 179 bases of the terminal 3'-noncoding sequence agrees with that reported by Koper-Zwarthoff et al. (17) with three base substitutions. The total length of AMV RNA3 is 2,037 bases and the base content is G, 22.7%, A, 26.5%, U, 29.2%, C, 21.6%.

AMV RNA3 is, in overall structure, similar to RNA3 of brome mosaic virus (BMV). Both are dicistronic with coding sequences for a 32K protein at the 5'-end and the coat protein at the 3'-end (18). We have, however, found little sequence homology between comparable regions of AMV RNA3 and BMV RNA3.

Terminal 5'-Noncoding Leader Sequence. When the 240 base 5'-noncoding sequence was compared with the published data of AMV strain 425 (16), three base substitutions at positions 3, 8, and 79 were found. The U to A change at position 79 eliminates the reported AUG codon from the sequence. None of these changes, however, affects the secondary structure proposed by Koper-Zwarthoff, et al. (16). Within the 5'-leader sequence there are three repeats of 28-30 bases (Fig. 3A). These regions have previously been reported for AMV strain S by Pinck, et al. (19). From the data presented here, three secondary structures can be postulated which include these regions (Fig. 3B). Regions 2 and 3 are identical, however, region 1 differs at two positions (Fig. 3A). These differences do not affect the secondary structure of region 1 (Fig. 3B). The initiation codon for the 32.4K protein occurs on the decending limb of the third structure. From analysis of ribosome protected fragments, Pinck et al. (19) found that two of the homologous regions that are part of stem-loops 2 and 3 (fig. 3B) were involved with ribosome binding. The first 22 nucleotides of AMV strain 425 and the first 37 nucleotdies of AMV strain S show little homology. The remaining nucleotides of the 5'-terminal noncoding region show 97% homology and the repeat sequences are maintained. This suggests some functional significance for the secondary structures and evolutionary conservation of the ribosome binding sites. Other plant viruses that have long 5'-terminal noncoding leader sequences have been reported to bind more than one ribosome under conditions where elongation of protein synthesis was inhibited (20,21).

When compared with the total base content of RNA3, the composition of the

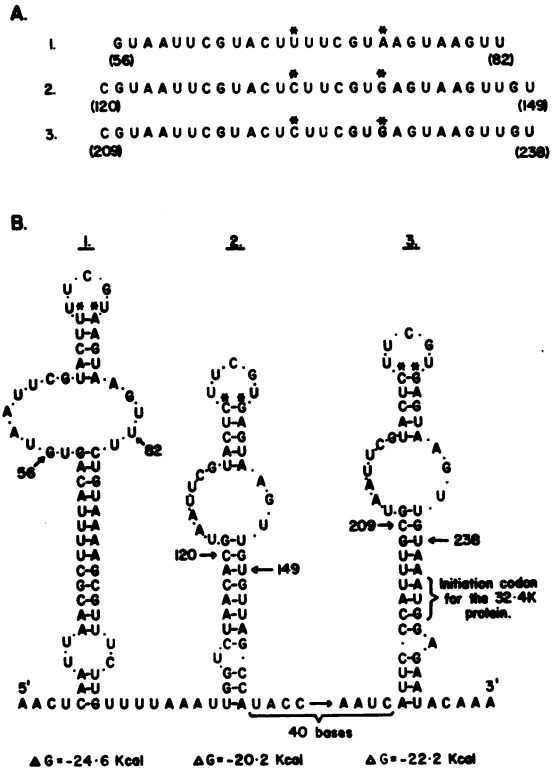

Fig. 3. (A) Nucleotide sequences of the homologous repeat regions at the 5'-terminal noncoding end of AMV RNA3. The numbers represent the nucleotide positions from the 5'-end. * indicates non-homologous bases. (B) Possible secondary structures showing the location of the homologous regions and their base substitutions. These structures may play a role in ribosome binding. The initiation codon for the 32.4K protein is indicated on stem-loop 3.

5'-noncoding sequence is markedly low in G residues (13%) and high in U residues (40%) (Table 1). Table 1 also shows that the regions surrounding the stem-loop structures account for the low G, high U content. The stem-loop regions have a base content similar to that of the whole RNA3. The reason for a prejudice against G residues in the terminal 5' noncoding sequence is not understood, but these results are in agreement with those for other 5'-noncoding leader sequences from plant viral RNAs (22,23).

The 32.4K Protein of AMV RNA3. The first AUG codon occurs at position 241 and initiation of translation is most likely to occur here (24). This corresponds to the AUG codon in strain S that Pinck et al. (19) suggest as being the initiation codon for the 32.4K protein. An open reading frame

Table 1. PERCENTAGE BASE CONTENT FOR VARIOUS REGIONS OF AMV RNA3

	Base Content			
	%G	%A	%U	%C
Total RNA3	23	26	29	22
Terminal 5'-noncoding sequence	13	30	40	17
Stem-loop 1	17	26	41	16
Stem-loop 2	24	24	31	21
Stem-loop 3	22	27	35	16
5'-noncoding sequence flanking the stem-loops	4	39	40	17
Intercistronic junction	14	19	50	17
32.4K protein cistron	24	28	28	20
Coat protein cistron	26	24	25	25
Terminal 3'-noncoding sequence	21	28	30	21

Table 2. CODON USAGE FOR AMV 32.4K PROTEIN CISTRON

Ala(A)	GCA	5	Gln(Q)	CAA	1	Leu(L)	CUU	7	Ser(S)	UCG	5
	GCC	3		CAG	4		UUA	8		UCU	7
	GCG	4	Glu(E)	GAA	10		UUG	6	Thr(T)	ACA	6
	GCU	12		GAG	6	Lys(K)	AAA	10		ACC	7
Arg(R)	AGA	2	Gly(G)	GGA	7		AAG	12		ACG	2
	AGG	4		GGC	0	Met(M)	AUG	8		ACU	2
	CGA	0		GGG	5	Phe(F)	UUC	1	Trp(W)	UGG	1
	CGC	1		GGU	8		UUU	8	Tyr(Y)	UAC	4
	CGG	1	His(H)	CAC	3	Pro(P)	CCA	7		UAU	3
	CGU	3		CAU	3		CCC	5	Val(V)	GUA	1
Asn(N)	AAC	3	Ilu(I)	AUA	5		CCG	2		GUC	5
	AAU	11		AUC	2		CCU	4		GUG	6
Asp(D)	GAC	4		AUU	11	Ser(S)	AGC	4		GUU	10
	GAU	12	Leu(L)	CUA	4		AGU	6	End(.)	UAA	0
Cys(C)	UGC	1		CUC	4		UCA	6		UAG	0
	UGU	1		CUG	4		UCC	3		UGA	1

extends from position 241 through the UGA stop codon to position 1143. This sequence of 903 bases codes for a protein having a of molecular weight of 32,400.

The codon usage for the AMV 32.4K protein is given in Table 2. Ala, Asn, Asp, Gln, Ilu, and Phe show some preference in codon usage. No preference is shown between purines and pyrimidines at the third base position. With codons

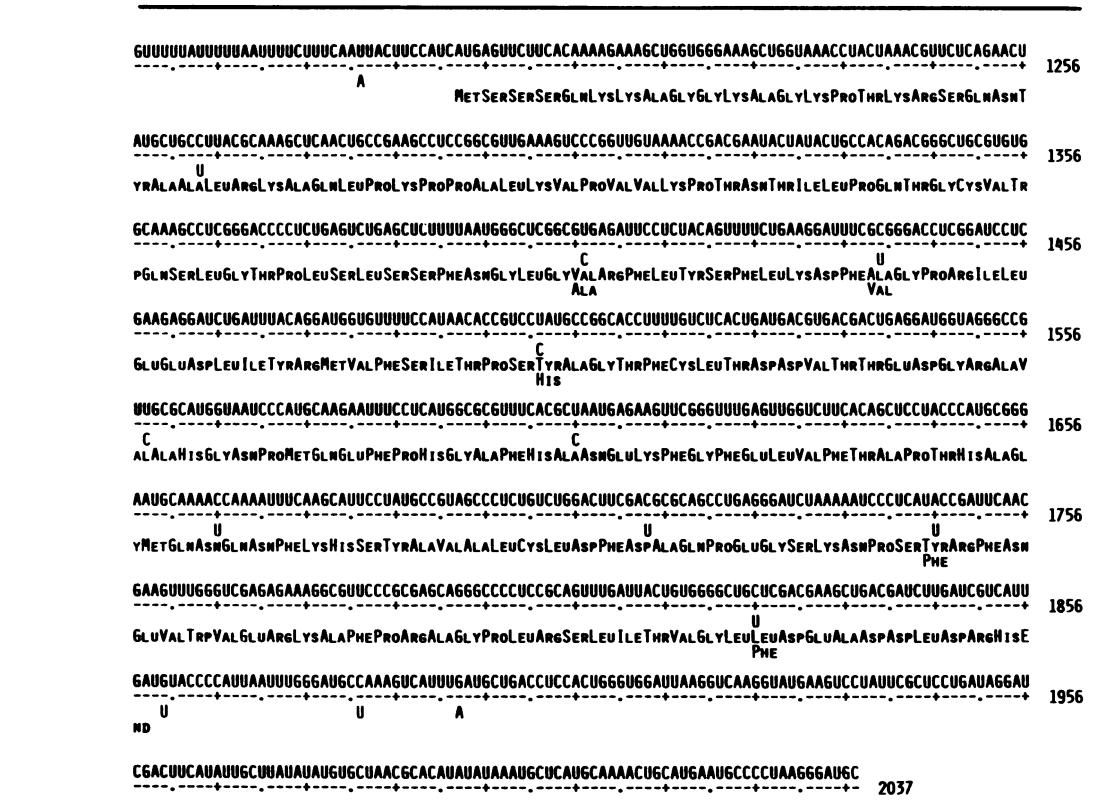

Fig. 4. The nucleotide sequence of RNA4 obtained from sequencing four individual cDNA clones. Where there are differences between this sequence and that published by Brederode et al. (3), the base changes are indicated. Amino acid differences are also shown. The nucleotide numbering corresponds to that for RNA3.

ending with pyrimidines, however, there is a significant preference for U over C (2.2:1).

Intercistronic Junction. The intercistronic junction in AMV RNA3 consists of 49 bases. Our sequence differs from that of Brederode et al. (3) by only one nucleotide (Fig. 4), and this is at a position where variation has been previously reported (25). A complex secondary structure, proposed by Koper-Zwarthoff et al. (26), that contains the intercistronic region is maintained. A large portion of the intercistronic junction is the 5'-leader sequence of AMV RNA4 which is characteristically low in G residues. This results in a low G (14%) and high U (50%) content for this junction (Table 1). We do not find a continuous reading frame between the 32.4K protein and the coat protein such as that suggested by van Vloten-Doting et al. (27).

Table 3. CODON USAGE FOR AMV COAT PROTEIN CISTRON

Ala(A)	GCA	1	Gln(Q)	CAA	6	Leu(L)	CUU	1	Ser(S)	UCG	0
	GCC	5		CAG	3		UUA	1		UCU	4
	GCG	7	Glu(E)	GAA	4		UUG	3	Thr(T)	ACA	2
	GCU	7		GAG	6	Lys(K)	AAA	8		ACC	3
Arg(R)	AGA	2	Gly(G)	GGA	3		AAG	6		ACG	3
	AGG	2		GGC	4	Met(M)	AUG	4		ACU	5
	CGA	2		GGG	6	Phe(F)	UUC	8	Trp(W)	UGG	2
	CGC	2		GGU	4		UUU	7	Tyr(Y)	UAC	3
	CGG	1	His(H)	CAC	1	Pro(P)	CCA	1		UAU	3
	CGU	2		CAU	5		CCC	3	Val(V)	GUA	2
Asn(N)	AAC	3	Ilu(I)	AUA	2		CCG	6		GUC	3
	AAU	6		AUC	1		CCU	7		GUG	5
Asp(D)	GAC	5		AUU	2	Ser(S)	AGC	2		GUU	3
	GAU	6	Leu(L)	CUA	0		AGU	4	End(.)	UAA	0
Cys(C)	UGC	1		CUC	8		UCA	2		UAG	0
	UGU	2		CUG	8		UCC	3		UGA	1

AMV Coat Protein. In AMV RNA3 the initiation codon for the coat protein is at position 1193. Following this is an open reading frame of 663 bases concluding with a UGA stop codon. This sequence codes for a protein with a molecular weight of 24,380. To confirm our sequence data for the coat protein cistron, three independent cDNA clones approximately the size of AMV RNA4 were also sequenced. Although these cDNA clones varied in length (745, 876 and 982 bases), they all contained the 3'-end. The sequences of these three independent cDNA clones were in complete agreement. The sequence shown in Fig. 4 shows 14 base and five amino acid substitutions from the published sequence (3). Although both cultures of AMV strain 425 originated from the same isolate (28), they had been maintained independently for many years. Our sequence data indicates that these cultures are now slightly different. However, the open reading frame of 138 amino acids in the minus strand of AMV RNA4 reported by van Vloten-Doting (29) was still maintained in the complement of the sequence presented here.

The codon usage for the coat protein cistron is shown in Table 3. This was similar to the published data of Brederode et al. (3) except for five amino acid substitutions. There was little preference for U over C (1.2:1) in codons ending with pyrimidines, whereas there was a significant bias for U over C (2.2:1) for the 32.4K protein.

3'-Terminal Noncoding Sequence. The 3'-terminal noncoding sequence consists of 179 bases (Fig. 4), and is similar to all of the published data on this region (17,26,30,31). The three base substitutions found in this area do not affect any of the postulated secondary structures for coat protein binding

(32,33). The G to U substitution at position 1860 created an RsaI site which was utilized during the sequencing strategy (Fig. 1).

In conclusion, from the entire sequence of AMV RNA3 presented here, we can identify highly homologous repeat sequences in the 5'-leader region which can form stable secondary structures incorporating ribosome binding sites described by Pinck, et al. (19). The clone pAMV170 contains sequences complementary to the entire RNA3; therefore, we will be able to investigate the effect of specific base changes on the function of the 5'-leader sequences as well as on the mechanisms of subgenomic RNA4 synthesis.

ACKNOWLEDGEMENTS

We thank Dr. Robert W. Fulton for providing the culture of AMV strain 425, Steve E. Nelson for excellent technical assistance, Conrad Brown for computer analyses, and Jo Adang for assistance with the figures. We are grateful to Annette Nassuth for providing us manuscripts prior to publication.

REFERENCES:

1. Heijtink, R. A., Houwing, C. J., and Jaspars, E. M. J. (1977) Biochemistry 16, 4684-4693.
2. van Vloten-Doting, L. and Jaspars, E. M. J. (1977) in Comprehensive Virology, Fraenkel-Conrat, H. and Wagner, R. R. Eds., Vol 11, pp 1-53, Plenum Press, New York, N.Y.
3. Brederode, F. Th., Koper-Zwarthoff, E. C. and Bol, J. F. (1980) Nucleic Acids Res. 8, 2213-2223.
4. Nassuth, A. and Bol, J. F. (1983) Virology 124, 75-85.
5. Smit, C. H. and Jaspars, E. M. J. (1982) Virology 117, 271-274.
6. van Vloten-Doting, L. and Jaspars, E. M. J. (1972) Virology 48, 699-708.
7. Brakke, M. K., and van Pelt, N. (1970) Anal. Biochem. 38, 56-64.
8. Sippel, A. E. (1973) Eur. J. Biochem. 37, 31-40.
9. Land, H., Grez, M., Hauser, H., Lindemaier, W., and Schütz, G. (1981) Nucleic Acid Res. 9, 2251-2266.
10. Dagert, M. and Ehrlich, S. D. (1979) Gene 6, 23-28.
11. Grunstein, M. and Wallis, J. (1979) in Methods in Enzymology, Wu, R., Ed. Vol 68, pp. 379-389. Academic Press, New York, N.Y.
12. Birnboim, H. C. and Doly, J. (1979) Nucleic Acids Res. 7, 1513-1523.
13. Maxam, A. M., and Gilbert, W. (1980) in Methods in Enzymology, Grossman, L. and Moldave, K., Eds. Vol 65, pp. 499-560. Academic Press, New York, N.Y.
14. Cooke, N. E., Coit, D., Weiner, R. I., Baxter, J. D., and Martial, J. A. (1980) J. of Biol. Chem. 255, 6502-6510.
15. Garoff, H. and Ansorge, W. (1981) Anal. Biochem. 115, 450-457.
16. Koper-Zwarthoff, E. C., Brederode, F. Th., Veeneman, G., van Boom, J. H., and Bol, J. F. (1980) Nucleic Acids Res. 8, 5635-5647.
17. Koper-Zwarthoff, E. C., Brederode, F. Th., Walstra, P., and Bol, J. F. (1979) Nucleic Acids Res. 7, 1887-1900.
18. Ahlquist, P., Luckow, V., and Kaesberg, P. (1981) J. Mol. Biol. 153, 23-38.

19. Pinck, M., Fritsch, C., Ravelonandro, M., Thivent, C., and Pinck, L. (1981) Nucleic Acids Res. 9, 1087-1100.
20. Filipowicz, W. and Haenni, A-L. (1979) Proc. Natl. Acad. Sci. USA 76, 3111-3115.
21. Ahlquist, P., Dasgupta, R., Shih, D. S., Zimmern, D., and Kaesberg, P. Nature 281, 277-282.
22. Goelet, P., Lomonossoff, G. P. Butler, P. J. G., Akam, M. E., Gait, M. J., and Karn, J. (1982) Proc. Natl. Acad. Sci. USA 79, 5818-5822.
23. Briand, J-P., Keith, G., and Guilley, H. (1978) Proc. Natl. Acad. Sci. USA 75, 3168-3172.
24. Kozak, M. (1978) Cell 15, 1109-1123.
25. Swinkels, P. P. H., and Bol, J. F. (1980) Virology 106, 145-147.
26. Koper-Zwarthoff, E. C. and Bol, J. F. (1979) Proc. Natl. Acad. Sci USA 76, 1114-1117.
27. van Vloten-Doting, L. (1976) Ann. Microbiol. (Paris) 127A, 119-129.
28. Hagedorn, D. J. and Hanson, E. W. (1963) Phytopathology 53, 188-192.
29. van Vloten-Doting, L., Dubelaar, M., and Bol, J. F. (1982) Plant Molecular Biology 1, 155-158.
30. Pinck, L. and Pinck, M. (1979) FEBS Lett. 107, 61-65.
31. Gunn, M. R. and Symons, R. H. (1980) FEBS Lett. 109, 145-150.
32. Koper-Zwarthoff, E. C. and Bol, J. F. (1980) Nucleic Acids Res. 8, 3307-3318.
33. Houwing, C. J. and Jaspars, E. M. J. (1982) Biochemistry 21, 3408-3414.